WORKING PAPER · NO. 2021-107

Energy and Environmental Markets, Industrial Organization, and Regulation

Ryan Kellogg and Mar Reguant

SEPTEMBER 2021

EPIC
ENERGY POLICY INSTITUTE
AT THE UNIVERSITY OF CHICAGO

An Affiliated Center of
Becker
Friedman
Institute

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

ENERGY AND ENVIRONMENTAL MARKETS, INDUSTRIAL ORGANIZATION, AND REGULATION

Ryan Kellogg
Mar Reguant

September 2021

This manuscript is an invited chapter for the forthcoming volume of the 4th Handbook of Industrial Organization by Elsevier. We thank the editors and four anonymous referees for helpful suggestions while preparing this chapter.

Energy and Environmental Markets, Industrial Organization, and Regulation
Ryan Kellogg and Mar Reguant
September 2021
JEL No. L0,Q2,Q3,Q4,Q5

ABSTRACT

This paper discusses contributions that industrial organization economists have made to our understanding of energy markets and environmental regulation. We emphasize the substantive contributions of recent papers while also highlighting how this literature has adopted and sometimes augmented theoretical and empirical tools from industrial organization. Many of the topics examined by this literature—especially auctions, investment, productivity and innovation, and regulation—also apply to a variety of settings beyond energy and the environment. We also indicate areas where future research is likely to be fruitful, with an emphasis on how industrial organization economists can help inform energy and environmental policies.

Ryan Kellogg
University of Chicago
Harris School of Public Policy
1307 East 60th Street
Chicago, IL 60637
and NBER
kelloggr@uchicago.edu

Mar Reguant
Department of Economics
Northwestern University
302 Donald P. Jacobs
2001 Sheridan Road
Evanston, IL 60208
and NBER
mar.reguant@northwestern.edu

Contents

1 Introduction

Since the 1990s, growing concerns about energy access, the costs of energy supply, and the environmental damages associated with fossil fuels have led to dramatic changes to energy and environmental markets and regulation. Examples include the restructuring of the electricity industry away from vertically integrated utilities and towards independent power producers that compete in wholesale power auctions, introduction of cap-and-trade markets for greenhouse gases (GHGs) and local air pollutants such as NO_x, tightening of fuel economy standards, and adoption of policies such as renewable portfolio standards to promote zero-emission energy sources. These policies have been enacted with varying stringency across jurisdictions over time, and it is clear that policy change will continue into the future, especially in the area of climate policy.

This evolution of regulation and policy has taken place concurrently with remarkable innovation and productivity improvements in energy production. For fossil fuels, the U.S. shale oil and gas revolution that started in the mid-2000s was driven by unprecedented productivity growth that led to booms in mineral leasing, oil and gas production, and crude-by-rail transportation. And for renewables, the costs of wind and solar electric generation have declined markedly in recent years, such that in some areas they can compete with fossil fuel based generation on the merits.

Changes in energy and environmental technology, industry structure, and regulation are likely to be even more dramatic in the years to come with the pressing need to reduce greenhouse gases (GHG) emissions. The sectors covered in this handbook chapter, either as inputs or as direct outputs, are responsible for over 70% of GHG emissions contributing to global warming. Learning about these sectors is therefore crucial for any researcher interested in climate change economics.

Understanding the changes in energy supply, energy markets, and regulations—in both a positive and normative sense—requires addressing questions that are well-suited for the theoretical and empirical tools from industrial organization. For instance, wholesale electricity markets take the form of high-frequency multi-unit auctions, vehicle markets can be characterized by differentiated Bertrand competition, extractors of exhaustible resources face dynamic problems, and energy distribution firms are regulated natural monopolies. These market structures and forms of regulation are all areas where IO economists hold expertise. Our main goal for this chapter is to illustrate how ideas and tools from IO have been marshalled to create insights into understanding energy and environmental economics and policy.

Our second goal in this chapter is to highlight areas where insights from energy and environmental markets are likely to be relevant for other settings of interest to IO economists, and cases where methods developed to address a particular question might have broader applicability. A convenient feature of energy and natural resource intensive industries for IO researchers is that they tend to be government regulated or at least government monitored, so that detailed administrative data are available on inputs, outputs, prices, costs, and investment. These data enable researchers to address broad IO questions that would be challenging if not impossible to answer credibly using data from other industries.

Given our primary goal of discussing how ideas and tools from IO have helped answer pressing, policy-relevant questions about energy and environmental markets and regulation, we have organized this chapter by substantive topic area. We begin in section 2 by discussing IO economists' contributions to our understanding of markets and regulatory policies in the primary energy resource extraction sector. Section 3 then discusses markets and policies for personal transportation, and section 4 discusses the electricity sector. Given the growing importance of environmental policies—especially policies targeting GHG emissions—for these markets, each of these three sections emphasizes contributions that speak, directly or indirectly, to the impacts of such policies. Then in section 5 we discuss environmental regulation and enforcement more broadly, including regulation of energy-intensive manufacturing industries and markets for water resources.

In each section of this chapter, most of our writing is dedicated to discussing the contributions of IO economists to the issues at hand, but we also take time as needed to provide industry background that is often a prerequisite to conducting research in these areas. Our discussions also include subsections that are forward-looking: what questions remain unanswered, and how might IO economists make contributions towards answering them? We then conclude the chapter in section 6 by summarizing what we see as the most promising areas for future contributions from IO to energy and environmental economics and policy, with an emphasis on research that helps inform policies aimed at accelerating a clean energy transition.

Our organizational approach will be most useful for researchers and students who are interested in understanding a particular substantive area. For instance, someone who is interested in fuel economy policy can proceed directly to that section to understand what the compelling research questions have been, how IO economists have contributed to those questions thus far, and what important unanswered questions remain. This organization is, though, perhaps less well-suited to those who are interested in applications of specific models or methods, such as auction models or productivity estimation methods.

Methods cross-walk. To help guide readers interested in particular methods and tools, rather than topical areas, we conclude this introduction with the following "cross-walk" that indicates where readers can find applications of—and in many cases, innovations to—models and tools that are broadly used across IO.

1. **Auction models** have been essential for understanding both the leasing of primary energy resources and the economics of wholesale electricity markets.
   (a) Section 2.2 discusses oil and gas leasing, which has been a long-standing setting for studies of auction models (Hendricks and Porter, 2007). This section includes discussions of auctions with contingent payments (oil and gas royalties), implications of bidders' uncertainty regarding the number of competitors, sequential versus bundled multi-unit auctions, comparisons between auctions and unstructured negotiations, and implications of post-auction secondary markets.
   (b) Section 4.2 discusses how the study of electricity markets has contributed to the advancing of our understanding of multi-unit uniform price auctions. Electricity markets provide a unique setting in which auctions happen very frequently, the players are well-known, and their cost structure and objective function are well-understood.
2. **Productivity estimation** has been central to understanding the dramatic recent reductions in the cost of producing shale oil and gas and of generating electricity from wind and solar resources. It has also been important for evaluating the consequences of electricity industry restructuring.

    (a) Section 2.4 discusses how research on the shale oil and gas industry has provided evidence for forms of productivity improvements that have previously not garnered attention from the literature on productivity and innovation, including learning about optimal input selection, improving the selection of which projects to complete, and learning how to work more efficiently with other firms in the vertical supply chain.
    (b) Production function estimation has also been used in the context of electricity markets to assess the productivity impacts of regulatory reform, discussed in section 4.1, and reductions in the costs of renewable power generation, discussed in section 4.3.
    (c) Section 5.3.2 discusses productivity analysis applied to agricultural production to quantify the misallocation of water.

3. **Differentiated product models** have been widely used in work on transportation and the environment to understand automobile consumers' demand for fuel economy (section 3.1), automakers' responses to fuel economy standards (section 3.2), the economics of regulations on vehicles' emissions of local pollutants (section 3.3), and markets for electric vehicles (section 3.5). This work frequently augments standard Berry et al. (1995) style models of automobile markets to include vehicle characteristics related to environmental performance and constraints imposed by environmental regulations.

4. **Models of consumer search** have been used to study how drivers search for retail gasoline and how households search for retail electricity providers.
    (a) Section 3.4 discusses how data from retail gasoline markets have been used to empirically test models of consumer search. This work has taken advantage of high-frequency spatial data on retail prices and on search intensity itself.
    (b) Section 4.4 discusses how consumer search models have been used to understand competition in liberalized electricity markets.

5. **Single-agent dynamic models** are key to modeling the dynamic aspects in energy and environmental markets.
    (a) Dynamic models have long been used, dating back to Hotelling (1931), to understand extraction of exhaustible resources. Theoretical and empirical applications of these models to the oil and gas industry are discussed throughout section 2 and especially in section 2.1.
    (b) Single-agent models have also been used extensively to model the behavior of electricity market participants, and discussions of such applications are provided in sections 4.3 and 4.4.
    (c) Section 5.2.2 discusses recent developments in structural models of regulation enforcement and monitoring that also use these dynamic methods.

6. **Models of dynamic games** are often important because energy industries are often characterized by long-lived capital investments and oligopolistic competition.

   - (a) Section 2.3 discusses how models of dynamic games have helped build an understanding of firms' oil and gas exploration behavior. Even though the large number of private oil and gas firms typically lends itself to modeling these firms using single-agent dynamic models, in local exploration contexts there can be important information spillovers across just a small set of firms, implying that models involving dynamic games are required to understand investment behavior.
   - (b) Section 5.1.2 shows how current state-of-the-art dynamic games modeling can be applied to understand environmental leakage applications, and section 4.3.3 discusses the importance of dynamics and learning-by-doing in the context of renewable technologies.

7. **Natural monopoly regulation** Many energy sectors have been traditionally regulated (and some still are) as natural monopolies, making it one of the leading applications of theoretical and empirical work in this area. In section 4.1, we discuss work examining the regulation of natural monopolies in the distribution of electricity and natural gas. Section 4.5 discusses how natural monopoly regulation interacts with environmental regulation.

8. **Network goods.** Electric vehicles and electric vehicle charging stations are subject to indirect network effects, and we discuss how network goods models have been applied to empirically study this industry in section 3.5.

9. **Vertical relationships** appear throughout the energy industry:
   - (a) Section 2.4 discusses mechanisms by which repeated interactions between well-matched oil producers and drilling rigs can improve productivity.
   - (b) Section 2.5 discusses a variety of ways in which vertical relationships between monopoly or near-monopoly railroads and coal-fired electric generators can create challenges for regulation of these generators' pollution emissions.
   - (c) Section 4.2 discusses the importance of vertical positions in the determination of incentives to exercise market power in the electricity sector.

2 Extraction of energy resources

This section discusses the industrial organization of primary energy resource extraction. This topic has recently experienced a surge in research, which we see as stemming from two high-level motivations: (1) topical importance that is driven by both the U.S. shale oil and gas boom and concerns about the environmental consequences of oil and gas production; and (2) unparalleled availability of data on investments in individual oil wells and on lease contracts between firms and mineral owners. These data can be used to answer questions about topics such as firms' investment behavior, oil and gas lease auctions and allocations, productivity growth, and impacts of environmental regulation.

The oil and gas industry is one of the largest sectors of the economy, has long been thought to be entwined with macroeconomic outcomes, and can experience dramatic short-run and long-run fluctuations. In 2019, total world oil production was 95 million barrels per day, worth \$6.1 billion per day (\$2.2 trillion per year) at 2019 average prices.[1] These large magnitudes, combined with oil price volatility arising from the inelasticity of short-run global demand and supply for oil, has long led economists and policy analysts to worry about macroeconomic impacts from changes in oil supply (Hamilton, 2013; Kilian and Vigfusson, 2017; Baumeister and Hamilton, 2019). Interest in understanding oil supply has only increased since the U.S. shale oil and gas boom began in the mid-2000s, more than doubling U.S. oil and gas production and leading the U.S. to become the world's largest oil producer in 2019.[2]

In addition, the negative environmental externalities associated with fossil fuel extraction have increasingly become a target for policy intervention. The GHG emissions associated with the ultimate consumption of extracted fuels have of course garnered a great deal of attention, but on top of that fossil fuel industries also generate both local and GHG emissions during the fuels' extraction and transportation. Industrial organization has a great deal to contribute to our understanding of the efficacy of regulations that target these externalities.

Beyond the oil and gas sector's substantive importance, industrial organization economists have been drawn to it because the wealth of available data—especially in the U.S.—make it an outstanding setting to study IO questions about firms' investment decisions and the economics of contingent contracts between asset sellers and well-informed buyers. Every oil and gas well in the U.S., whether drilled on public or private land, has a publicly-recorded drilling permit and completion report associated with it that describes who operates it, precisely where it was drilled and to what depth, when drilling began, and when the well was completed. In many cases, researchers can also observe drilling inputs and the identities of contractors who worked on the well. Moreover, wells' monthly production is also observable because it, too, is recorded by state agencies. The U.S. oil and gas industry therefore offers a unique opportunity to observe a large number of discrete investments and the return on those investments. For instance, in 2019 approximately 17,000 oil and gas wells were drilled in the U.S. by hundreds if not thousands of firms.[3] This granularity of investment data is difficult to find in almost any other setting. On top of that, a wealth of data are available regarding leases between mineral owners and extraction firms. For U.S. state and federal minerals, these leases are typically auctioned, and available data often include information on both winning and losing bids. Private mineral leases are instead typically negotiated, but because these leases must be publicly recorded, data are usually available about major lease terms such as the royalty and primary term (though not the bonus). Together, these data enable researchers to answer questions about how firms compete to acquire development contracts, and about how alternative allocation mechanisms affect the owner's revenues, resource development, and total surplus.

While this section of the chapter is titled "Extraction of energy resources", it will follow the recent literature by focusing the vast majority of its attention on the oil and gas sector. The subsection on environmental regulation will, however, also discuss coal in the context of downstream emissions regulations and exercise of market power by railroads. Industrial organization research on coal extraction itself has unfortunately

---

[1] The 95 million bbl/d figure is from BP (2020), page 16. Valuations are based on the 2019 average Brent crude price of $64.30/bbl, from the EIA at `https://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm`.

[2] See `https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPUS2&f=M` for U.S. oil production data.

[3] Per Patel and Geary (2020), approximately 1,400 new oil and gas wells were drilled per month in the U.S. in 2019.

lagged behind that studying the oil and gas sector, despite coal's large role in the GHG emissions inventory and despite a wealth of available data on U.S. coal extraction from agencies such as the U.S. Geological Survey, Energy Information Administration (EIA), and the Mine Safety and Health Administration.

In terms of methods, tools for simulating and estimating models of dynamic behavior feature prominently in the research we discuss in this section. IO research on resource extraction also draws heavily on tools for simulation and estimation of auctions, as we discuss in subsection 2.2, and productivity estimation, as we discuss in subsection 2.4. In general, our discussion of research in this area will emphasize papers' substantive contributions rather than describe the methods used in detail, in part because the methods themselves are discussed extensively in other chapters in this volume.

2.1 Dynamics of oil production and drilling timing

Since Hotelling (1931), economists have recognized that the question of how to optimally produce natural resources out of an initial stock is inherently a dynamic problem. In most cases, the relevant question is not whether or not to extract the reserves at all, but instead how much to extract today and how much to leave for tomorrow. Hotelling's famous 1931 article presents a prescriptive model of the optimal extraction path (or, equivalently, the extraction path that would prevail in competitive equilibrium with forward-looking extractors), culminating in what has become known as Hotelling's Rule that the difference between the resource's price and its marginal extraction cost should rise at the rate of interest. In this sub-section, we begin by discussing the core intuition and model that underlies this result. We then highlight the empirical shortcomings of overly-simple Hotelling-style models when applied to oil and gas extraction, and discuss a series of recent papers in industrial organization that augment the basic version of the Hotelling model with important real-world features of firms' extraction problems: capacity constraints on production from drilled wells, exercise of market power by Organization of Petroleum Exporting Country (OPEC) producers, and uncertainty about future oil and gas prices. These model augmentations yield vast improvements in the ability of Hotelling-style models to explain observed market behavior, raising the prospects that such models can be used to help evaluate the likely consequences of policies affecting the oil and gas industry.

2.1.1 The "standard" Hotelling model

In the standard Hotelling (1931) style model, as commonly taught in graduate resource economics courses, the central planner (or equivalently, a price-taking, resource-owning firm) is assessing how to optimally extract an exhaustible resource, given static and deterministic functions for flow utility and extraction costs. Consider a continuous time version of such a model, letting $y_t \geq 0$ denote the extraction rate at time t, and letting $x_t \geq 0$ denote the remaining stock at t. Utility is given by $u(y_t)$, and the extraction cost is $c(y_t)$, such that the difference $u(y_t) - c(y_t)$ is strictly increasing at $y_t = 0$ and strictly concave. The initial stock $x_0 = S > 0$.

Given a discount rate $r > 0$, the planner's problem is given by:

$$\max_{\{y_t\}} \int_0^\infty (u(y_t) - c(y_t))\, e^{-rt} dt, \quad \text{s.t. } \dot{x}_t = -y_t,\, y_t \geq 0,\, x_t \geq 0,$$

where $\dot{x}_t$ denotes the rate of change of the stock x_t. Letting μ_t denote the current-time shadow value of the resource stock, the current-value Hamiltonian and first-order conditions (FOCs) are then:

$$H = u(y_t) - c(y_t) - \mu_t y_t \tag{1}$$

$$\text{FOC}_{y_t}: u'(y_t) - c'(y_t) - \mu_t \leq 0, \quad y_t \geq 0, \quad y_t \left(u'(y_t) - c'(y_t) - \mu_t\right) = 0 \tag{2}$$

$$\text{FOC}_{x_t}: \dot{\mu}_t = r\mu_t \tag{3}$$

$$\text{TVC}: \lim_{t\to\infty} x_t \geq 0, \quad \lim_{t\to\infty} \mu_t e^{-rt} \geq 0, \quad \lim_{t\to\infty} x_t \mu_t e^{-rt} = 0. \tag{4}$$

FOC_{y_t} and FOC_{x_t} together tell us that on the optimal extraction path, the difference between marginal utility (i.e. price) and marginal cost must be rising at the interest rate r so long as the extraction rate is strictly positive. This result is commonly known as the Hotelling Rule. In a competitive equilibrium setting, the rule relates to an intuitive indifference condition. In order to spread out extraction continuously over time, extraction firms must be indifferent about extracting their resource across time. This indifference condition can only hold if the marginal return to extraction is increasing over time at the interest rate, thereby holding the present value constant.

The transversality condition (TVC) then requires that either the resource be completely exhausted in the limit as $t \to \infty$, or alternatively that the resource be worthless on the margin in the limit. In most practical applications with an infinite horizon it is the former condition that will hold, and the TVC then puts a boundary on the problem and determines the level of the initial extraction rate.

The standard Hotelling Rule prediction is strong and has been shown to be at odds with most natural resource price data (and, in some studies, implied data on in-situ shadow values). See reviews of these studies in Krautkraemer (1998) and Slade and Thille (2009). In the context of crude oil prices, figure 1 shows that the time series of U.S. price of crude oil exhibits no strong upward trend over the long run.

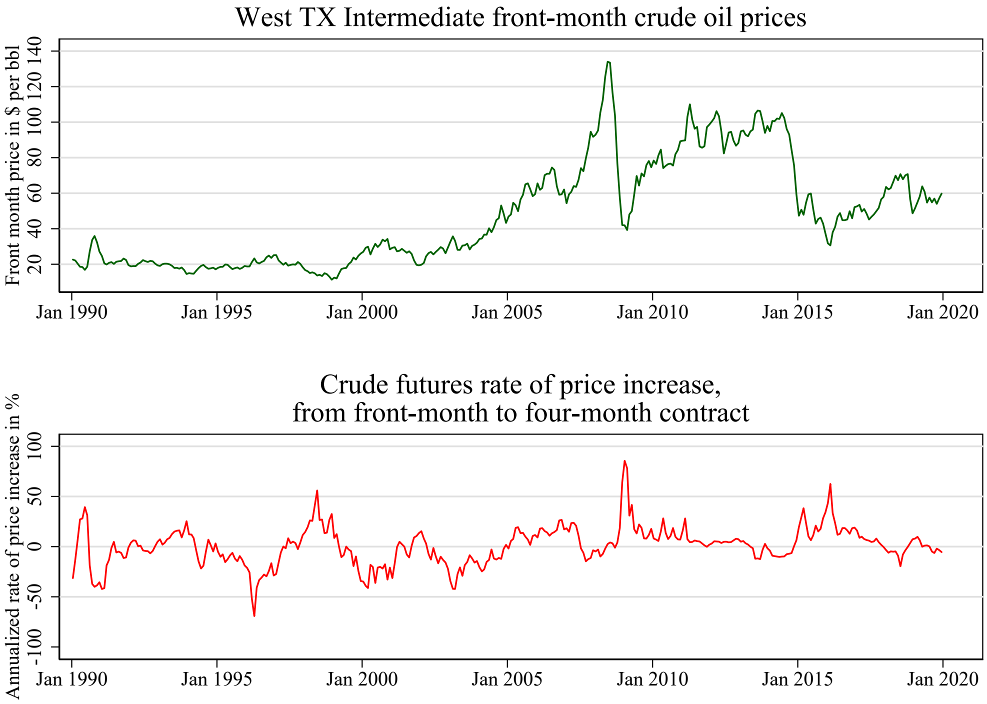
One clear gap between the simple Hotelling model presented above and real-world data is that the model is fully deterministic, whereas real-world demand ($u'(y_t)$) and marginal extraction costs ($c'(y_t)$) of crude oil are stochastic. Incorporating demand and cost shocks into the simple Hotelling model would naturally lead the model to predict price volatility (particularly if demand and marginal cost are inelastic in y_t), but Hotelling's Rule would still hold in expectation. That is, the model would continue to predict that rational agents at any time t should forecast that the difference between the expected future price and marginal extraction cost of oil will increase at the interest rate r.

The bottom panel of figure 1 shows that even the expectations version of Hotelling's Rule fails to hold in crude oil markets. Using futures prices as a proxy for expected futures prices,[4] this panel shows the evolution over time of the percentage difference between the 12-month oil futures price and the front-month futures price. A positive value indicates that the market is *in contango* and that the price is expected to increase, whereas a negative value indicates that the market is *backwardated* and that prices are expected to

---

[4] As explained in Pindyck (2001), futures prices will not be equivalent to expected future prices if the risk-adjusted discount rate that applies to investments in crude oil stocks differs from the risk-free rate (e.g., if the CAPM beta for crude oil differs from zero). Even if futures prices are systematically biased away from expected future prices for this reason, the variation in futures curves between backwardation and contango is not consistent with Hotelling. Also, Anderson et al. (2018) estimates a CAPM beta of nearly zero for West Texas Intermediate crude oil using data from April 1983 through April 2015.

Figure 1: Crude oil front-month prices and futures curve growth rates


West TX Intermediate front-month crude oil prices

Crude futures rate of price increase, from front-month to four-month contract

Note: Futures prices shown are NYMEX West Texas Intermediate crude for delivery to Cushing, Oklahoma. Data source: `https://www.eia.gov/dnav/pet/pet_pri_fut_s1_d.htm`. The front-month price is for delivery in the upcoming month. We compute the annualized rate of price increase as the difference between the futures price for delivery four months out and the front-month price, divided by the front month price, times 4 (and converted to percent by multiplying by 100).

fall. The long periods in which oil futures are backwardated, the fluctuations between backwardation and contango, and the periods of extreme contango when the oil price is expected to rise more quickly than any plausible discount rate are inconsistent with a Hotelling-style expectation of steadily increasing resource prices.

2.1.2 Augmenting Hotelling by separating drilling and production

As noted in Slade and Thille (2009), the theoretical and empirical literature on Hotelling-style models had largely fizzled by the year 2000. Even though there existed some theoretical work such as Pindyck (1978) that could yield initial periods of decreasing price expectations by requiring initial investments in reserves, the empirical literature that focused on explaining price time series struggled to fit these models to the data.

Anderson et al. (2018) attempts to address the shortcomings of Hotelling-style models—while retaining the core dynamic optimization intuition—in the context of oil and gas by focusing attention on micro-level development and extraction decisions rather than macro-level price time series. The paper begins by examining lease-level administrative oil production data from Texas. Using data from 16,159 oil leases that did not experience any new well investment after 1990, the paper finds that production from established, drilled wells is almost completely unresponsive to oil price shocks and instead smoothly falls over time along a roughly exponential decline curve. This finding prevails despite large oil price swings during the paper's 1990–2007 sample and a period in 1998–1999 when 12-month futures prices exceeded the front-month price by more than 20%. This production behavior violates the predictions of standard Hotelling-style models.

In contrast, when Anderson et al. (2018) studies data on drilling of new wells, it finds a strong response of drilling to oil price shocks (the estimated elasticity of drilling to the oil price is 0.73). In addition, the marginal cost of drilling, measured using data on the rental rate for drilling rigs, positively co-varies with oil prices. Taken together with the initial finding that oil production from drilled wells is *not* responsive to prices, these empirical results signal a need to reformulate Hotelling models in a way that treats production decisions differently from drilling decisions.

The core innovation that Anderson et al. (2018) uses to explain these results is to follow the petroleum geology engineering literature by modeling the production rate from drilled wells as subject to a capacity constraint that is proportional to the remaining reserves available to the well. The idea is that the motive force that propels oil through the rock, to the wellbore, and up the wellbore is the underground pressure of the reservoir fluids, and as oil is produced that pressure declines.[5] Anderson et al. (2018) then assumes that the marginal production cost is zero for production rates that are less than this hard constraint. Together, this additional structure on the extraction problem can deliver the empirical result that firms' production from existing wells follows a decline curve and is not responsive to price shocks.

To formally show this result, Anderson et al. (2018) models the choice of production rate F_t, conditional on some initial stock of drilled wells with production capacity K_0, as the problem of maximizing $\int_0^\infty U(F_t)e^{-rt}dt$, subject to the constraints $F_t \in [0, K_t]$ and $\dot{K}_t = -\lambda F_t$, where λ denotes the exponential production decline rate whenever $F_t = K_t$. Letting θ_t denote the current shadow value of capacity K_t, the FOC governing the optimal production rate is then that the capacity constraint binds when $U'(F_t) > \lambda\theta_t$. That is, it will be optimal for firms to produce at full capacity if $U'(F_t)$—the oil price—exceeds the value of the capacity that is reduced in the course of production (producing at a rate F_t reduces capacity by λF_t, and that capacity is valued at θ_t dollars per unit).

A testable implication of this production model is that, during the 1990–2007 time period studied in Anderson et al. (2018), the $U'(F_t) > \lambda\theta_t$ condition should hold, since firms appear to have set $F_t = K_t$ throughout this time. To conduct this test, the paper estimates $\lambda = 0.1$ based on the observed production decline rate and estimates θ_t using a second implication of the model: the value θ_t of a marginal unit of capacity should equal the expected future stream of oil prices, discounted at $\lambda + r$. Anderson et al. (2018) then shows empirically that, under a variety of reasonable assumptions on the appropriate discount rate, the $U'(F_t) > \lambda\theta_t$ condition holds throughout the sample.

The upshot is then that, barring extreme periods in which the price of oil is forecast to increase rapidly for a long period of time, the problem of optimizing oil production from previously-drilled wells is actually rather boring. Price-taking firms should produce wells at their geologic capacity constraint, and production

---

[5] For an accessible primer on petroleum geology and decline curves, see Hyne (2001).

will then naturally decline over time. And micro-data indicate that this is exactly what U.S. oil producers do. Moreover, this empirical result has been replicated for modern shale oil and gas wells by Newell and Prest (2019) and Newell et al. (2019).

Drilling decisions, in contrast, are where the action is and where the dynamic Hotelling-style incentives lie. In Anderson et al. (2018), firms drill new wells to increase their production capacity, facing a marginal drilling cost that is strictly increasing in the rate of drilling, and facing a finite stock of wells that they can drill. Denoting the drilling rate by a_t, the initial capacity of a new well by X, and marginal drilling cost by $d(a_t)$, the return to drilling a marginal well is given by $\theta_t X - d(a_t)$. Anderson et al. (2018) shows that, because the stock of wells to drill is a scarce resource, this marginal value must increase over time at the discount rate r. Thus, the Hotelling logic applies to drilling timing decisions, rather than to production decisions once drilling has taken place.

While the model in Anderson et al. (2018) is deterministic, the paper closes by simulating drilling, production, and price outcomes when the equilibrium path is subjected to one-time, unanticipated demand shocks. The simulated responses to these shocks can result in price paths that, if anticipated following the shocks' impact, would naturally lead to switches between backwardation and contango in futures markets. Namely, following a positive demand shock, the oil price must increase on impact because production cannot. Firms will then respond by increasing their rate of drilling, consistent with observed drilling data. This increased drilling will eventually increase the production rate, causing the price of oil to gradually fall after the initial impact of the shock. If agents anticipate this behavior, oil futures markets should be backwardated immediately following the shock, which is consistent with the backwardation observed in historical futures price data.

Anderson et al. (2018) therefore makes two high-level contributions. First, it shows that a Hotelling-style model actually can provide useful empirical predictions for the oil and gas extraction industry, provided that the model incorporates the industry's essential institutional features and constraints. Second, it provides a theoretical and empirical foundation for why economists studying the oil and gas industry should focus their attention on drilling investments rather than on production decisions once wells have been drilled. Many of the papers we discuss below that examine mineral leasing policies, productivity, and firms' strategic behavior adopt this approach.

2.1.3 Real options: drilling in the presence of stochastic oil prices

Crude oil markets are of course stochastic, rather than deterministic as modeled in Anderson et al. (2018). Because drilling an oil well is a classic example of an irreversible investment—there is no way to "un-drill" a well and recover the investment cost—firms holding drillable acreage should account for this volatility when deciding when to drill their wells. Specifically, real options theory (Dixit and Pindyck, 1994) highlights that high anticipated future oil price volatility should, all else equal, increase the value of holding investment for the future, making drilling today less attractive. The intuition for this result does not come from risk aversion, but rather the fact that high volatility increases the potential upside from sinking the investment in the future, while the downside of waiting is bounded by the fact that the firm can always choose not to drill if the oil price turns out to be low (achieving profits of zero rather than strictly negative profits).

Kellogg (2014) examines whether the drilling decisions of oil and gas companies actually comport with the prescriptions of real options theory. The paper aims to test not only whether firms reduce their rate of drilling when expected oil price volatility is high, but whether the magnitude of this effect matches what is implied by dynamic optimization. The paper ultimately answers this question in the affirmative, helping to underscore the usefulness of dynamic models for explaining firms' investment behavior.

To execute its empirical test, Kellogg (2014) takes advantage of two helpful features of the oil and gas setting. First, the presence of both oil futures markets and oil futures options markets mean that it is possible to obtain independent measures of the market's expected future oil price and expected future price volatility (the latter of which can be inverted out from options prices, using a variant of the Black and Scholes (1973) financial options pricing model). Second, the paper takes advantage of the fact that U.S. oil producers behave competitively in the output market, so that firms' investment problem can be modeled as a single-agent dynamic program rather than as a more challenging dynamic game.[6]

In a simplified version of the model employed in Kellogg (2014), the optimal drilling program should be given by the solution to the following Bellman equation:

$$V(P,\sigma) = \max\{\pi(P), \delta E[V(P',\sigma')]\}, \tag{5}$$

where P and σ are the state variables denoting the current price of oil and the firm's belief about oil price volatility, respectively. $\pi(P)$ denotes the current-period payoff to drilling, and δ is the discount factor. The oil price P directly affects the profits from drilling, but volatility σ does not. Instead, σ affects the variance of the distribution of oil prices that the firm believes it will face next period.

Kellogg (2014) then tests whether firms' beliefs about oil price volatility, as implied by their drilling behavior, line up with actual oil price volatility data from the futures options market. The paper models firms' volatility belief at a given time t as a linear combination of the sample average logged volatility, $\overline{\log\sigma}$ and the difference $\log\sigma_t^d$ between the actual volatility at t and this average. That is:

$$\log\sigma_t = \overline{\log\sigma} + \beta\log\sigma_t^d. \tag{6}$$

If $\beta = 1$, then firms' beliefs (as implied by their drilling behavior) align with the market. In contrast, the value $\beta = 0$ would imply that firms behave as if oil price volatility does not vary over time.

Kellogg (2014) estimates β using monthly data on drilling, oil prices, and implied price volatility from 1993–2003. Both the price and price volatility vary substantially during this time period, with volatility experiencing notable increases during the 1998–1999 oil price crash and following the 9/11 attacks. To enable the model to rationalize the observed drilling data, the final empirical model also includes a time-varying unobservable that represents firms' beliefs about the quantity of oil that will be produced from each well.

Kellogg (2014) finds that firms respond to volatility shocks in a way that closely matches the theory: the estimated value of β is 1.12. The paper then surmises that firms may behave this way because there is sub-

---

[6] To further ensure that a single-agent model is appropriate, Kellogg (2014) limits its sample to established Texas oil fields that are operated by a single firm, thereby avoiding issues of information externalities (we discuss papers tackling these issues in subsection 2.3).

stantial value at stake in getting drilling decisions right. For instance, the paper constructs an example using in-sample variation in volatility, showing that recognizing that volatility is relatively high—and therefore delaying drilling—can increase a drilling opportunity's value by 27%.

2.1.4 OPEC, market power, and drilling sequencing

The model in Anderson et al. (2018) assumes that oil and gas firms act as price-takers. This assumption is reasonable for the many privately-owned firms that produce oil and gas, but it is likely problematic for the large oil-producing nations with nationally-owned oil companies that participate in the Organization of Petroleum Exporting Countries (OPEC) cartel. The market power possessed by this large cartel, or even unilaterally by a very large oil producer like Saudi Arabia, might plausibly lead to distortions in crude oil extraction.

In a competitive market, low-cost resources should be produced before high-cost resources, as shown by Herfindahl (1967) using a Hotelling-style model in which different deposits produce perfectly substitutable products but at different marginal costs. But if the low-cost producers possess market power, they may find it profitable to delay extraction, leading to high-cost resources being produced simultaneously with or even after low-cost resources. This out-of-merit production ordering is a form of allocative inefficiency.

Asker et al. (2019) attempts to quantify the extent of this inefficiency using data from the global oil market from 1970 to 2014, during which time OPEC held a roughly 40% share of world oil production. Because well-level data are not available at a global scale, Asker et al. (2019) uses proprietary oil field-level data on production and extraction costs from Rystad Energy, an energy consultancy. A necessary limitation of these data is that it is not possible to separately model the costs of new drilling investment versus production from existing wells and facilities.

Asker et al. (2019) uses the Rystad cost data to construct a counterfactual, surplus-maximizing global oil extraction plan in which extraction proceeds in strict order from the lowest-cost to highest-cost fields. The key assumption the paper makes to enable this approach is that marginal costs are constant within-field (else, it may be optimal to produce from multiple fields simultaneously). The paper can then quantify the extent of mis-allocation by comparing costs under the observed extraction order to costs under this counterfactual, holding constant the total amount of extraction each year.

Because the Rystad data indicate that production costs in many of the major OPEC countries are substantially lower than costs outside of OPEC, Asker et al. (2019) finds that the costs of OPEC-driven mis-allocation are substantial. Asker et al. (2019) values OPEC-driven mis-allocation at \$105 to \$163 billion during 1970–2014.[7] The paper therefore highlights how OPEC's exercise of market power can lead to large allocative efficiency losses by distorting extraction timing relative to that implied by a Hotelling-style benchmark model in which firms behave competitively. Note that because the paper's counterfactuals hold total global extraction constant, this welfare loss is in addition to the standard welfare loss from market power associated with any decrease in the total volume of oil produced each year (though such a loss may be substantially offset by the environmental damages associated with oil production and consumption).

---

[7] These values are deflated to 2014 dollars.

Asker et al. (2019)'s estimated OPEC-driven distortion omits distortions related to mis-allocation within OPEC itself as well as distortions within the set of non-OPEC countries. The paper's estimate of the overall surplus distortion that includes these other forms of mis-allocation (including within-country mis-allocation) is $744 billion. However, some of this figure may be due to measurement error in the Rystad data or to expectational errors that would occur even in the absence of any market failure.[8] Nonetheless, Asker et al. (2019) implies that a variety of frictions—including but not limited to market power—may impede oil and gas development from proceeding in a fully efficient manner. Some of the papers we discuss below build on this result by pointing to specific frictions that inhibit efficient lease allocation or rapid productivity improvements, thereby diminishing the sector's overall efficiency.

Finally, we note that the results in Asker et al. (2019) also have rather negative implications for climate policies aimed at reducing global oil consumption. The OPEC-driven mis-allocation in Asker et al. (2019) arises from the result that, for many large OPEC nations, the marginal cost of producing additional crude oil is substantially less than recent oil prices. This result will make it harder for alternative fuels to effectively compete with oil, since displacing oil at scale will ultimately require out-competing oil at cost, not at current prices. Quantifying the climate policy implications of the gap between oil prices and the low marginal extraction costs in OPEC nations therefore strikes us as a valuable contribution for future research.

2.2 Oil and gas mineral leasing

The owners of underground mineral resources—whether they be sovereign governments, state governments, or private individuals—typically do not themselves possess the necessary expertise or capital required to carry out extraction in a cost-effective manner (or at all).[9] To realize value from their resources, mineral owners instead lease them to private extraction firms. These leases play a central role in determining when particular resources are developed, which firm develops them, and how resource rents are shared.

U.S. federal oil and gas leases—and in particular lease auctions in the Outer Continental Shelf (OCS) of the Gulf of Mexico—have historically received a great deal of attention from IO economists. An earlier volume of this handbook (Hendricks and Porter, 2007), as well reviews in Porter (1995) and Haile et al. (2010), summarizes the catalog of research on bonus bidding in U.S. federal OCS auctions through 2010. Three important lessons from this body of work are that: (1) exploratory ("wildcat") lease sales in the OCS exhibited large bid dispersion, substantial competition, and the government capturing a large share of the tracts' value; (2) sales of "drainage" tracts located adjacent to producing tracts were less competitive, since the neighboring lessees could use their information advantage to capture rents; and (3) firms submit bids consistent with their recognition of the "winners' curse" in common value settings. We refer readers interested in this literature to the excellent reviews highlighted above.

In this sub-section, we will discuss recent work in industrial organization that extends the literature above by expanding our understanding of bidding firms' valuations and informational environment, and of

---

[8] Expectational errors are out-of-order extraction that occurs because ex-post costs turned out to differ from costs that extractors expected to incur prior to investing in a field.

[9] Countries with national oil companies are an exception. However, because little data are available regarding the operations of such companies, the IO literature has instead overwhelmingly focused on settings—especially the United States—where governments or private owners lease their interests to extraction firms.

aspects of oil and gas leasing beyond the up-front bonus payments. This research has leveraged new datasets from states' oil and gas lease auctions, which exhibit experimentation in auction formats that is largely absent from federal OCS auctions, and from private oil and gas leasing, which takes place via decentralized negotiations rather than formal auctions.

2.2.1 Royalties and primary terms

We begin by discussing recent work oil and gas leases' royalty clauses, which dictate that the lessee firm must pay a royalty share of its oil and gas revenue to the mineral owner. While royalty clauses often allow for some deductions related to transportation and processing costs, they do not allow the firm to take deductions for the major, up-front costs of drilling and completing wells.[10] Thus, the royalty effectively acts as a tax on revenue, distorting the firm's incentives relative to the case where it is the residual claimant on all drilling and extraction activity. In particular, the royalty will induce the firm to delay drilling (or perhaps not drill at all) and may potentially also induce the firm to reduce other inputs—such as the intensity of a hydraulic fracturing treatment if one is being applied—conditional on drilling and completing a well.

So why then do leases include royalties at all? The answer is provided by theory developed in Laffont and Tirole (1986), Riley (1988), and Hendricks et al. (1993). If there are only a handful of firms who are interested in leasing the parcel, and if those firms have differing valuations for the parcel, then selling the lease for only the bonus bid, without a royalty, will necessarily leave the winning firm with information rents. One way to reduce these information rents is to link the payments made by the winning firm to the realized value from the tract (Milgrom and Weber, 1982). The royalty clause in an oil and gas lease does precisely this. However, including the royalty comes at the cost of distorting the firm's extraction effort.

These effects can be illustrated in a simple, stylized model. Suppose there is only one potential lessee firm, and the mineral owner can make the lessee a take-it-or-leave-it offer that is a combination of an upfront payment p and royalty r. The lessee's profits from the lease, should it accept the offer, are determined by:

- v: the firm's private value on the reserves. The mineral owner believes v is distributed $F(v)$, with support on $[v_L, v_H]$, where $v_L > 0$.
- E: the extraction "effort" undertaken by the firm. In practice, effort will take the form of drilling the well sooner rather than later, or increasing the intensity of hydraulic fracturing. The empirical papers that we discuss below will model these forms of effort explicitly, but for now we simply consider effort as measured by the scalar $E \geq 0$ and abstract away from any dynamics.

The firm's profits π are then given by

$$\pi(E|v, p, r) = (1 - r)vE^{\alpha} - CE - p, \tag{7}$$

where $\alpha \in (0, 1)$ and $C > 0$. Conditional on accepting the lease, and on $r < 1$, the optimal effort E^* is then

---

[10] Allowing cost deductions—effectively making royalties more like profit taxes—would improve the efficiency of these contracts. In practice, however, accurate cost reporting can be difficult to monitor, audit, and enforce, especially when the lessee firm operates leases across many different owners, and when the owners are private individuals with limited expertise and resources.

strictly greater than zero, and it is a strictly decreasing function of the royalty r. The firm will then accept the lease offer if $\pi(E^*|v,p,r) \geq 0$.

Let $\underline{v}(p,r)$ denote the lowest participating firm type given the bonus p and royalty r. The expected value V_o of the lease to the mineral owner is given by:

$$V_o(p,r) = \int_{\underline{v}(p,r)}^{v_H} (p + rvE^{\alpha})dv \tag{8}$$

To see that the optimal royalty is non-zero, we can differentiate equation (8) with respect to r and evaluate it at $r = 0$. Doing so yields:

$$\frac{dV_o(p,0)}{dr} = \int_{\underline{v}(p,0)}^{v_H} vE^{\alpha}dv, \tag{9}$$

which is strictly greater than zero. From the mineral owner's perspective, the revenue gains from increasing the royalty rate (starting from 0%), and thereby reducing the firm's information rents, are first-order. The cost of the induced effort distortion, however, is second-order. Thus, the revenue-maximizing royalty rate for the owner will generally lie strictly between 0% and 100%.

Two questions then follow from this intuition. First, in typical oil and gas lease auctions that fix the royalty and ask firms to submit bonus bids, what is the revenue-maximizing royalty rate to set? And second, what are the benefits or drawbacks of alternative auction formats that ask bidders to submit royalty bids (either with a fixed bonus or with a bonus that is also a bid variable)? We next discuss a set of recent papers that have been the first to address these questions empirically.

Optimal royalties in fixed-royalty bonus bid auctions

In the commonly-used fixed-royalty bonus bid auction format, the revenue-maximizing royalty rate depends on the amount of private information possessed by firms (which is competed away as the number of bidding firms increases) and the extent to which the royalty distorts drilling activity after the lease is awarded. Bhattacharya et al. (2020) and Ordin (2019) address this question using data from lease auctions and drilling activity on state-owned mineral parcels in New Mexico. The workhorse of both papers is a structural model that links: (1) a lease auction in which firms with heterogeneous valuations bid on tracts, taking the royalty rate as given; and (2) a model of the winning firm's drilling timing problem. The papers' goal is to use this combined model to simulate counterfactuals with different royalty rates and assess how royalties affect bidding, the state's revenues, and drilling activity.

In Bhattacharya et al. (2020)'s and Ordin (2019)'s model, the firm's choice of when to drill (or to not drill at all) is the analog to the effort E in the simple model presented above. This drilling timing problem is represented in the model as a single-agent, finite-horizon dynamic optimization problem, via a Bellman equation. A simplified version of this problem is given by:

$$V(p,q,c,r) = \max\{(1-r)pq - c, \delta E[V(p',q,c,r)]\}, \tag{10}$$

where p denotes the oil price (which is stochastic), q denotes the firm's belief about quantity, c is the drilling cost, and r is the royalty rate. The solution to this problem not only gives drilling probabilities each period as a function of output prices and lease terms, but also firms' valuation of each tract at the time of the auction. The model then uses these valuations in the first-stage auction game. This auction is modeled a a standard common value auction, given the solution to the Bellman equation (10) and a parameter governing firms' uncertainty about the tract's quantity q at the time of bidding.

Bhattacharya et al. (2020)'s and Ordin (2019)'s estimation and simulation of the combined auction and drilling timing model are themselves a methodological innovation, since to the best of our knowledge previous work has not combined these models into a single package. This approach is likely to be useful not just for auctions of natural resource extraction rights, but also for other settings in which a principal grants an agent the right to develop or execute a project, the realized value is ex-post contractible, and the agent has some discretion as to the quantity or quality of the work. Both papers use data on firms' bonus bids (New Mexico uses sealed-bid auctions and makes information on winning and losing bids available), drilling timing, oil production, oil prices, and royalties (which vary across tracts) to jointly estimate the parameters of the model in a single step by matching simulated moments to the data.[11]

Unlike many auction papers, the distribution of bidders' values at the time of the auction is not itself a primitive to be estimated, since these values are the solution to the Bellman equation (10). Instead, the papers must estimate the distribution of actual tract quantities, along with the noise in bidders' signals of tract quantities and firms' drilling costs. Obtaining the distribution of quantities conditional on drilling is straightforward because this object is observed directly in the data. The unconditional distribution is identified using information on firms' drilling timing decisions (or decision not to drill at all), where the oil price p is used as an excluded variable: it is a determinant of drilling time that does not affect output.

The estimation method used in Bhattacharya et al. (2020) proceeds roughly as follows. Given a set of parameters, the model can be simulated via backwards induction. That is, the paper first solves the drilling timing problem given by equation (10) for a range of quantities q. Then, given a candidate distribution of q, the paper repeatedly draws firms' signals of q and simulates common-value auctions (and then the drilling decisions that follow). Estimation of the parameters and the distribution of q then proceeds by matching simulated moments to actual moments in the data. These moments include the highest, second-highest, and average bid; moments of the quantity distribution; moments for drilling delays; and interactions.

Both Bhattacharya et al. (2020) and Ordin (2019) use the estimated model to find that the revenue-maximizing royalty rate for the state of New Mexico is 29%, which is modestly higher than the royalty rate the state uses in practice. Ordin (2019) emphasizes that the revenue-generating property of the royalty comes at the cost of significantly reducing drilling: it estimates that eliminating the royalty entirely would increase the unconditional probability that a tract is drilled from 0.096 to 0.154 (a 60% increase).

Taken together, these papers therefore provide quantitative guidance for how changing royalty policies will influence the government's revenues and drilling activity. Providing results of this nature is not possible in the absence of a model that links firms' behavior in the leasing auction to their ex-post drilling actions, and these papers are the first to provide an econometric approach to doing so.

---

[11] Following other work using optimal stopping problems to model drilling, such as Kellogg (2014), Bhattacharya et al. (2020) and Ordin (2019) estimate the oil price transition process in a preliminary step.

Auctions with royalty bidding

Another advantage of the model that Bhattacharya et al. (2020) and Ordin (2019) develop is that they can use it to simulate counterfactual policies other than the standard bonus bid auctions that New Mexico (along with other states and the U.S. federal government) uses. These simulations are motivated by theoretical work in DeMarzo et al. (2005) and Skrzypacz (2013) on securities auctions that link the bidders' total payments to cash flows that are realized in the future. One alternative security auction form that could be used in oil and gas leasing is an "equity" auction, in which firms would bid in a royalty rate rather than a bonus. DeMarzo et al. (2005) shows that in the absence of moral hazard, the equity auction is likely to maximize the seller's revenue. Skrzypacz (2013), however, emphasizes that equity auctions can decrease both the seller's revenue and total surplus if the high rate of revenue sharing significantly distorts the agent's actions after the auction.

Bhattacharya et al. (2020) and Ordin (2019) are two of the first papers to quantify the impacts of these alternative auction forms using a model estimated from real-world data. Both papers find that equity auctions would perform terribly in New Mexico oil and gas leasing. In the simulated counterfactuals, firms bid high royalty rates that can exceed 40%, ultimately resulting in little drilling taking place and therefore little revenue flowing to the government. Kong et al. (2020) finds similar results using data from state-run auctions in Louisiana. There, firms bid in both a bonus and a royalty rate, and Kong et al. (2020) finds that a simpler bonus bidding mechanism with a fixed royalty would increase the state's revenue, the likelihood that tracts are allocated to the firm with the highest valuation, and the likelihood that tracts are drilled (Kong et al. (2020)'s model involves private rather than common valuations).

Primary terms

In addition to bonus bids and royalties, oil and gas leases also include primary terms that limit the number of years the lessee has to drill and complete at least one productive well. If the firm does not drill by the end of the primary term, it must relinquish the lease, and the mineral owner is then free, if desired, to re-lease the parcel (either to the same firm or to a different one). This deadline can substantially alter firms' drilling incentives. For instance, in Bhattacharya et al. (2020)'s and Ordin (2019)'s simulations, firms' likelihood of drilling, conditional on not yet having drilled, increases sharply as the deadline approaches.

Herrnstadt et al. (2020) studies the impacts of primary terms in the context of private natural gas leasing in the Haynesville Shale in northwest Louisiana. One feature of this environment is that private parcels are typically too small on their own to support drilling. The state therefore prescribes a process by which leases may be force-pooled into square-mile drilling units, and that drilling a productive well anywhere in the unit holds all the leases in the unit. Importantly, the operating firm may then drill subsequent, follow-up wells elsewhere in the unit. The incentive to preserve the option for future drilling is strong, and Herrnstadt et al. (2020) begins by showing that in Haynesville drilling units there is clear bunching of drilling timing into the months just prior to the first lease expiration date in the unit.

While the bunching of drilling timing may seem distortionary ex-post, Herrnstadt et al. (2020) next illustrates why primary terms may actually enhance both efficiency and the owner's revenues ex-ante. Starting with an analytical model that builds on Laffont and Tirole (1986) and Board (2007), Herrnstadt et al. (2020) shows that if the owner induces the firm to accelerate drilling by including a drilling subsidy in the lease contract, expected revenue can increase because the subsidy counteracts the delay incentive generated by the royalty. Although an explicit subsidy may often be impractical due to liquidity constraints or other factors, primary terms can fulfill the same objective, albeit in a coarse way (similar to "notched" tax policies in the public finance literature (Kleven, 2016)). This intuition, while most directly applicable to oil and gas leases, may also be relevant to other settings, such as master franchise contracts or licenses to adapt creative works, in which principals sell time-limited development options to agents.

Herrnstadt et al. (2020) then develops and estimates a structural model to explore how revenues, total surplus, and drilling are affected by primary terms. The core of the model, like in Bhattacharya et al. (2020) and Ordin (2019), is the firm's optimal stopping problem, which has a finite horizon when there is a primary term or an infinite horizon if not. Herrnstadt et al. (2020) adds to this framework the feature that, in the event the primary term expires, the owner and firm can sign a new lease, requiring another bonus payment.

Consistent with the paper's analytical results, Herrnstadt et al. (2020) finds that primary terms can increase the owner's expected revenue by 1%–3%, relative to a contract with just a royalty and no primary term deadline. Total surplus also increases because drilling is pulled forward in time, counteracting the delay induced by the royalty.

Herrnstadt et al. (2020) also finds, however, that primary terms are likely to reduce the owner's revenue when the first well drilled in a unit creates an indefinite option to drill additional wells in the future. In that situation, owners are better off if they can impose primary terms on all potential wells that might be drilled, not just the first one. Regulations in Louisiana proscribe such contracts, but these results relate to the increasing use by other states' private mineral owners of "retained acreage" clauses that allow firms to hold only lease acreage that is proximate to a drilled well, and not acreage that might be used for future wells.

2.2.2 Information, common value auctions, and royalties

The leasing of wildcat (exploratory) tracts is a classic common-values auction setting. In a common-values environment, the bids of losing firms help inform the winning firm's beliefs about how productive the tract will be. These beliefs in turn affect the winning firm's likelihood of drilling. An important question for the mineral owner is then whether to release the values of losing bids to the winner. Doing so might increase the winner's probability of drilling—and therefore also the owner's royalty revenue—if the losing bids are relatively high (i.e. not too far below the winning bid), but they may decrease the probability of drilling if they are relatively low. Moreover, the owner's policy on bid data release might plausibly affect firms' bidding strategies in the auction. Which of these effects dominates is an empirical question, and this question is, to the best of our knowledge, addressed for the first time in Nguyen (2021).

Nguyen (2021) studies U.S. federal OCS leasing from 2000–2019, during which time only 24.5% of leased tracts were actually drilled. Throughout this period, the government's policy was to release the values of losing bids, and the paper begins by showing that the likelihood of drilling is increasing in the value of the highest losing bid (conditional on the winning bid), suggesting a role for information transmission. The paper's goal is to evaluate what drilling and government revenues would have been in a counterfactual in which the government did not disclose the losing bids.

As is standard in models of common value auctions, each bidding firm i in Nguyen (2021) is modeled as receiving a signal S_i of the tract's true expected oil and gas revenue Q. If the winning firm drills, it must pay a drilling cost C that is drawn after the auction from a distribution F^C that is independent of Q. This timing and independence assumption is important to the tractability of the model (and ultimately identification) because it means that firms receive no additional information about Q after the auction, and that there is no private value component to firms' valuations of the tract. The model further simplifies the winning firm's drilling problem by treating it as a static binary choice between not drilling at all versus drilling (and realizing profits equal to $(1-r)Q - C$, where r is the royalty rate).

The key object that must be identified in order to conduct the counterfactual analysis is the function governing firms' posterior beliefs about Q, conditional on the private signals S_i, i.e. $V(s_1, ..., s_N) = E[Q|S_1 = s_1, ..., S_N = s_n]$. However, the fact that Q is not observed for most tracts (since they are not drilled) poses an identification challenge. To identify the distribution of Q, Nguyen (2021) relies on variation in drilling costs as measured by the rental cost for drilling rigs, arguing that these costs affect drilling timing but not Q.

With its model estimated, Nguyen (2021) then examines how outcomes would have differed had the government not disclosed the losing bids. The main result from this counterfactual is that bonus bid revenue modestly declines (because firms recognize that they will be more likely to make "mistakes" when making drilling decisions), but drilling and royalty revenue increase substantially, so that the government's revenue per deepwater tract increases by $734,000 on average. The drilling rate increases because disclosure of losing bids is more likely, on average, to stop a winning firm from drilling than to induce a winning firm to drill. This asymmetry is especially strong on more productive tracts, leading to the substantial increase in production from adopting a non-disclosure policy.

2.2.3 Uncertainty about the number of bidders

Government-run auctions of oil and gas leases sometimes attract many bidders, sometimes attract just one bidder, and sometimes attract none at all. In the commonly-used sealed bid auction format, the number of bidding firms is likely to be not precisely known by the bidding firms. How might this uncertainty affect firms' bidding behavior, particularly when firms are risk-averse?

Kong (2020) addresses this question by studying oil lease auctions for state-owned land in New Mexico, from 2005–2014. The paper begins by documenting that there are many sealed-bid auctions in the data in which there was only one bid, yet that bid was several times greater than the publicly-announced reserve price. A firm that knew it was the only bidder would never bid this way (since it could win the tract by simply bidding the reserve), so these auctions are evidence of firms' uncertainty regarding the number of competitors they face.

Kong (2020) then develops a private-value model with non-selective entry, in which bidders are permitted to be risk-averse.[12] There are N potential bidders, which have values v distributed $F(v)$. Each potential bidder enters the auction with probability p and has a monotonic bidding strategy $b(v; p)$. In this model, if a bidder is risk-averse, it may be optimal to place a bid b that is substantially higher than the reserve, even

---

[12] The paper also allows for asymmetric valuations and entry probabilities for subgroups of bidders. We simplify the discussion by abstracting away from this feature of the paper.

if the expected number of other entrants is not much greater than zero. Placing such a high bid effectively insures the firm against the risk of not winning the parcel. Importantly, this effect exists in first-price sealed bid auctions but not in an English ascending-bid or second-price auction, in which firms' weakly-dominant strategy is to bid their value v if they enter.

To estimate the model, the Kong (2020) first estimates entry probabilities as a function of tract characteristics, using information on the number of potential bidders versus the number of actual bidders in each auction. To separately identify the distribution of firms' valuations from firms' risk aversion, the paper then takes advantage of the fact that New Mexico uses both English auctions and sealed-bid auctions. Under an assumption that the selection of auction format is as good-as-random, conditional on tract observables, the open-outcry auctions pin down the distribution of valuations (Athey and Haile, 2002), and then the sealed bid auction data identify the degree of risk aversion.

Using the estimated model, Kong (2020) shows that the combination of risk aversion with uncertainty about the number of bidders leads to the sealed-bid format generating substantially larger revenues than the open-outcry format, especially in situations where the expected number of bidders is low. Overall, Kong (2020) estimates that converting the sealed-bid auctions in the data to an open-outcry format would decrease the average bonus bid by nearly \$27,000 per tract, equal to 21% of the average winning sealed-bid. The impact of the format change is often on par with that obtained by reducing the number of bidders by one. The large magnitudes of these results indicate a substantial revenue advantage of sealed-bid auction formats, relative to English auctions, in settings with little competition and uncertain bidder entry.

2.2.4 Sequential lease auctions and firms' valuations of neighboring tracts

Governments will sometimes auction neighboring oil and gas tracts in a sequential manner. That is, after a tract is auctioned in a lease sale, a the government may auction a neighboring tract in a following sale. Firms' valuations of the two tracts are likely to be related for at least two reasons. First, because the underlying geology is spatially correlated, firms' valuations for the two tracts are likely to be affiliated (i.e., each firm's value for one tract will be correlated with its value for the other). Second, because there may be synergies (e.g., economies of scale) from developing neighboring tracts together, the winner of the first tract may increase its valuation of the second tract. Is auctioning such tracts sequentially the revenue-maximizing way to go for governments, or should they instead auction the tracts together as a bundle? The answer to this question is likely to be relevant to settings other than oil and gas leases. For instance, FTC spectrum auctions feature licenses of spectrum for geographically adjacent areas, for which bidders' values are likely to be correlated due to affiliation, synergies, or both.

Kong (forthcoming) studies this question using data from New Mexico auctions for state-owned tracts. New Mexico often adopts the practice of offering "pairs" of neighboring tracts on the same day, with one tract being offered first in a sealed-bid auction and the other subsequently offered in an English auction. Thus, firms will know whether or not they won the first tract before bidding on the second.

The core empirical challenge in Kong (forthcoming) is to separately identify affiliations in values from synergies across the tracts. This challenge is akin to the classic problem discussed in Heckman (1981) of separating persistent heterogeneity from state dependence in panel choice data. Kong (forthcoming) addresses this problem via a regression discontinuity design. The paper defines, for each firm in the first auction of each pair, a running variable z that is the difference between the firm's bid and the highest bid among the other firms. The paper then shows that the probability the firm wins the second auction, as a function of z, is discontinuous at $z = 0$, consistent with significant cross-track synergies.

To assess the counterfactual of a bundled auction, Kong (forthcoming) then nests its regression discontinuity design within a private values model of bidding behavior, wherein bidders in the first auction anticipate that winning will yield benefits in the second auction. The estimated model indicates that both synergies and affiliated values are important in this setting, though the latter are relatively more important. Kong (forthcoming) then simulates counterfactuals to show that auctioning tracts as a bundle would increase the state's revenue by 7% on average. This result is driven by bundling's guarantee of achieving synergies, as well as the traditional effect of bundling that it reduces overall dispersion in values (so long as bidders' values for the two tracts in each pair are not too strongly affiliated, see for instance Adams and Yellen (1976)). On the other hand, bundling decreases the total surplus from the tracts, by 2 to 3%, since the allocative efficiency loss from being able to separately allocate each tract to the highest-value firm outweighs the gain from locking in synergies. Thus, Kong (forthcoming) highlights that the optimal bundling strategy for a seller in a multi-unit auction may depend on whether the seller's goal is revenue-maximization or total surplus maximization.

2.2.5 Auctions vs unstructured oil and gas leasing

The vast majority of empirical IO papers on oil and gas leasing study organized government-run auctions of tracts (Herrnstadt et al. (2020) discussed above is one exception). Such settings are ripe for study because the auction rules are publicly announced, bonus bid data can be obtained by researchers, and a single government office can typically provide data on a large number of auctions.

Nonetheless, most oil and gas leases in the United States are for privately, not publicly owned mineral resources. These leases overwhelmingly do not transact in organized auctions but are instead sold in a decentralized marketplace via negotiations. Typically, negotiations for an oil and gas lease of private minerals are initiated by the firm rather than the mineral owner, and survey evidence suggests that many mineral owners execute a lease contract with the first firm that knocks on their door (Ward and Kelsey, 2011).

While leases for privately-owned minerals share the same basic bonus + royalty + primary term structure as do leases for publicly-owned minerals, the markedly different way in which these leases are allocated might plausibly lead to substantially different outcomes for the owner's revenue and resource development. Yet evaluating differences in outcomes between these two formats is challenging. Direct comparisons between private and public parcels are subject to confounds from omitted variables that might be correlated with mineral ownership. And simulating counterfactuals is burdened by the problem that there is a dearth of applicable models of bilateral negotiation of contingent contracts when the buyer has private information.

Covert and Sweeney (2019) makes progress on this question by taking advantage of an unusual situation in Texas whereby some parcels were plausibly exogenously assigned to an auction format, while others were not. For reasons dating back to Texas's time as a Spanish colony, its status as sovereign nation in the 19th century, and the Texas Relinquishment Act of 1919, some of Texas's state-owned minerals exist on split estates where private individuals own the surface rights. On some of this land the surface owners negotiate mineral leases on their own in exchange for half of all proceeds, while on the remainder the state retains a full 100% interest and allocates leases via organized, fixed-royalty bonus bid auctions.

Covert and Sweeney (2019) argues that the assignment of split estate parcels to each of these lease allocation formats is plausibly exogenous with respect to other determinants of the value of shale oil and gas deposits (which have only recently become exploitable). Then, because the bonus bids for the privately negotiated leases are publicly recorded (since the state has a 50% interest in the revenues that flow from them), Covert and Sweeney (2019) can, unusually, make a direct comparison between the revenues from privately negotiated versus publicly auctioned leases. The paper finds that bonuses on auctioned leases are dramatically larger: \$584 to \$1,009 more per acre on average. These values are equivalent to 55%–95% of the average per-acre bonus on private leases. Auctioned leases also have, on average, slightly higher royalty rates, so these bonus differences actually understate the revenue effects of organized auctions.

Covert and Sweeney (2019) also finds that auctioned leases are 8 to 18 percentage points more likely to be drilled and increase oil and gas production by 44 to 74% of the average output of negotiated leases. These effects are consistent with auctions being able to better select firms that will be productive lessees.

Covert and Sweeney (2019) hypothesizes that these results reflect the decentralization and high search costs of private oil and gas leasing markets, where firms must undertake effort to identify leasable parcels and perform title searches. Thus, this market is potentially characterized by "non-sequential" search in which more productive firms are not necessarily more likely to be the first to contact the lessor. The upshot is then that there may be large gains—for both lessors' revenues and total surplus—from policies that either encourage or mandate a more centralized lease allocation mechanism for private oil and gas mineral interests.

2.2.6 Reassignment of leases after the initial sale

One question raised by Covert and Sweeney (2019)'s results on allocative inefficiencies in the private leasing markets is why, given the potentially large productivity gains, the initial lessees do not sell their leases to more productive firms. Such sales—termed "lease reassignments" in the industry—are generally permitted by both private and public oil and gas leases, but the results in Covert and Sweeney (2019) suggest that reassignments are not occurring a frictionless manner.

Progress on understanding the reassignment process has been made by Brehm and Lewis (forthcoming), which takes advantage of variation generated by 1970s lease allocations of federal land in Wyoming that used lotteries rather than auctions. Firms and individuals were able to enter these lotteries for the low fee of just \$10, which naturally attracted a large number of bidders. As a consequence, the lotteries were often won by individuals who were completely unqualified to actually develop the parcel.

The dataset used by Brehm and Lewis (forthcoming) includes information on each tract's lottery winner as well as the second and third-place winners (whose lots were drawn in case the first-place winner did not pay the fee). The paper's empirical strategy is then to regress outcomes (lease reassignment, drilling, and production) on a treatment indicator for whether the winning bidder was an genuine oil and gas firm rather than an unqualified individual. To address potential selection problems associated with endogenous entry into the lotteries, the main specifications in the paper restrict attention to observations in which exactly one of the three "winners" was a firm.

Brehm and Lewis (forthcoming) first shows that the vast majority of leases won by individuals are eventually reassigned, relative to roughly half of leases won by firms. These reassignments preview the paper's next result: the impact of a firm, rather than an individual, winning the lottery on drilling and production outcomes is estimated to be a precise zero. These results are consistent with the lease reassignment market being effective at reallocating leases from low-productivity lessees to high-productivity lessees.

Brehm and Lewis (forthcoming) then studies an interesting subset of the data: leases sold by lottery that are near a lease that was already producing. In these leases, Brehm and Lewis (forthcoming) finds that when a firm wins the lottery, the lease is actually *less* likely to be drilled (by about 10 percentage points) and *less* likely to be productive (by about 4 percentage points). That is, leases that are close to nearby oil and gas production turn out to be more productive when an unqualified individual wins the lottery.

Brehm and Lewis (forthcoming) shows that this empirical result can be explained using a model of private information. If the newly-sold lease is best developed by the firm operating the nearby parcel, the fact that that firm likely has private information about the new lease can be a barrier to exchange. However, if the winner of the lottery is an unqualified individual with a reservation value of zero for the lease, the tremendous surplus gain that would be realized from trade overcomes the informational barrier, and the lease is successfully reassigned to the neighboring firm. Consistent with this story, Brehm and Lewis (forthcoming) finds that reassignment to the nearby firm is significantly more likely if the lottery is won by an unqualified individual than by another firm.

The upshot of Brehm and Lewis (forthcoming) is then that when a potentially productive lease is held by an agent that is completely incapable of developing it, the reassignment market is effective at reallocating the parcel to the most productive user. However, in the presence of private information—which is pervasive in settings where nearby tracts have already been developed—trade may fail to reallocate the lease to the most productive user if the current owner is reasonably productive, even if it isn't the very best firm to develop the lease. These results highlight that the allocation created by the primary leasing stage—when mineral owners initially lease their mineral rights to firms—can have long-run impacts on allocative efficiency despite the existence of a secondary lease market. This implication is consistent with the large efficiency differences between privately negotiated and publicly auctioned leases documented in Covert and Sweeney (2019).

2.2.7 Opportunities for future work on oil and gas leasing

IO researchers have made remarkable progress in recent years in deepening our understanding of U.S. oil and gas leasing. Much of this progress has benefited from new datasets covering state and private leasing activity. Nevertheless, there remain unanswered questions. We see the following topics as especially likely to be fruitful opportunities for future research:

1. The leasing process is initiated in a variety of ways. Many states and the Bureau of Land Management (which manages federal onshore oil and gas leasing) ask firms to nominate tracts for regularly-scheduled lease sales, and then auction off a subset of those tracts. In the federal OCS and some other states, area-wide leasing is used in which firms can bid on any of a large number of tracts in a large

area. And in the private market, leasing is typically initiated in a decentralized manner when a firm approaches a mineral owner. A handful of papers have explored differences between area-wide leasing versus nominations (Hendricks and Porter, 2007) and unstructured (private) vs structured (public) leasing (Covert and Sweeney, 2019), but more work is needed to understand the implications of these differences in how tracts are selected for leasing.

2. Recent work has found evidence supporting both a common-values framework for lease auctions (see, e.g. Bhattacharya et al. (2020)) and evidence supporting private values (Covert and Sweeney, 2019). While oil and gas leasing is traditionally viewed as a canonical example of a common values setting, characteristics of the shale revolution, in which different firms may use different fracking and completion techniques (see section 2.4 below), suggest that private values are likely to be important. One valuable agenda for future work would therefore be to develop econometric models of auctions in which bidders' values have both common and private value components.

3. The literature on oil and gas lease auctions ubiquitously models the mineral owner as trying to maximize some convex combination of revenue, drilling activity, and total (owner + firm) surplus. However, some mineral-owning governments are increasingly concerned about the greenhouse gas emissions associated with oil and gas produced from public lands and waters. Research is needed to understand how leasing policies—including not just the bonus, royalty, and primary term clauses discussed above but also novel tools such as fixed per-barrel fees—might impact leasing, drilling, and emissions, and how such policies might interact with other policy instruments like broad cap-and-trade programs or carbon taxation.

2.3 Information spillovers and externalities

In many situations, firms control leases that lie within the same or nearby pools of oil and gas. Thus, the production from wells drilled on one lease is likely to be correlated with the production from wells drilled on another. This correlation implies that drilling a well generates value not just from the well's actual production but also from the information that is revealed about the pool's quality. The upshot is then that forward-looking firms should consider these information spillovers when making decisions about when to drill exploratory wells.

This subsection discusses how IO economists have modeled firms' strategic exploration incentives given the likely importance of information spillovers in practice. We begin by discussing models of situations in which different firms own nearby leases in the same pool. In this situation, information externalities imply that firms' drilling timing decisions should not be modeled as a single-agent dynamic problem but rather as a dynamic game. This body of work generally shows that "free-riding"—delaying investments to wait and see how well other firms' wells do—is a common behavior in exploratory drilling.

These findings on free-riding are important not just for their direct application to oil and gas exploration but also because they connect to broader a literature on free-riding in research and development. See, for instance, the fairly general model of experimentation and free-riding proposed in Bolton and Harris (1999).

The normative implication is that the overall level of knowledge-generating investment by firms is likely less than that which a full-information, surplus-maximizing social planner would implement.

2.3.1 A theoretical framework for free-riding and the incentive to delay exploration: the "war of attrition"

Economists' understanding of information externalities in oil and gas exploration was first developed in Hendricks et al. (1987), Hendricks and Kovenock (1989), and Hendricks and Porter (1996). Our discussion of this work will focus on Hendricks and Porter (1996), which posits a model similar to that in Hendricks and Kovenock (1989) and expands on the empirical work in Hendricks et al. (1987). We will walk through the model in Hendricks and Porter (1996) in some detail because it is sufficiently simple that it leads to an analytic solution to firms' dynamic drilling game that nicely highlights the mis-coordination and free-riding problems that are present.

Hendricks and Porter (1996) considers a model of drilling in what the paper terms an *area-cohort* of leases: leases that are in a geologically similar area and were awarded in the same auction (and therefore all expire on the same date, five years after the auction). Each tract has a deposit whose size is given by e^X, where X is normally distributed with mean θ and a normalized standard deviation of 1. Capturing the idea that leases in an area-cohort are geologically similar, θ is common within an area-cohort, but its value is unknown by the firms.

If a firm drills an exploratory well on its lease, it incurs a cost c and then its deposit size x is publicly revealed. The value of a deposit is given by $\pi(x)$, which is equal to zero for deposits too small to be commercially viable ("dry holes"). The model abstracts away from time-varying oil prices or drilling costs, so this value is independent of when the firm drills.

Each lease i has T periods in which it may be drilled, where a "period" should be thought of as the time required to make and then execute an exploratory drilling decision (about three months, according to the paper). In the first period, the only information available to each firm about its lease is the signal s_i that is revealed by the auction (which is not modeled). s_i is defined to have a normal distribution with a mean of x_i and precision τ_i. The density function for firms' beliefs about θ is then normal, with mean μ that is a weighted average of the s_i, and a precision ρ, where the weights and ρ are functions of the τ_i. As firms drill their leases, revealing the x_i, Hendricks and Porter (1996) then applies Bayes rule to show that the posterior mean of θ will then shift to be a weighted average of the revealed x_i with the signals s_i on the undrilled leases. Moreover, the beliefs about the size of the deposit on each undrilled lease i will also shift, becoming a weighted average of the posterior mean of θ with the initial signal s_i.

Hendricks and Porter (1996)'s model therefore generates an option value of waiting to drill—even though the oil price is effectively constant—because each firm obtains payoff-relevant information from observing the drilling outcomes of other firms. So when should each firm drill? If a firm's signal s_i is sufficiently high and precise, the firm should drill immediately rather than wait, since other firms' outcomes are unlikely to change its beliefs. But what if several firms each control marginal tracts, for which new information could plausibly shift beliefs about whether the tract is profitable to explore or not? In this situation, firms' drilling problem becomes a dynamic game in which firms would like to wait and gain information from other firms' outcomes, but face risk in doing so since those firms may choose to wait as well.

Hendricks and Porter (1996) at this point focus attention on a simple case of their model in which there are just two firms, i and j. The paper first shows that the subgame in which one firm has already drilled is straightforward: the remaining firm has a belief about its deposit size that is informed by the other firm's outcome, and then given that belief it is optimal to either drill immediately or not drill at all (since no other variables are time-varying). The subgame for period T in which neither firm has yet drilled is also straightforward, since in that case each firm drills if its belief about its deposit size (based on the initial signals s_i and s_j) is such that drilling is expected to be weakly profitable.

Specifically, let V_{it} denote firm $i's$ expected payoff from drilling its lease in period t. V_{it} will be a function of the firm's belief about x_i, and in either of the two subgames discussed above, firm i should drill if $V_{it} > 0$. But what about subgames in which neither firm has yet drilled, and the final period T has not yet been reached? Here, Hendricks and Porter (1996) proceed by backward induction. Consider period $t = T - 1$. Let $W_{i,T-1}$ denote the expected payoff, in period T, to i from waiting one period and then responding optimally to firm j's drilling outcome. Note that we must have $W_{i,T-1} > V_{i,T-1}$ since there is some (perhaps quite small) probability that firm j's outcome will be so bad that it is optimal for firm i to not drill at T. In addition, let $q_{j,T-1}$ denote the probability firm j drills during period $T - 1$. Firm i's expected payoff from waiting is then given by $\beta(q_{j,T-1}W_{i,T-1} + (1 - q_{j,T-1}) \max\{0, V_{i,T-1}\})$, where β is the firms' common discount factor, and the expression takes advantage of the fact that $V_{i,T} = V_{i,T-1}$ if firm j does not drill in period $T - 1$.

Firm i is then indifferent between drilling in period $T-1$ (and obtaining $\max\{0, V_{i,T-1}\}$) versus waiting in period $T - 1$ iff $q_{j,T-1}$ is given by q_j^*, as defined by:

$$q_j^* = \frac{(1 - \beta) \max\{0, V_{i,T-1}\}}{\beta(W_{i,T-1} - \max\{0, V_{i,T-1}\})} \tag{11}$$

Equation (11) tells us that the probability q_j^* of firm j drilling that is required to make firm i indifferent in period $T - 1$ decreases with the discount factor β, so that firm i is more likely to drill immediately if it is impatient. Moreover, the higher is firm i's belief about its deposit size, then the higher is $V_{i,T-1}$, the closer $V_{i,T-1}$ is to $W_{i,T-1}$, and the greater is q_j^* (and q_j^* may exceed 1, in which case firm i drills for sure and obtains expected payoff $V_{i,T-1}$).

A remarkable feature of the model in Hendricks and Porter (1996) is that q_j^* is constant for all periods $t = 1, 2, ..., T - 1$. To see this result, consider period $T - 2$. Because there are no time-varying variables in the model, it must be that $W_{i,T-2}$ is the same as $W_{i,T-1}$. For the same reason, firm i's value $V_{i,T-2}$ from drilling in period $T - 2$ must be the same as the value $V_{i,T-1}$ from drilling in $T - 1$ in the event that firm j has not yet drilled. Thus, the value q_j^* required to keep firm i indifferent in period $T - 2$ must be the same as in $T - 1$. This argument can then be repeated all the way back to period $t = 1$.

The upshot of Hendricks and Porter (1996)'s model is then that, barring a situation in which one of the firms has a signal that is so strong that it drills in period 1 for sure, the two firms play a mixed strategy equilibrium that can be characterized as a war of attrition, "since the payoffs from following (letting the other firm drill first) exceed the payoffs from leading" (p. 393). This equilibrium results in a host of inefficiencies relative to the optimal program that would have been implemented by a single owner of both leases. In that program, any drilling is always completed by the second period, and the decision to drill at least one well accounts for the information that will be generated regarding the profitability of drilling a second well. In the non-cooperative equilibrium, however, drilling may be delayed all the way to the last period, the leases may be drilled simultaneously rather than sequentially, and they may even be drilled in the wrong order.

Hendricks and Porter (1996) then show that the delays predicted by their model are borne out in data from "wildcat" (exploratory) leases in the federal OCS that were awarded from 1954–1979. Hendricks and Porter (1996) defines 270 area-cohorts based on lease sale dates and administrative geographic classifications. Within each area-cohort, Hendricks and Porter (1996) then computes quarterly exploratory drilling hazards. Consistent with the model, these hazards exhibit a strong U-shape, with many leases drilled either in the first few quarters or just before lease expiration. In addition, whether a tract is drilled promptly is positively correlated with a high winning bid for that tract. There is some suggestive evidence that drilling (and successful drilling) on other tracts increases the probability of drilling a given tract, but these results are unfortunately quite noisy. One institutional problem the paper notes is that, while drilling activity is clearly publicly available information, production outcomes may not be (at least not without a significant lag).

2.3.2 Empirical evidence on cross-firm information spillovers and strategic delay

Since Hendricks and Porter (1996), a suite of papers have examined empirically whether, and how strongly, the "war of attrition" affects exploratory drilling in practice. Levitt (2016) studies the determinants of wildcat drilling in Alberta, Canada, using comprehensive data from 1930–2005. Using fixed effects regressions to try to control for unobserved firm-by-play heterogeneity, Levitt (2016) finds that a firm's exploration rate increases with both the number of successful exploration results it had in the past and the number of successful results obtained by other firms. The latter, spillover effect is roughly an order of magnitude smaller than the own-firm effect, but it is nonetheless consistent with information spillovers per a model like that of Hendricks and Porter (1996). An important identification assumption underlying a causal interpretation of the result in Levitt (2016) is that there are no time-varying unobserved factors (e.g., technology improvements) that affect firms' exploration success rates.

In contrast to the empirical work in Hendricks and Porter (1996) and Levitt (2016), Lin (2013) and Hodgson (2018) develop and estimate structural models of firms' dynamic exploration game. The models used in these papers are richer than that of Hendricks and Porter (1996) in a variety of ways, thereby increasing the models' fidelity to the setting at hand but imposing the cost that the model must be solved computationally rather than analytically.

Lin (2013), like Hendricks and Porter (1996), uses data from the federal OCS. Lin (2013)'s model specifies that the profits a firm earns from drilling are directly affected by neighboring firms' exploratory and development drilling. The main objective of Lin (2013) is to estimate the signs and magnitudes of the parameters governing these effects, using data on drilling activity and production outcomes.

Hodgson (2018), in contrast, studies oil and gas exploration in the UK portion of the North Sea during 1964–1990, with the objective of estimating counterfactuals that quantify how the "war of attrition" delays drilling and reduces industry surplus. The model in Hodgson (2018) is explicitly spatial. The paper begins by estimating the spatial correlation of exploration success across blocks using a Gaussian process regression, finding that the probability of successful exploration in any given block is informed by outcomes occurring within one or two neighboring blocks, but not further than that. Motivated by this correlation, the paper then runs regressions similar to those in Levitt (2016), finding that a firm is more likely to explore a particular block if nearby blocks experienced successful exploratory drilling, regardless of whether the success was accomplished by the same firm or a different firm (and controlling for firm, block, and time fixed effects).

To quantify the extent to which free-riding dampens drilling, Hodgson (2018) then develops and estimates an econometric model of firms' exploration decisions. Rather than make each firm's profits a direct function of other firms' drilling, as in Lin (2013), Hodgson (2018) builds a model in which the relevant state variable for each firm in each period (month) is its belief—as generated by the Gaussian process model given the industry's exploration history—about the probability of exploration success on each block. The model simplifies firms' action space each period by specifying that each firm can explore in and then develop at most one block in each period. The structure of the game, like the simpler analytic game in Hendricks and Porter (1996), gives firms an incentive to delay exploration and development in the hope of free-riding on information revealed by other firms' drilling.

In both Lin (2013) and Hodgson (2018), firms make drilling decisions as part of a dynamic exploration game, the solution concept for which is Markov perfect equilibrium. Specifically, each period each firm makes exploration and development decisions based on publicly available information and its own private information about tract-level profitability. Both papers estimate their models' structural parameters using a two step procedure; Lin (2013) uses the estimator from Pakes et al. (2007), while Hodgson (2018) adopts the Bajari et al. (2007) estimator. The core assumption used in both papers to identify information spillover effects is that there are no time-varying unobserved factors that are geographically correlated and affect the probability of successful drilling.

Lin (2013) finds no evidence supporting significant externalities from exploratory drilling. This result echoes the noisy correlations found in Hendricks and Porter (1996) between drilling and outcomes on nearby tracts in the federal OCS. Lin (2013) does, however, find that profits from drilling are positively, and significantly, increased after a neighboring firm drills a *development* well. A natural interpretation of this result is that drilling a development well is a strong signal that the deposit contains a sufficiently large quantity of reserves that development is profitable.

Hodgson (2018) uses its estimated model to simulate a counterfactual that shuts down the incentive to free ride by forcing firms to believe (incorrectly) that other firms will never drill. Doing so increases the number of blocks developed by 1990 by 28%, and industry profit increases by 31%. These substantial effects highlight the importance of strategic interactions in governing firms' exploration behavior, and they illustrate how these interactions can substantially reduce exploration investment and firms' surplus.

Taken as a whole, this literature has developed a useful theoretical framework for understanding the incentives generated by information externalities in oil and gas exploration, and it has taken initial steps at empirically quantifying the importance of these effects in practice. The strongest evidence for significant drilling delays comes from the North Sea (Hodgson, 2018), whereas the evidence from the Gulf of Mexico is more mixed (Hendricks and Porter, 1996; Lin, 2013).

Much remains to be learned on this topic, particularly from the onshore U.S. where the shale boom is taking place. In addition, the work discussed in this sub-section has generally taken lease assignments as given. However, a firm's decision to place a high bid on a particular tract may be related to its beliefs about when (or whether) other firms will develop their leases. Beyond inducing a form of selection into firms' participation in a dynamic drilling game, firms' anticipation of how the game will play out may enter into their behavior during the auction, and thereby impact the sellers' revenues. The implications of links between lease auctions and strategic interactions during subsequent drilling could be developed in future research. Such work would of course be useful for understanding leasing and drilling behavior in this important industry, and at the same time it may also help shed light on investment games in other settings where knowledge spillovers are important.

2.3.3 Exploration spillovers across tracts governed by different extraction policies

The papers discussed thus far in this subsection have studied situations in which different firms control leased exploration tracts, but the mineral owner of all the tracts is the same (a sovereign government). In the onshore United States, however, it is commonly the case that mineral ownership of a pool is fragmented between the federal government, state governments, and private entities. These different owners will typically write leases with different terms and have different policies that govern how wells may be drilled. This "patchwork" of fragmented ownership and governance policies within a pool, within which production outcomes are likely to be geologically correlated, leads to a situation where one owner's policy choices can affect outcomes for other owners. This form of policy spillover is related to but distinct from the economics of the dynamic drilling games discussed above. It is also a form a spillover that has parallels in settings beyond oil and gas exploration. In the pharmaceutical industry, for example, different countries can impose different policies (patents, subsidies, etc.) to support research and development. Because capital can flow between countries, and because innovations can often be applied globally, changes to one country's innovation policy can affect research expenditures elsewhere.

Lewis (2019) examines the effects of patchwork state and federal mineral ownership using data on exploration, drilling, and production in Wyoming. Wyoming is an enticing setting for this question due to the peculiar way in which land is often divided between federal and state owners.[13] Typically, comparing drilling and production outcomes on federal versus state land would be problematic because differences in ownership may have been selected based on underlying geological productivity. Lewis (2019) takes advantage of the fact that, in many places in Wyoming, the pattern of state versus federal ownership is effectively random due to the Land Ordinance Act of 1785, which established a regular land grid and allocated 2 of every 36 square mile sections to the state.

Lewis (2019) begins by developing a model of firms' exploratory and development drilling decisions when facing land in which drilling costs are heterogeneous, motivated by the idea that federal land is likely to be more costly than state land due to federal environmental regulations. A key feature of this model is the fact that the underlying oil and gas productivity on any given section of land is correlated with the productivity of nearby sections. Thus, when drilling costs are relatively low on state land, exploratory

---

[13] The remoteness and harsh climate of the region studied in Lewis (2019) has led to very little private ownership.

drilling on state land increases not just because of the direct cost reduction, but also because exploration is diverted away from nearby federal land. Lewis (2019) then shows that this prediction is borne out in the data: exploratory drilling is particularly likely on state land but particularly unlikely on federal land near to state land.

Lewis (2019) also considers how spatial cost heterogeneity affects the drilling of development wells (which follow exploratory wells if they are successful) and ultimate oil and gas production. Here, while the theoretical predictions are ambiguous, Lewis (2019) again finds evidence of increased drilling and production on state land, and reduced drilling and production on federal land close to state land (relative to federal land far from state land). It is further noteworthy that Lewis (2019) shows that the differences in drilling and production activity on state versus federal land do not materialize strongly until the 1970s, coinciding with the establishment of the EPA and passage of the Endangered Species Act.

A key implication of Lewis (2019)'s findings is that tighter environmental restrictions on federal land cause a non-trivial amount of leakage of drilling and production activity onto nearby state land. Thus, any regulatory cost-benefit analysis that looks only at federal land will fail to account for both the substitutability of nearby state land (which reduces the regulations' costs to firms) and the additional environmental damages occurring on state land (which reduces the regulations' environmental benefits).

2.4 Productivity, innovation, and learning

The most important driver of the shale oil and gas revolution has been the remarkable increase in the productivity of shale wells. In this subsection, we discuss research that has endeavored to understand the sources of productivity growth in the oil and gas industry. Some of these studies have defined and studied "productivity" in the way economists have traditionally done when studying other industries: a Hicks-neutral additive production shifter in an equation specifying the determinants of output. Other work, however, has shed light on forms of productivity improvements—for instance, how firms learn about the production function itself—that have garnered less attention in the economic literature. Together, this body of work has substantive importance for understanding the shale boom itself, and for understanding its future trajectory as oil and gas compete with renewable sources of energy that are themselves becoming less costly to produce. In addition, this work has value for understanding productivity dynamics more broadly, particularly as the well-level investment and output data that are accessible to researchers in the oil and gas business typically do not have an equivalent in most other industries.

2.4.1 Learning about the production function and optimal input choices

Economic models of firms' production and input mix decisions often assume that firms are flawless profit-maximizers and cost-minimizers, given the input and output prices that they face. But for new technologies, the form of the production function may initially be unknown to firms, and it must be instead learned through experience. In such situations, firms' initial input choices will not be consistent with profit-maximization, or even with cost-minimization conditional on output. Moreover, the process of learning about the production function may then be an important mechanism for increases in firms' output, efficiency, and profits.

Covert (2015) makes some of the very first progress on evaluating this learning mechanism by studying firms' fracking input choices, and the resulting production outcomes, in the Bakken shale oil formation in North Dakota. As with essentially all modern shale wells, wells drilled in the Bakken are "fracked" by injecting large volumes of water, sand, and other fluids into a well at high pressure. Doing so fractures the oil-bearing shale in the vicinity of the wellbore, releasing the oil so that it can be produced. The amount of water and sand used, along with other details of the fracking process, are important determinants of each well's ultimate production.

To understand the novel form of learning studied in Covert (2015), consider a simplified version of the paper's production function, in which Q_i denotes well i's output, W_i is injected water, S_i is injected sand, and lat_i and lon_i denote the well's latitude and longitude:

$$\log Q_i = f(lat_i, lon_i) + \beta_i^W \log W_i + \beta_i^S \log S_i + \varepsilon_i \tag{12}$$

A standard econometric evaluation of productivity growth would estimate equation (12) and then measure increases in the total factor productivity residuals ε_i over time. The premise of Covert (2015) is instead that an important way firms increase output Q_i (and profits) over time is by learning about the true values of β_i^W and β_i^S, and then making better input choices.

The first step in Covert (2015) is to estimate the "true" production function using data on all wells drilled during the paper's 2005–2012 sample. The production function model, and the Gaussian process regression used to estimate it, allows the marginal impacts of water and sand, given by β_i^W and β_i^W in equation (12) above, to vary flexibly with each well's geographic location. This flexibility turns out to be important, since the estimates imply that the shape of the production function—and the output-maximizing volumes of water and sand—vary substantially over space. The identification assumption is that, conditional on flexible spatial controls, water and sand inputs are exogenously determined (by, for instance, firms' experimentation with different input mixes).

Once the true production function is estimated, Covert (2015) moves on to study how the profitability of firms' input choices, given the estimated production function, has varied over time. Doing so requires knowledge of the costs of additional water and sand inputs. Covert (2015) obtains this information using data on drilling and completion costs that are available for a subset of wells in the dataset. Covert (2015) finds that in 2005, at the start of the shale boom, firms' input choices are far lower than those that would maximize profits. The average share of profits firms obtain versus what they could have obtained if they chose the optimal input levels is just 21%. The share of profits obtained steadily rises over time as firms use more water and sand in their frac jobs, reaching 60% by 2012. Thus, Bakken firms were able to achieve large increases in profits captured (and production) over seven years by increasing their use of inputs.

Covert (2015) also considers the extent to which firms make good input choice decisions, given the information they have available at the time they drill. That is, for a well drilled on a particular date t, Covert (2015) creates an estimated production function that just uses information from wells prior to t. In this framework, firms' input choices look initially less bad in the early part of the sample. However, differences between the optimal and actual sand and water use increase over time, as optimal sand and water use increase more quickly than actual sand and water use.

Covert (2015)'s results suggest that firms are not making optimal use of the data on outcomes from previously drilled wells that they have available to them. The paper finds that one possible mechanism for this result is that firms focus too heavily on information from their own wells rather than wells drilled by other firms. The paper arrives at this conclusion by showing that firms' input choices are consistent with a model in which each firm estimates a production function in which its own wells receive a weight that is larger than wells drilled by others. Covert (2015) also finds that firms have an aversion to experimenting, since they tend to avoid input choices for which the current-information production function provides a noisy estimate of what the outcome may be. This lack of experimentation may be responsible, at least in part, for why the industry overall was performing below the optimal frontier even in the last year of the sample. Nonetheless, Covert (2015) shows that even with this shortfall, improved input choices drove substantial production and profitability growth in the Bakken shale over 2005–2012.

Another implication of the results in Covert (2015) is that firms considering completing wells in shale formations potentially benefit from waiting for other firms to move first, so that they can learn more in advance about the production function. That is, there is a free-riding incentive based not just on learning about the size of the underground deposit—as in the papers discussed in subsection 2.3—but also on learning about the production function.

These free-riding incentives are considered in Steck (2018), which also uses the Bakken shale as its empirical setting. At its core, Steck (2018) combines the production function estimation of Covert (2015) with the estimation and simulation of a dynamic investment game, as in Hodgson (2018) and Lin (2013). Similar to Covert (2015), Steck (2018) finds that wells' input use and production have increased considerably over time, concluding that firms were learning about the production function. This evidence of learning then motivates a dynamic model in which firms choose both when to drill a well on each of their leases and the inputs with which to frack it. The publicly-observable state in each period includes parameters that dictate how production is likely to respond to increased input use. Firms rationally update their beliefs about these parameters in response to observed input use and production outcomes from wells drilled by other firms.

Steck (2018) finds that, overall, the incentive to learn from other firms' input choices is not strong enough to substantially delay drilling. This finding may be related to the fact that Covert (2015) does not find that firms experiment a great deal, so there is limited value to a firm from waiting to see what other firms do. Steck (2018) does find that, in situations where firms are highly uncertain about the production function—for instance, when few wells have been drilled overall—the delay incentive can become economically large.

In a similar vein, Fetter et al. (2020) examines the extent to which firms learn from observing other firms' choices of fracking inputs, focusing on chemical use in the Marcellus Shale in Pennsylvania. The paper examines a policy change that required firms to publicly disclose the chemicals they include in their fracking fluid, starting on April 2012. Fetter et al. (2020) shows that after the policy change, firms' fracking "recipes" became more similar to one another, using a Jaccard similarity index to measure the similarity between the high-dimensional vectors of chemical concentrations between any given pair of wells. Fetter et al. (2020) also finds that experimentation with new chemical mixes falls following the 2012 policy change.

2.4.2 Learning where to drill

The papers discussed in section 2.4.1 above point to better input selection as an important driver of productivity growth in shale oil and gas. Agerton (2020) examines another potentially important factor: firms learning about where to drill, not just how to drill. This idea, in other settings, would be analogous to firms becoming better at selecting good investment opportunities, not just better at executing them.

The paper studies the Haynesville Shale in Louisiana, in which individual leases are force-pooled into administrative units that are typically one square mile (and in practice line up with Public Land Survey System sections). These units are sufficiently large that firms can drill, frack, and produce from more than one well. The analysis in Agerton (2020) centers on firms' decision of whether, after drilling the first well, they should drill additional wells.

In Agerton (2020)'s model, firms have an initial, noisy signal of how productive the pooling unit is likely to be. Firms will generally find it advantageous to drill at least one well, even if this signal is somewhat poor, since doing so reveals information and indefinitely holds the leased acreage in the entire unit. Given the information revealed from this first well, firms can then decide whether to drill subsequent wells.

Agerton (2020) solves its model by backwards induction. The drilling decision for subsequent wells is an infinite horizon problem, and the drilling decision for the initial well is a finite horizon problem in which drilling results in an immediate (and possibly negative) expected payoff from drilling and in unlocking the continuation value associated with future drilling. The key parameters of the model—drilling costs, time-varying technological productivity and drilling costs, and the noisiness of the initial signal—are estimated via the nested fixed point algorithm of Rust (1987), using data on both drilling timing and wells' output. Agerton (2020) does not explicitly model fracking input choices, so improvements in input use are incorporated into the model's time-varying technological productivity term. To avoid the complexity of modeling a dynamic game, Agerton (2020) assumes that information obtained from drilling does not spill over across firms.

Agerton (2020) finds that changes in firms' drilling location selection over time explain most of the productivity improvements in the Haynesville shale over 2008–2016. In the early years of the Haynesville's development, firms often drilled wells in poor locations, both to hold acreage and because they were not fully informed about local reservoir quality. Subsequently, as most pooling units became held by their initial well, drilling activity shifted to follow-up drilling, but only in units that were highly productive. The upshot is that a naive estimator that doesn't account for this effect would find that technological progress improved wells' productivity by 7% per year. After accounting for site selection, however, technological progress accounts for productivity improvements of just 2% per year.

2.4.3 Productivity in vertical relationships

A feature of many industries—and especially oil and gas—is that production requires contributions from multiple firms that are vertically linked to one another. In oil and gas, the exploration and production companies that acquire leased acreage and organize the drilling and production process actually outsource the drilling of wells to specialized firms that own and operate drilling rigs. Because data are available on both the production companies and drilling companies (and indeed the individual rigs) responsible for drilling a given well, the oil and gas industry offers an opportunity to empirically understand productivity dynamics in vertical relationships.

First, Kellogg (2011) studies data from Texas during 1991–2005, mostly pre-dating the shale boom. The paper's measure of productivity for a given well is the number of days required to drill it. The idea is that an important driver of drilling cost is the time it takes to drill, and in the absence of comprehensive drilling cost data, drilling time is a good proxy for drilling costs. Kellogg (2011) then models drilling time as a function of the producer's own experience, the rig's own experience, and the joint experience between the producer and the rig. To address the possibility that producers and rigs that are especially good matches may work together more frequently, biasing the estimate of the effect of producer-rig experience, the model includes fixed effects for each producer-rig pair (along with field fixed effects to control for geologic heterogeneity and time fixed effects to control for technological progress).

Kellogg (2011) finds that relationship-specific learning is an economically important driver of productivity improvements. Rigs that stick with one producer can be expected to improve their productivity (i.e. reduce their drilling times) twice as quickly as rigs that frequently change producers. The average realized benefit from relationship-specific learning in the data is a nearly 5% reduction in drilling times, which Kellogg (2011) estimates to be worth $12,400 in cost savings per well.

In addition to relationship-specific learning, vertical relationships can also drive variation in productivity over time if the quality of matches between firms varies over time. Vreugdenhil (2020) studies this possibility in the context of producer-rig matching in the U.S. Gulf of Mexico during 2000–2009. In this setting, wells vary in the level of their complexity. For instance, some wells are deeper than others or may require horizontal drilling techniques. Rigs also vary in their ability to drill and complete relatively deep or complex wells. Vreugdenhil (2020) groups rigs into low, medium, and high efficiency based on their depth ratings: the maximum water depth at which they can operate. The paper then shows that, on average, high-efficiency rigs tend to be matched to relatively complex wells. This positive assortative matching is intuitively efficiency-enhancing since the capabilities of high-efficiency rigs are not needed for simple wells.

Vreugdenhil (2020) then documents an intriguing pattern: the strength of positive assortative matching co-varies positively with oil and gas prices. That is, wells and rigs are likely to have better matches during a price boom than a price bust.

To explain this boom-bust pattern, Vreugdenhil (2020) proposes a model of costly search in which new wells ("projects") enter the market each period and must search to find a rig. The paper models search as flexibly falling on a spectrum between fully random and perfectly targeted. The probability of successfully contracting a rig is modeled as an increasing function of the ratio of available rigs to the number of searching projects. This ratio falls during booms and rises during busts.

Upon being contacted by a project, the rig can decide whether to accept the match—in which case the price is determined by Nash bargaining—or wait until the next period. This accept / reject decision then gives rise to the covariance between match quality and boom / bust periods. In a boom, rigs have a strong incentive to pass on poor matches because they have a high probability of being matched in the following period. Thus, matches that are formed are likely to be of high quality. In busts, however, rigs are willing to accept poor matches because the probability of being contracted by a project in future periods is low.

2.4.4 Lessons learned and paths for future work on oil and gas productivity

Understanding productivity growth in the oil and gas business is important for its own sake. The U.S. shale oil and gas boom was a direct consequence of productivity improvements, and the extent to which productivity continues to improve will help determine the extent to which zero-emission energy sources can successfully compete with oil and gas as substitute fuels. In addition, the wealth of data on investments into individual wells, and on wells' ultimate output, has allowed researchers to empirically examine novel mechanisms for productivity growth, such as improvements in input selection. These mechanisms are likely important in other industries but difficult to empirically evaluate because typical datasets aggregate data to the establishment-year level rather than providing data at the level of individual investments.

The most impactful way that IO researchers can build off the progress made thus far would be to investigate whether the ideas and tools used to study oil and gas productivity growth can be applied to technologies like wind, solar, and geothermal energy production. Improvements in these industries' productivity is crucial to future decarbonization. And like oil and gas, these industries feature data on the location and timing of individual investments, so there is potential for the ideas and tools that have been used to study productivity in the oil and gas sector to be transferable to these newer, zero-emission technologies.

2.5 Environmental regulation of resource extraction and transportation

Every stage of the fossil fuel industry—including extraction, transportation, and final consumption—is associated with emissions of local pollutants and greenhouse gases. Tightening regulation of these emissions is arguably the largest issue facing this industry going forward. We highlight in this section how industrial organization research has helped shed light on the impacts (both prospectively and retrospectively) of environmental regulations on resource extraction and transportation.

2.5.1 Regulation of environmental damage at production sites and of site decommissioning

One set of important environmental regulations concerns preventing local harm in the vicinity of production sites. We already discussed in section 2.3.3 one paper—Lewis (2019)—that shows how regulations governing drilling on federal lands decreased federal exploratory drilling but increased exploratory drilling on adjacent state land. These effects amount to a form of *leakage* of drilling activity and (potentially) pollution away from a stringently regulated area to a less-stringently regulated area.[14]

Another recent paper considering leakage is Vreugdenhil (2021), which studies how deepwater drilling rigs respond to increased stringency of regulations aimed at reducing the risk of oil spills and blowouts in the Gulf of Mexico (like the BP Deepwater Horizon disaster in 2010). In this setting, the potential for leakage is international, as deepwater rigs can travel across the ocean to other basins. The main goal of Vreugdenhil

---

[14] In related work in the downstream refining sector, Sweeney (2015) discusses evidence of leakage and environmental policy spillovers in the context of reformulated gasoline standards that only apply in some U.S. gasoline markets.

(2021) is to evaluate the potential leakage from U.S. spill prevention policies that (per engineering estimates) would increase the cost of drilling by 20% while decreasing spills per well drilled by 20%.

To assess this counterfactual, Vreugdenhil (2021) adopts most of the model from Vreugdenhil (2020) (discussed in section 2.4 above). To that earlier framework, Vreugdenhil (2021) adds a model of rigs' location choice. In this discrete choice model, the value of moving to a location j is given by the expected price-cost margin at j, minus the cost of moving to j, plus an idiosyncratic type I extreme value error term. The value of the moving cost is proportional to distance, with the constant of proportionality given by data on rates charged by marine transport ships. Vreugdenhil (2021) then estimates location-specific drilling costs and the scale of the error term by fitting the discrete choice model to empirical choice probabilities. The assumption needed for identification of the scale term—which dictates how responsive rig moves are to price or cost changes in the long-run—is that there are no unobserved time-varying cost changes across location.

After estimating the model, Vreugdenhil (2021) finds in its main policy counterfactual that for every 1 barrel reduction in U.S. oil spills, movement of rigs away from the U.S. causes 0.48 barrels of spills elsewhere in the world. Most of this leakage is driven by the most technically sophisticated rigs, which are estimated to be especially sensitive to the regulation and are associated with larger wells and higher spill volumes. Leakage is not 100% for two reasons. First, rigs are not perfectly price sensitive, leading some rigs to stay in the U.S. market even though U.S. margins are depressed, relative to elsewhere in the world, in the long run. Second, the crowding of rigs into other regions decreases their utilization rates. Still, Vreugdenhil (2021)'s estimate of 48% leakage is large, especially considering that the paper does not incorporate any mechanism for leakage via increases in oil prices.

While Lewis (2019) and Vreugdenhil (2021) are primarily concerned with environmental hazards during the drilling or operation of oil and gas wells, wells at the end of their economic life—when their production rate has declined so far that production revenue no longer covers fixed operating and maintenance costs—also present an environmental challenge. Such wells must be properly decommissioned, not just simply shut in by closing the surface valves, in order to avoid risks from leaks of methane (a potent greenhouse gas) or toxic chemicals into the local environment (Environmental Protection Agency, 2021; Raimi et al., 2021). This decommissioning process typically involves installing a plug in the wellbore and filling it with cement, which can cost $20,000 alone before any required surface reclamation (Raimi et al., 2021).

These decommissioning costs are problematic because, in the absence of strong policy incentives, firms will avoid or postpone decommissioning. Muehlenbachs (2015) illustrates this problem using data from 84,000 wells in Alberta, Canada and a dynamic model of firms' decision whether to continue operating a well, temporarily shut it in, or permanently decommission it. In this model, temporarily shutting in a well preserves the option value of resuming production in the future should oil prices increase. The firm values this option, but crucially does not value the externality imposed by the presence of a non-decommissioned wellbore. Meanwhile, the Alberta government requires decommissioning but does not impose a deadline on firms, who can leave wells temporarily shut-in indefinitely. These circumstances lead Muehlenbachs (2015) to examine firms' incentives to leave wells in a temporarily shut-in state rather than decommissioning them.

A simplified version of the model in Muehlenbachs (2015) is as follows. If a well is producing in a given period, it earns profits of $(P - C)Q$, where P is the oil price, C is the marginal extraction cost, and Q is the production quantity (which declines over time). A well that is temporarily shut-in carries an inactivity cost of M, and a decommissioned well carries no annual cost. The model includes switching costs for transitions between any of these states, leading to hysteresis in firms' decisions. Decommissioning is an absorbing state, and the switching cost into this state represents the decommissioning cost.

Muehlenbachs (2015) estimates its model using the Rust (1987) nested fixed point method. The amount of hysteresis in the data pins down the switching costs, and then the extent to which firms actually decommission wells is informative about the cost M of maintaining a well in a temporarily shut in state. The paper then uses the estimated model to shed light on why so many wells are temporarily shut in rather than decommissioned. The key counterfactual finds that even if natural gas prices were doubled, the number of active wells would increase by just 6% (via switching out of the temporarily shut-in state). This result suggests that the main value firms derive from temporarily shutting in wells is not the option value of restarting them in the future, but rather the value of avoiding decommissioning costs.

A seemingly natural policy solution to the problem of avoided decommissioning would be to force firms to promptly decommission wells following a period of inactivity, or to expose them to full liability for the environmental harms from leaking, non-decommissioned wells. These solutions are unfortunately confounded by difficulties. The liability-centered solution is only effective if gas and fluid leaks are detectable and attributable to specific wellbores, which is challenging in practice. And both solutions face the "judgment-proof problem" (Shavell, 1986) that decommissioning and environmental liabilities can be avoided via bankruptcy. This problem may be especially severe in the oil and gas industry, where wells that are nearing the end of their productive lives can be sold to poorly-capitalized firms who are unlikely to be able to fund decommissioning.

The classic solution to the judgment-proof problem is to require firms, prior to drilling, to post a bond that is large enough to cover the well's decommissioning. Yet in the U.S., state and federal bonding requirements have generally been far short of expected decommissioning costs, leading to a proliferation of "orphaned" wells that do not have a solvent owner (Ho et al., 2018).

Would increasing bonding requirements help slow the creation of orphaned wells? The evidence presented in Boomhower (2019) suggests that the answer is yes. Boomhower (2019) studies a substantial increase in Texas's bonding requirements that was enacted in 2001. This increase led to: (1) the sale of a large number of wells from small operators with poor environmental records to larger firms; (2) a 70% decrease in the number of new orphaned wells; and a 25% reduction in violations of state water protection rules.

The economically large effects documented in Boomhower (2019) provide the first large-scale evidence of the effectiveness of bonding requirements that we are award of. Beyond highlighting the value of increased oil and gas well bonding requirements for jurisdictions outside of Texas, these results also speak to the likely importance of decommissioning bonds for investments in other assets that may create environmental liabilities absent decommissioning and site remediation.

2.5.2 Regulation of emissions from hydrocarbon transportation

The majority of the oil and gas produced in the shale boom is located in rural areas, so that long-distance overland transportation is required to bring these products to market. Considerable popular and policy attention has been drawn to increases in the volume of crude oil transported by rail, which carries the risk of explosive derailments in populated areas (such as the disastrous explosion in Lac-Mégantic in 2013) and is associated with considerable local air pollution from locomotive emissions (Clay et al., 2019).

At the same time, several long-distance oil pipeline projects have drawn attention and concerns regarding both the potential for local spills and the pipelines' effects on upstream production and total CO_2 emissions. The Keystone XL pipeline (from Alberta's oil sands to the U.S. Midwest) and the Dakota Access Pipeline (DAPL, from the Bakken shale in North Dakota to the Gulf of Mexico) have been especially controversial and extensively litigated.

These policy debates provoke the question of how regulation of pipeline construction or crude-by-rail transportation might affect production and transportation of crude oil, along with the associated emissions. Covert and Kellogg (2018) considers this issue in the context of the Bakken shale, DAPL, and crude-by-rail. The paper develops a model that captures the essential characteristics of pipeline vs rail transportation, with the goal of quantifying the degree of substitution between these two modes. Pipeline transport has nearly zero marginal cost but large up-front sunk construction costs, and construction financing typically requires that prospective pipeline customers ("shippers") commit to long-term capacity payments, whether they actually use the capacity or not. These commitments can be as long as 10 years; thus, to spur pipeline construction shippers must commit to the line even though oil prices over such a long time horizon are highly uncertain. Rail transport, in contrast, requires much smaller up-front commitments but does involve non-trivial marginal costs. In addition, rail shipments can be directed to multiple downstream destinations, whereas the pipeline is locked into its terminal once built. The trade-off between pipeline and rail transportation is then that rail transport offers shippers greater temporal and spatial flexibility, but at a higher per-barrel cost.

The simple version of the model in Covert and Kellogg (2018) is as follows. Let the upstream supply curve be given by the function $Q(P)$, and suppose pipeline capacity is equal to K. If the downstream price P_d is relatively low, then that will induce little upstream supply, so that $Q < K$, and with zero marginal transportation cost on the pipeline, the upstream and downstream prices will be equal. But for high downstream prices, enough supply is induced that the pipeline flows at capacity ($Q = K$), and there can be a wedge between the upstream and downstream prices. It is this price wedge that is the reward to pipeline shippers who committed to pipeline capacity in advance. The magnitude of the price wedge is limited, however, by the ability of shippers to move crude-by-rail at a marginal cost r. Were the price wedge to exceed r, crude-by-rail flows would increase, pulling up the upstream price and (after a modest lag due to crude-by-rail adjustment costs) eventually equating the price wedge to r.

The willingness of pipeline shippers to commit to constructing a pipeline of capacity K is then determined by the expected price wedge that will be realized during the duration of the commitment. This expected value is an increasing function of the cost r of crude-by-rail and a decreasing function of K. Thus, regulations that increase the cost of crude-by-rail lead to increases in capacity investment.

Covert and Kellogg (2018) estimates each of the components of its model using monthly data on oil prices, crude production and flows, and costs of rail transportation. The paper's main results then stem from the finding that pipeline and rail transport of crude oil are highly substitutable. An increase in rail costs of $2 per barrel is estimated to motivate an increase in DAPL's capacity of 12–29%, depending on the specification. This substitution implies that the elasticity of rail flows with respect to rail costs is quite large: -0.9 to -2.2. Conversely, if DAPL's construction had been blocked, then crude-by-rail would replace 82–91% of the lost pipeline flows.

The upshot of Covert and Kellogg (2018) is then that environmental regulations targeting just one transportation mode will primarily displace crude oil transportation to the other mode, rather than reduce the overall volume of crude oil transported. Such substitution can be beneficial if the regulated mode is associated with substantially greater environmental externalities than the other mode. However, if the main objective of the regulation is to reduce overall crude oil production, then regulating just one of these two closely substitutable transportation modes is unlikely to be successful.

2.5.3 Interactions between downstream environmental regulations and market power in resource transportation

Reducing emissions from coal-fired electric power generation is essential for controlling both local pollution and CO_2 emissions. A series of papers has demonstrated, however, that the exercise of market power by railroads—which are responsible for nearly all U.S. long-distance coal movements—can substantially complicate the impacts of downstream environmental regulations.

Busse and Keohane (2008), Preonas (2019), and Hughes and Lange (2020) all document evidence that railroads engage in substantial price discrimination on coal movements.[15] In Busse and Keohane (2008), the basis for price discrimination is the enhanced regulation of sulfur dioxide (SO_2) emissions per the Clean Air Act Amendments (CAAA) of 1990. The CAAA specified that a subset of coal power plants ("table A" plants) were newly required to participate in an SO_2 emission allowance trading market, giving them strong incentives to reduce their emissions. One way to do so was to invest in scrubbers—an expensive undertaking. The other option was to switch coal supply to low-sulfur coal in the Powder River Basin (PRB) of Wyoming. This latter option required railroad transportation, and essentially all table A plants were connected to the PRB by just one or two potential rail carriers.

Using data on generators' coal procurement costs, Busse and Keohane (2008) shows that railroads engaged in price discrimination in response to the imposition of the SO_2 trading program. In particular, delivery costs for PRB coal increased for table A plants relative to non-table A plants, and these cost increases were especially large for table A plants that were relatively close to the PRB.

The implication of Busse and Keohane (2008)'s findings is that, in response to a regulation that substantially increased the value of PRB coal, railroads were able to take advantage of their market power to capture that value, at the expense of generators (and likely, ultimately ratepayers). Preonas (2019) and Hughes and Lange (2020) in essence consider the reverse of this situation: the decline in generators' value of coal caused by the shale boom's dramatic expansion of natural gas supply. The reduction in the price of natural gas over

---

[15] Hughes (2011) similarly finds evidence consistent with price discrimination on ethanol movements.

the past decade has enabled gas-fired electric generators to sell power at low prices, curtailing the revenues that coal-fired generators can earn. Both Preonas (2019) and Hughes and Lange (2020) exploit heterogeneity in coal plants' exposure to natural gas competition to show that more exposed plants experienced significant decreases in their coal input prices, consistent with railroad price discrimination. Preonas (2019) shows that coal price decreases were especially large for "captive" plants reliant on only a single railroad, while Hughes and Lange (2020) finds especially large coal price decreases for plants exposed to deregulated wholesale power markets, where competition with gas-fired generators would be most severe.

Preonas (2019) goes on to draw an analogy between its results and the potential impacts of carbon pricing in the electricity sector, which (like a decrease in the price of natural gas) would disadvantage coal-fired generators. Preonas (2019) points out that if an analyst were to forecast changes in emissions from such a tax under an assumption that coal were priced at marginal cost, the analyst would likely over-estimate the extent of coal-to-gas switching—and therefore emissions reductions—that would be caused by the policy. The upshot of this set of papers is then that the impacts of environmental policies that impact fuel use at the point of consumption (here, electric power generators) can be influenced substantially by firms (here, railroads) who are upstream in the supply chain.

2.5.4 Opportunities for future work on impacts of environmental regulation on fossil fuel extraction and transportation

Industrial organization economists have made excellent progress in understanding the behavioral responses of resource extraction and transportation firms to environmental regulations. We view this literature as still being in an early stage of development, and the likelihood of continued increases in regulatory stringency (particularly regarding greenhouse gas (GHG) emissions) makes this area ripe for future work. In particular, we see the following areas as especially high-potential for research contributions:

- **Methane emissions.** Methane is an especially potent greenhouse gas, and oil and gas extraction has been identified as an important source of methane emissions. Methane emission control is bedeviled by the difficulty of leak detection and substantial heterogeneity across firms and sites in leak rates. IO research can be valuable in designing and evaluating regulatory schemes, and the enforcement of such schemes, that aim to reduce methane leaks.

- **Leasing policy and climate change.** Policies that put a price on carbon emissions will interact with oil and gas leasing, which historically has put an implicit price on oil and gas production, via the imposition of royalties. In addition, the response of lease terms to carbon prices will determine how the loss of producer surplus is split between mineral owners and firms, and the extent to which carbon pricing reduces drilling and production activity versus reducing the royalties and bonuses imposed by mineral owners.

- **Applications to zero-emissions energy sources.** Are economic models and lessons learned from the oil and gas sector applicable to zero-emissions technologies that, like oil and gas, rely on geographic-specific inputs? Like oil and gas, large-scale geothermal, wind, and solar energy production all require contracts with landowners, and as was the case with the shale boom, productivity improvements will be central to the success of these technologies. As data on these energy sources becomes more widespread, application of models and research designs originally conceived for oil and gas are likely to be fruitful in understanding the economics and these zero-emissions technologies.

3 Personal transportation, energy use, and environmental regulation

This section discusses the industrial organization of transportation, energy use, and the environment. We focus our attention on how ideas and methods from industrial organization have helped (and going forward, can continue to help) answer questions about the impacts of policies aimed at reducing emissions from the transportation sector. These policies include fuel taxes, fuel economy standards, regulation of emissions of local pollutants, and incentives for adoption of alternative fuel vehicles—all of which have been implemented or received serious consideration in a variety of jurisdictions.

While our discussion will emphasize the substantive contributions of industrial organization to these policy-relevant topics, we also note that work in this area has also made positive contributions to broadly relevant IO topics and methods, including estimation of consumers' valuation of goods' attributes, models for how firms set product characteristics in differentiated oligopoly markets, models of consumers' search behavior in retail markets, and evaluation of the impacts of incentives and standards in network goods markets (here, electric vehicles (EVs) and EV charging stations). The contributions we discuss also relate to the IO literature on automobile markets that endeavors to understand consumers' vehicle demand, firms' behavior, and market equilibrium. We will not extensively discuss this broader literature here—much of it is covered in other chapters of this handbook volume—though many of the papers we discuss will use tools from that literature, especially the differentiated products demand and Bertrand oligopoly model in Berry et al. (1995). Because the methodological details behind implementing these tools are discussed elsewhere in this volume, here we will emphasize how these tools enable papers' substantive contributions rather than devote significant time to the tools themselves. That said, we will highlight instances where papers at the intersection of IO, transportation, and the environment augment IO methods with new features that are important for capturing phenomena relevant to the question at hand. For instance, we will discuss how the constraints imposed by fuel economy standards affect firms' pricing decisions in Bertrand competition, as modeled in Jacobsen (2013).

The section proceeds in five subsections: (1) consumers' demand for fuel economy and associated implications for fuel economy policy; (2) economic impacts of fuel economy standards; (3) regulating emissions of local air pollutants; (4) consumer search in retail gasoline markets; and (5) markets for EVs and EV charging stations.

3.1 Estimating consumers' demand for fuel economy, and implications for fuel economy policy

Vehicle fuel economy is an important vehicle attribute that consumers value. The distance a particular vehicle can be driven per dollar spent on gasoline (miles per dollar) was included as a demand-side attribute in the original Berry et al. (1995). Vehicle fuel economy has also long been a focus of policy attention, given interest in reducing the externalities associated with gasoline consumption.

Policy-makers have a variety of tools at their disposal to improve vehicles' fuel economy. Two prominent such tools have been gasoline taxes and fuel economy standards. Gasoline taxes have long been upheld by environmental and public finance economists as the most efficient approach for reducing greenhouse gas emissions from automotive fuel consumption. As Anderson and Sallee (2016) notes in a recent review of this literature, an important reason behind economists' favorable view of gasoline taxes is that that they provide incentives to reduce fuel use via not just improving vehicles' fuel economy, but also by reducing miles traveled.[16]

Anderson and Sallee (2016) also notes, however, that gas taxes may not be fully effective if consumers fail to fully internalize future fuel costs when they purchase vehicles. That is, if consumers are myopic and under-value future fuel costs relative to the up-front vehicle price, then their response to a gasoline tax (in terms of purchasing more efficient vehicles) will be attenuated relative to what a model assuming fully rational, forward-looking behavior would predict. An argument for fuel economy standards is then that they are not adversely affected by this consumer "myopia" because they can directly mandate a high level of fuel economy. Moreover, under this argument fuel economy standards can be viewed as actually improving consumers' welfare—in a paternalistic sense—since they can correct the "internality" caused by consumers' myopic decision-making.

An essential input to the discussion of fuel taxes versus fuel economy standards is then empirical evidence on the extent to which consumers actually value future fuel costs when they purchase a vehicle. This subsection discusses a suite of papers that employ demand models to answer this question, using a variety of research designs. Before discussing individual papers, however, we think it will be useful to highlight several of the empirical challenges that must be overcome (or assumptions that must be made) in order to credibly address the question.

To begin, consider the following model of consumers' utility from purchasing a vehicle, letting i denote a consumer and j denote a particular model:

$$U_{ij} = -\alpha p_j - \gamma F_j + \beta X_j + \xi_j + \epsilon_{ij} \tag{13}$$

In equation (13), p_j denotes the vehicle's price, F_j denotes its expected lifetime fuel costs, and X_j denotes other vehicle characteristics. ξ_j then denotes an unobserved vertical product characteristic, and ϵ_{ij} is an idiosyncratic error term. The goal is to test whether γ is equal to α. If that null hypothesis is rejected, and instead $\gamma < \alpha$, we would take that as evidence that consumers are myopic. And in that case, the ratio γ/α tells us how many cents on the dollar consumers value future fuel costs relative to the up-front vehicle purchase price.

One immediate obstacle to estimating equation (13) is that doing so requires a computation of expected

---

[16] Anderson and Sallee (2016) offers a brief survey of the literature that estimates the elasticity of gasoline consumption or vehicle miles traveled with respect to the price of gasoline. This literature can rightly be described as enormous, and it spans the fields of environmental economics, public finance, macroeconomics, and industrial organization. We therefore do not cover gasoline demand estimation in this chapter but rather direct readers to Anderson and Sallee (2016) and the reviews and papers cited therein.

future fuel costs (at the time of vehicle purchase) F_j. Computing F_j requires estimates of how many miles the vehicle will be driven each year, the vehicle's expected remaining life before scrappage, the vehicle's fuel economy, expected future fuel prices, and the consumer's discount rate. None of these components is straightforward to obtain or compute, and the papers we describe below use a mix of approaches and datasets, as we will discuss.

Second, the test of consumer myopia as we described it above assumes that consumer tastes—or at least the parameters α and γ—are homogeneous. All of the papers in this literature make this assumption (sometimes implicitly, sometimes explicitly).[17] Doing so certainly adds value in terms of yielding a sharp answer to the research question, but of course in reality tastes or the degree of myopia across consumers may vary. In that case, the object of interest would be the distribution of the relevant preference parameters, and therefore myopia, across the population.

Finally, identification of γ in equation (13) requires an unconfoundedness assumption that future fuel costs F_j are not correlated with unobserved vehicle characteristics that constitute ξ_j. An important component of F_j is the vehicle's fuel economy, which will typically be correlated with a variety of other performance characteristics of the vehicle. Because available and operationalizable data on vehicle characteristics are typically limited—often to variables like horsepower, weight, acceleration, and footprint—this unconfoundedness assumption can be difficult to defend. Thus, papers in this literature have tended to adopt empirical strategies that isolate the variation used to identify γ to sources other than cross-sectional variation in vehicles' fuel economy.

3.1.1 Identifying consumers' valuation of fuel costs from used vehicle prices

This subsection discusses three papers closely-related papers—Busse et al. (2013), Allcott and Wozny (2014), and Sallee et al. (2016)—that study consumers' valuation of fuel economy by using data on used vehicle markets, and leveraging used vehicle price variation induced by shocks to the price of gasoline. All three of these papers conclude that consumers fully, or nearly fully, value future fuel use when they are purchasing a vehicle.

A motivation for studying the used car market is that the supply of a particular vehicle model by model-year (for instance, a 1996 Honda Accord) can be approximated as fixed. With fixed supply and with homogeneous preferences for vehicle attributes, the preference parameters in equation (13) can then be estimated directly using the hedonic pricing equation (14):

$$P_j = -\frac{\gamma}{\alpha}F_j + \frac{\beta}{\alpha}X_j + \xi_j \tag{14}$$

The benefit of this hedonic approach is that it is not necessary to estimate consumers' price sensitivity α, so that the standard equilibrium price endogeneity problem of demand estimation can be avoided.[18] There are at least two caveats to the "fixed supply" assumption underlying this approach, however. First, to the

---

[17] Grigolon et al. (2018), which we discuss below, does somewhat weaken this assumption by modeling heterogeneity in consumers' vehicle miles traveled.

[18] Busse et al. (2013) includes analyses of new car transactions, in which the paper estimates fuel use valuation after imposing a new car demand elasticity from the literature. The paper then concludes that buyers of new cars also fully value future fuel use.

extent that consumers substitute between used cars and new cars, quantity changes in the new car market will affect equilibrium outcomes in the used car market. Second, any price-sensitivity of the vehicle scrappage rate (which Jacobsen and van Benthem (2015) finds evidence for) will violate the assumption. Allcott and Wozny (2014) relies on the hedonic model for its main estimates but explores the potential importance of both of these caveats in an appendix, finding that its results do not change substantially.

To address potential omitted variables biased caused by unobserved vehicle characteristics that are correlated with fuel economy, each of Busse et al. (2013), Allcott and Wozny (2014), and Sallee et al. (2016) adopts a strategy of including detailed vehicle model by model-year (or in the case of Allcott and Wozny (2014) model by model-age) fixed effects in its pricing equation and then leveraging variation in future fuel costs that is induced by changes in fuel prices. Thus, these papers estimate equations like (15) below, where t indexes the month-of-sample, and μ_j denotes the model by model-year (or model-age) fixed effects:

$$P_{jt} = -\frac{\gamma}{\alpha} F_{jt} + \mu_j + \tau_t + \varepsilon_{jt} \tag{15}$$

Equation (15) includes month-of-sample fixed effects τ_t to capture market-wide changes in vehicle prices (and preference shifts to or from the outside good), and the error term ε_{jt} accounts for period-to-period changes in model-specific tastes. Model j's lifetime fuel costs F_{jt} are now a function of fuel prices at t, and the underlying fuel price variation is assumed to be uncorrelated with ε_{jt}.

To illustrate the calculation of F_{jt}, we follow Allcott and Wozny (2014); other papers in this literature make similar calculations. Allcott and Wozny (2014) obtains information on vehicle miles traveled (VMT), differentiated by vehicle class and age, from the National Household Travel Survey (NHTS). It obtains survival probabilities from vehicle registration data. The paper uses the Environmental Protection Agency (EPA) combined (city and highway) fuel economy rating for each vehicle. For a discount rate, Allcott and Wozny (2014) averages interest rates associated with equity returns and vehicle loans, in order to capture the opportunity cost of funds for cash transactions and financed transactions, respectively.

The final component of F_{jt} is the expected path of gasoline prices over the lifetime of the vehicle, which is not necessarily the same as the gasoline price at the time of the vehicle's purchase. One approach to modeling future expected fuel prices—and that used by Allcott and Wozny (2014) in its headline specification—is to use futures market prices. Another is to use survey data on consumers' beliefs from the Michigan Survey of Consumers, which Anderson et al. (2013) finds correspond closely to a belief that gasoline prices are a martingale.

Allcott and Wozny (2014) finds that consumers modestly under-value future fuel costs, estimating a value of $\gamma/\alpha = 0.76$ in its baseline specification. In alternative specifications this value ranges from 0.46 to 1.01, though the full valuation result requires a large (15%) assumed discount rate. Busse et al. (2013) finds results that generally imply full valuation. The difference in findings between Allcott and Wozny (2014) and Busse et al. (2013) is not large and may be attributable to either their use of different datasets (Allcott and Wozny (2014) uses prices from wholesale transactions while Busse et al. (2013) uses retail transactions) or differences in their specifications.[19]

---

[19] In addition to the difference in model by model-year versus model by model-age fixed effects noted above, Busse et al. (2013) differs from Allcott and Wozny (2014) in that it replaces F_{jt} with an interaction between vehicle fuel economy and the gasoline

Finally, Sallee et al. (2016) also finds results consistent with consumers fully valuing future fuel costs, but identifies γ/α using a different source of variation: heterogeneity in odometer readings within vehicles of a particular model by model-year. For instance, consider two 2002 Toyota Priuses (a very efficient model), one of which has a lot of miles on it and the other very few. Following a gasoline price increase, we would expect the price of the vehicle with fewer miles on it to be more responsive to the gas price shock, since it is likely to have a longer remaining life. Sallee et al. (2016) isolates this variation in the data by carefully computing F_{vjt} in a way that accounts for the odometer reading of each particular vehicle v, and by including fixed effects for all interactions between model, model-year, and month-of-sample. The identifying variation in Sallee et al. (2016) is therefore more restricted than that used in Allcott and Wozny (2014) and Busse et al. (2013), but nonetheless Sallee et al. (2016) finds a similar headline result: approximately full valuation of future fuel costs.

3.1.2 Consumer valuation of fuel costs for new vehicles

An important limitation of the studies discussed in section 3.1.1 above is that they only speak to consumers' behavior when purchasing used vehicles, not new vehicles. Because new vehicles are, by definition, only new for one year, research designs based on leveraging only variation in fuel prices over time (e.g., by including model by model-year fixed effects in the estimating equation) will not work when only data from new vehicle markets are used. So alternative research designs are necessary to study consumer behavior in these markets.

Grigolon et al. (2018) studies new vehicle transactions in seven European countries during 1998–2011. The paper leverages the fact that even within a given model by model-year, there exists variation in vehicles' fuel economy because consumers can choose between different engine types. In particular, many models are offered with both diesel and gasoline engine types, and diesels achieve considerably better fuel economy than do gasoline-powered vehicles (at the cost of a higher sticker price). At the same time, Grigolon et al. (2018) also allows for heterogeneity in consumers' VMT, so that consumers who anticipate driving more will be more likely to select relatively efficient vehicles.

Grigolon et al. (2018)'s embeds this within-model variation in fuel economy into the following model of consumers' indirect utility function:

$$u_{ijkt} = -\alpha p_{jkt}/y_t - \gamma\rho\beta_i^m e_{jkt} g_{kt}/y_t + x_{jkt}\beta_i^x + \xi_j + \xi_t + \xi_{jkt} + \epsilon_{ijkt}, \tag{16}$$

where i indexes consumers, j indexes vehicle models, k indexes engine variants, and t indexes country by month-of-sample. p_{jkt} denotes the vehicle's sticker price. The marginal utility of money is assumed to be inversely proportional to income, so the utility loss from paying for vehicles and fuel is scaled by market-level income y_t. The second term on the right-hand-side of equation (16) denotes discounted expected future fuel costs, calculated as the product of the consumer's expected annual VMT β_i^m with the vehicle's fuel use per mile e_{jkt}, fuel prices (either gasoline or diesel) g_{kt}, and a capitalization coefficient $\rho \equiv \sum_{s=1}^{S}(1+r)^{-s}$

---

price in the main estimating equation. Then, after estimation, Busse et al. (2013) transforms the parameter estimate to an implied discount rate using information on VMT and scrappage rates.

(where S is the vehicle's expected lifetime).

As with the literature discussed in section 3.1.1 above, a goal of Grigolon et al. (2018) is to test whether $\alpha = \gamma$. The crucial identification assumption is that each vehicle's fuel economy is uncorrelated with the vehicle's other unobserved characteristics. To help buttress this assumption, the utility specification (16) includes other observed characteristics x_{jkt} (such as horsepower and size), model fixed effects ξ_j, and market fixed effects ξ_t. The identification assumption is that other vehicle-specific unobservables ξ_{jkt} are uncorrelated with inverse fuel economy e_{jkt} and other characteristics x_{jkt}.

Grigolon et al. (2018) estimates equation (16) using the methods and instrumental variables discussed in Berry et al. (1995), with the twist that the distribution of consumers' VMT β_i^m is restricted to follow the distribution of reported VMT from survey data. It finds an estimate of $\gamma/\alpha = 0.91$, consistent with nearly full valuation of future fuel costs by consumers. This result is robust to alternative specifications that shut down preference heterogeneity over vehicle characteristics and fuel use.

Grigolon et al. (2018) also leverages the estimated heterogeneity in VMT to make an important point about the efficacy of fuel taxes versus product taxes (or fuel economy standards). Fuel taxes induce high-VMT consumers to differentially select efficient vehicles, whereas product taxes do not. The model in Grigolon et al. (2018) allows the paper to quantify this effect, finding that a fuel tax reduces total fuel use by 18%, relative to a reduction of just 12% from a revenue-equivalent product tax.[20]

Another recent paper that studies the new vehicle market, but comes to a different conclusion, is Gillingham et al. (forthcoming). This paper studies the U.S. market, exploiting variation in fuel economy ratings induced by a 2012 scandal in which two automakers, Hyundai and Kia, were caught overstating the fuel economy of several of their top-selling vehicle models. As one of several consequences, the U.S. EPA required Hyundai and Kia to restate the fuel economy of these vehicles. Gillingham et al. (forthcoming) exploits the changes this restatement had on new vehicle prices and sales volumes—using transaction-level sales data—to estimate consumers' valuation of future fuel costs.

Gillingham et al. (forthcoming) begins by providing narrative evidence that the fuel economy restatement was sudden and likely unexpected by consumers. It then estimates the effect of the restatement on equilibrium vehicle prices using a panel fixed effects (difference-in-difference) regression that focuses the identifying variation on changes in prices for affected versus unaffected models produced by Hyundai and Kia. The result is that the restatement reduced the price of affected vehicles by 1.2%, equal to about \$300 on average. The paper invests considerable effort in showing that this estimate is primarily driven by decreases in the price of affected vehicles rather than increases in the price of these vehicles' close substitutes.

Based on estimates of VMT, future fuel prices, expected vehicle lifetimes, and discount rates that follow the assumptions used by the papers discussed in section 3.1.1, Gillingham et al. (forthcoming) finds that this sticker price decrease is substantially less than change in the affected vehicles' future fuel expenditures. With a 4% discount rate, consumers are estimated to value future fuel costs by only 17% of how they value the up-front purchase price.

An alternative explanation for this under-valuation result is that the quantity supplied of the affected vehicles also declined, attenuating the decrease in the equilibrium price. However, Gillingham et al. (forth-

---

[20] To compute these counterfactuals, Grigolon et al. (2018) assumes Bertrand competition by automakers, per Berry et al. (1995).

coming) estimates a small and statistically insignificant positive effect of the restatement on equilibrium quantities: +5% with a standard error of 4%. Moreover, even if one postulates that quantities actually decreased by 5% (5 times the size of the estimated price decrease), consumers estimated valuation of future fuel costs doubles at most.[21]

3.1.3 Lessons learned and paths forward for research on consumers' valuation of fuel economy

The extent to which vehicle consumers value future fuel expenditures when making purchase decisions is a crucial input for evaluating fuel economy policy. Using a variety of demand models, estimation methods, and datasets, research on this question has made substantial progress over the past 15 years. Three papers—Busse et al. (2013), Allcott and Wozny (2014), and Sallee et al. (2016)—all find that in the U.S. used vehicle market, consumers fully or nearly fully value future fuel costs at the time of purchase. Grigolon et al. (2018) finds the same for the European new vehicle market, but Gillingham et al. (forthcoming) arrives at a starkly different result—substantial undervaluation—for the U.S. new vehicle market.

A potential explanation for the discrepancy between Gillingham et al. (forthcoming) and the other papers we discussed is that the population of new car buyers in the U.S. behaves differently than the population of U.S. used car buyers (or European new car buyers). Another is that consumers are more attuned to changes in gasoline prices or differences in engine types than they are to changes in fuel economy ratings. Yet another is that the undervaluation result is simply specific to the case examined in Gillingham et al. (forthcoming): a scandal-induced fuel economy restatement affecting two automakers. Future research is needed to help distinguish between these explanations for these papers' diverging results.

In addition, more work is needed on heterogeneity in consumers' valuation of fuel economy. With the exception of Grigolon et al. (2018)—which allows for valuation heterogeneity that is induced by heterogeneity in VMT—all of the papers discussed in this section use models in which consumers' valuation of fuel economy is homogeneous. As pointed out in Anderson et al. (2013), consumers' valuations are likely to be heterogeneous not just due to VMT, but also due to heterogeneity in local gasoline prices, beliefs about future gasoline prices, discount rates, and inattention. The welfare effects of fuel economy policies depend on this heterogeneity and which mechanisms lie behind it, so research is needed to better understand it.

Finally, it would be useful to extend the literature on consumers' valuation of future fuel costs to study alternative fuels, and in particular electric vehicles (EVs). One component of the appeal of EVs is that the price per unit energy for electricity is likely to be lower than that for gasoline. Even if the literature that studies consumers' valuation of fuel economy for gasoline-fueled vehicles reaches a consensus, it is not obvious that consumers' behavior when choosing between vehicles with high versus low future gasoline expenditures will translate directly to choices between gasoline-powered versus electric-powered vehicles. Research is needed to help inform how variation in the relative prices of electricity and gasoline will affect consumer adoption of EVs.

---

[21] To complete the calculation, Gillingham et al. (forthcoming) also uses estimates of consumers' demand elasticity from the literature.

3.2 Economic impacts of, and firms' responses to, fuel economy standards

Many countries have tightened their fuel economy regulations over the past decade, often with reducing CO_2 emissions as the stated main policy objective. Understanding the cost-effectiveness and distributional impacts of fuel economy standards, especially relative to fuel taxation, is essential for informing fuel economy policy-making. One input for understanding these impacts is consumers' valuation of fuel economy, as discussed in subsection 3.1 above. That input alone is not sufficient, however, as the effectiveness of fuel economy standards will also be heavily influenced by firms' behavioral responses.

Ideas and methods from industrial organization—especially differentiated products demand and Bertrand competition models related to Berry et al. (1995)—have proven themselves useful for research into how firms react to fuel economy standards. This subsection discusses the progress made by this literature. We begin by discussing Jacobsen (2013), which illustrates how binding fuel economy standards alter vehicle manufacturers' pricing decisions relative to what would be predicted by a standard Bertrand pricing model. We then discuss a series of papers that follow up on Jacobsen (2013) by exploring other mechanisms with which firms can comply with fuel economy standards, with particular emphasis on the trade-offs firms face between improving fuel economy versus improving other vehicle characteristics.

3.2.1 Fuel economy standards and automakers' pricing and fleet mix decisions

Fuel economy standards reduce fuel use and CO_2 emissions by setting a ceiling on the fuel use per mile of newly sold vehicles. When the U.S. first introduced its Corporate Average Fuel Economy (CAFE) standard in 1978, compliance was evaluated at the original equipment manufacturer (OEM) level. That is, the sales-weighted average fuel use per mile within an OEM needed to fall weakly below the constraint. Since 2009, regulations have permitted CAFE compliance credit trading across firms.

Jacobsen (2013) develops a model for how the pre-2009 CAFE standards affected OEMs' pricing decisions, equilibrium outcomes in vehicle markets, and the surplus received by consumers and producers. The paper's model is rich, as it includes models of consumers' demand and driving behavior, Bertrand price competition among OEMs, and the used car market. Our discussion will refrain from extensively discussing all of these components but rather focus on the paper's key innovation regarding modeling and estimating the impacts of binding CAFE constraints on the market's Bertrand pricing equilibrium.

Perhaps most importantly, Jacobsen (2013) accounts for the fact that the U.S. CAFE standard, in the absence of credit trading across firms, had different impacts on three distinct types of firms. First, some OEMs, such as Honda and Toyota, produced vehicles that were so efficient that the standard was not binding. Second, another group of firms, such as BMW and Mercedes, chose not to comply with the standard and instead paid fines that were a linear function of their non-compliance. The behavior of these first two groups of firms had been considered by earlier work (Goldberg, 1998). But Jacobsen (2013) argues that a third group of firms—consisting of the "Big 3" U.S. automakers Chrysler, Ford, and GM—treated the CAFE standard as a binding constraint, since violating the standard may have caused them to incur reputational and political costs that substantially exceeded the explicit fines. Consistent with this view, these three automakers' weighted-average fuel economy closely matched the standard throughout the period studied in the paper.[22]

Jacobsen (2013) shows that these three types of firms face profit maximization problems that differ in important ways. First, consider firms for whom CAFE does not bind. Jacobsen (2013) models them as solving a standard differentiated product Bertrand profit maximization problem, per equation (17):

$$\max_{p_j, j \in \mathcal{J}} \sum_{j \in \mathcal{J}} (p_j - c_j) q_j(P), \tag{17}$$

where j denotes vehicle models and $\mathcal{J}$ denotes the set of vehicles produced by the firm. p_j and c_j denote vehicle prices and marginal costs, and $q_j(P)$ denotes the quantity sold of each vehicle as a function of the full vector P of all vehicle prices in the market (including vehicles sold by other firms).

Now consider a firm that is out of compliance and pays fines. The firm's problem now includes a fine term that is proportional to the product of total sales with the difference between the fuel economy standard d (in miles per gallon) and the firm's sales-weighted average fuel economy $CAFE$. Thus, per equation (18), the firm is effectively taxed on its sales of inefficient vehicles.[23]

$$\max_{p_j, j \in \mathcal{J}} \sum_{j \in \mathcal{J}} (p_j - c_j) q_j(P) - \gamma(d - CAFE) \sum_{j \in \mathcal{J}} q_j(P). \tag{18}$$

Finally, consider the "Big 3" firms that treated the standard as a binding constraint. These firms faced the constrained maximization problem in equation (19) below, where mpg_j denotes the fuel economy of a particular model j:

$$\max_{p_j, j \in \mathcal{J}} \sum_{j \in \mathcal{J}} (p_j - c_j) q_j(P) \quad \text{s.t.} \quad \frac{\sum_{j \in \mathcal{J}} q_j(P)}{\sum_{j \in \mathcal{J}} \frac{q_j(P)}{mpg_j}} - d \geq 0. \tag{19}$$

The constraint in equation (19) forces automakers to decrease their sales of their inefficient vehicles and increase sales of efficient vehicles. To do so, they must set markups on efficient vehicles that are smaller than what would be optimal in standard Bertrand competition (equation (17)), and markups on inefficient vehicles that are larger.

Jacobsen (2013) then integrates the firms' problems (17), (18), and (19) into an equilibrium model of the vehicle market. The paper adopts a demand model from Bento et al. (2009) that is similar to that in Berry et al. (1995). It estimates the demand system in a first step and then estimates the supply model using the FOCs implied by equations (17), (18), and (19). In addition to the usual estimation of each vehicle model's marginal cost, this step also involves estimating the shadow value of the CAFE constraint for the three firms for which it binds. Doing so requires additional restrictions, since for any shadow value there exists a set of marginal cost values that could rationalize the data. Jacobsen (2013) therefore restricts OEMs' markups to be proportional to markups imposed by dealers (which are observable), per an argument from Bresnahan and Reiss (1985). The extent to which these markups are relatively low for efficient vehicles (versus what would

---

[22] In practice, the regulation allows some banking and borrowing of credits across years, so these firms do not have to satisfy the standard every year.

[23] We simplify the exposition here by ignoring the different standards imposed for cars versus light trucks. Sales-weighted average fuel economy is computed as a sales-weighted harmonic mean of each model's fuel economy in miles per gallon. During the period studied in Jacobsen (2013), the fine parameter γ was equal to 50.

be implied by the inverse demand elasticity) is informative about the importance of the CAFE constraint.

Jacobsen (2013) estimates that the shadow values of the CAFE constraint for Chrysler, Ford, and GM are quite large. For instance, the marginal effect of loosening the standard by one MPG for GM's passenger car fleet is estimated to be \$438 per vehicle. Note, however, that this shadow value describes the effect of tightening the standard for just GM, while leaving the standard for other OEMs unchanged. Quantifying the effects of an overall tightening of the CAFE standard requires a counterfactual simulation.

To assess counterfactuals, Jacobsen (2013) integrates its model of the new vehicle market with a model of the used vehicle market from Bento et al. (2009), in which vehicles are probabilistically scrapped as a function of their equilibrium price. Jacobsen (2013) then simulates a one MPG increase in the CAFE standard. An important feature of these simulations is that the policy change takes over a decade to achieve its full effect, since its effects percolate slowly through the used vehicle market. Used car fuel economy improves by less than new car fuel economy, since inefficient used cars have high prices (because they are scarce) and are scrapped more slowly. The reductions in surplus are initially split nearly evenly between consumers and producers, but after a decade become overwhelmingly incurred by consumers as new vehicles percolate through the used market. The producer surplus impacts are borne entirely by the "Big 3" firms on which the standard binds. Other firms actually benefit because they face weakened competition in the sub-market for large, inefficient vehicles. Overall, the cost of tightening the CAFE standard is \$616 per ton of CO_2 avoided. Taken together, the results in Jacobsen (2013) highlight the value of integrating standard models of equilibrium in vehicle markets with institutional details—such as different firm types and used car markets—to capture important implications of fuel economy policies.

3.2.2 Fuel economy standards and vehicle attributes

Most of the modeling in Jacobsen (2013) permits OEMs to respond to CAFE only by changing vehicle prices and quantities sold. The end of the paper considers an extension in which OEMs can incur additional costs per vehicle to improve fuel economy, using technology cost curves from the engineering literature. Allowing for technology adoption in this way reduces Jacobsen (2013)'s estimated cost of tightening CAFE, from \$616 to \$222 per ton of CO_2 avoided.

This last result from Jacobsen (2013) raises the question of how the full extent of OEMs' possible responses to CAFE, including sales-mix shifting, technology adoption, and trading off fuel economy against other vehicle attributes, affects the policy's outcomes. The potential importance of attribute trade-offs is also highlighted by Knittel (2011), which estimates OEMs' technological frontier that determines the trade-offs between vehicle weight, engine power, and fuel economy. Knittel (2011) finds that this frontier has steadily pushed outward over time. It also finds, however, that during 1980–2004 automakers slid their choices along this frontier so that there were large increases in the size and power of new vehicles sold—for instance, new vehicle horsepower nearly doubled—but only a 6.5% increase in new vehicle fuel economy.

To further understand the incentives behind attribute trade-offs, and how fuel economy choices affect vehicle attributes and welfare outcomes, Klier and Linn (2012) and Whitefoot et al. (2017) augment the Bertrand pricing model from Jacobsen (2013) by endogenizing OEMs' attribute choices. Doing so introduces additional modeling and estimation challenges. First, instead of representing each vehicle model's marginal cost by some constant c_j, these papers need to estimate marginal costs as a function of attributes x_j. This function can in principle be estimated using firms' FOCs, via a strategy like that in Fan (2013)'s study of newspaper markets. However, both Klier and Linn (2012) and Whitefoot et al. (2017) instead use engineering models to estimate the cost function.

A second challenge is that the standard demand estimation strategy of using non-price characteristics as instruments is invalid because these characteristics are chosen simultaneously with the unobserved characteristics in ξ_j. Klier and Linn (2012), for instance, states that "the firm may choose a higher price and greater horsepower for a vehicle that consumers perceive as being 'sporty' or of higher 'quality"'. Whitefoot et al. (2017) addresses this problem by partitioning the set of characteristics into those that can be adjusted in the short-to-medium run and those that are determined by long-run planning schedules, such as vehicle dimensions, powertrain type (e.g. hybrid vs conventional), and drive type. It then uses the latter set of characteristics (of both same-manufacturer and different-manufacturer vehicles) as instruments. Klier and Linn (2012) instead exploits the fact that OEMs often offer models in different vehicle classes (e.g. SUVs versus full-size sedans) that share an engine platform. The instruments in Klier and Linn (2012) are then characteristics of same-manufacturer models from different classes but sharing an engine platform. The idea is that these characteristics will be highly correlated for cost and technology reasons, but the characteristics of vehicles in a different class should not be correlated with the demand shifter ξ_j of the vehicle under consideration.

Whitefoot et al. (2017) uses its model to simulate the effects of replacing the 2006 U.S. fuel economy standards (27.5 mpg for cars and 21.6 mpg for trucks) with the 2014 standard (34.0 mpg for cars and 26.3 mpg for trucks). It finds that OEMs will engage in both sales-mixing and attribute-shifting. For instance, to meet the tightened requirement the average 0 to 60 miles per hour acceleration time will increase by 0.7 seconds. Unlike Jacobsen (2013), Whitefoot et al. (2017) finds that almost all of the additional compliance costs will be borne by consumers, whether or not firms are permitted to adjust vehicle attributes (it is not clear to us why these two papers arrive at different results on CAFE's cost incidence). But the ability to adjust attributes decreases the simulated loss of consumer surplus in 2014 from $12.8 billion to $7.4 billion.

Klier and Linn (2012) examines the effects of a one mile per gallon increase in CAFE standard stringency, finding results that are rather different from those in Whitefoot et al. (2017). Klier and Linn (2012) finds that OEMs constrained by the standard do not respond by decreasing attributes like horsepower or size, but instead adopt technology that improves fuel economy while holding these other attributes constant. In equilibrium, the sales-weighted vehicle horsepower and size still decrease, but only because the sales of large, powerful cars decrease. Furthermore, like Jacobsen (2013), Klier and Linn (2012) finds a more even split in compliance costs between consumers and producers. Producers are found to benefit substantially from the ability to adjust vehicle attributes, while consumers actually wind up slightly worse off.

It is not clear why Whitefoot et al. (2017) and Klier and Linn (2012) arrive at such different conclusions. One possibility is simply that the effects of a marginal change in CAFE standard stringency (as in Klier and Linn (2012)) are different than those of a large change (Whitefoot et al., 2017). The two papers also model different vehicle attributes, use different engineering models as a basis for their cost functions, and use different identification strategies for their demand models. Additional research that explored these differences would be valuable. Moreover, the industry has evolved substantially since these papers were published, so a refresh with up-to-date data would be useful for informing current fuel economy policy.

Finally, Anderson and Sallee (2011) takes a different approach to estimating the cost to firms of CAFE standards. Instead of estimating an equilibrium model of the vehicle sector, as in Jacobsen (2013), Klier and Linn (2012), and Whitefoot et al. (2017), Anderson and Sallee (2011) exploits a "loophole" in fuel economy regulations together with a condition implied by firms' profit maximization. The loophole is that, prior to the Obama revision of fuel economy standards, automakers were able to overstate the fuel economy of vehicles with technology that enabled them to be "flexible-fuel", typically meaning that they could operate on fuel that was 85% ethanol. The cost of this technology is known from engineering estimates to be roughly $100 to $200 per vehicle.

Under a set of assumptions—including that automakers must be at an "interior solution" (i.e., not fully exhausting the loophole) and that consumers do not directly value flex-fuel vehicles—Anderson and Sallee (2011) shows that automakers should equate, on the margin, the cost of flex-fuel technology with the cost of increasing fuel economy. Anderson and Sallee (2011) then argues that these assumptions hold in practice and concludes that the cost to automakers of complying with CAFE is merely $9–$27 per vehicle. This estimate is far lower than that from Jacobsen (2013), potentially because automakers can comply by changing vehicle attributes or investing in technology—channels that are absent from Jacobsen (2013)'s main estimates. Jacobsen (2013) also notes that Anderson and Sallee (2011) draws on data from a period when gasoline prices were high (so that consumers were willing to pay for fuel economy anyway), whereas gasoline prices during most of Jacobsen (2013)'s sample were low.

3.2.3 Attribute-based fuel economy standards

Fuel-economy standards in many countries are "attribute-based", in that the fuel economy target for a particular vehicle model is a function of that model's weight or footprint. Japan and much of Europe use weight-based fuel economy standards, and since 2008 the U.S. CAFE standard has been footprint-based. Ito and Sallee (2018) shows that standards structured in this way give automakers an incentive to up-size their vehicles in order to slacken the fuel economy constraint that they face, leading to a deadweight loss distortion. Ito and Sallee (2018) then examines vehicle characteristics data from Japan, where the fuel economy target for vehicle models discontinuously decreases at discrete vehicle weights. Ito and Sallee (2018) shows that the precise weight of Japanese vehicles is tightly bunched just above these discrete points, providing direct evidence that automakers respond to the incentives created by attribute-based regulation.

3.2.4 Gaming of fuel economy standards

Another way that firms can respond to fuel economy policy is to cheat. Reynaert and Sallee (forthcoming) documents that, following the enactment of stringent fuel economy regulations in Europe in 2007, a large gap developed between automakers' laboratory fuel economy ratings of their vehicles and actual on-road performance, using driver-level panel data from the Netherlands. In 2014, Reynaert and Sallee (forthcoming) documents an average fuel consumption performance gap that exceeds 50%. To provide a welfare interpretation of this gaming of the standard, the paper then develops a model that captures automakers' incentives to game and allows for two possible beliefs on behalf of consumers: that they know the gaming is occurring, or that they are fooled at the time of purchase. Taking the model to the data, Reynaert and Sallee (forthcoming) shows that even when consumers are fooled (and therefore buy cars that do not maximize their ex-post utility), consumers still benefit from gaming because the cost reductions from non-compliance with the true fuel economy standard partially pass through into lower vehicle prices. However, these consumer surplus gains come at the expense of increased CO_2 emissions.

Reynaert (forthcoming) develops a model that incorporates the full spectrum of ways discussed above that automakers may respond to fuel economy standards: changing their sales mix, adjusting vehicle attributes (downsizing), adopting fuel-saving technology, and gaming. The paper focuses its attention on the EU's adoption of a stringent fuel economy standard that phased in over 2007–2015. It begins by following Knittel (2011)'s approach to estimate the technological frontier firms faced during this time period, concluding that firms responded to the standard by decreasing officially-reported emissions without engaging in significant downsizing. As in Reynaert and Sallee (forthcoming), Reynaert (forthcoming) finds that the majority of the officially reported fuel economy improvements are not reflected in real-world driving data. To better understand these effects and evaluate impacts on consumer and producer surplus, Reynaert (forthcoming) then develops a model of the EU auto industry, incorporating consumer preferences, marginal costs, costs of technology adoption, and costs of gaming. The paper concludes that the EU fuel economy standard induced firms to comply through a combination of technology adoption and gaming, with little contribution from sales mix shifting or downsizing of other attributes. The standard reduces both consumer and producer surplus, though by less than if firms were restricted to comply only through sales mix shifting. Relating to Ito and Sallee (2018), Reynaert (forthcoming) also shows that the fact that the standard is weight-based substantially curtails the CO_2 emission reductions achieved by the policy, while redistributing surplus away from French and Italian automakers and towards German automakers.

3.2.5 Lessons learned on fuel economy standards, and paths for future work

The past decade has seen tremendous research progress into firms' responses to fuel economy standards and the implications of these responses for consumer surplus, firms' profits, and GHG emissions. Across settings in Europe, Japan, and the U.S., the papers discussed above offer a useful set of tools for modeling the variety of ways firms might respond to increasingly stringent standards: fleet-mix shifting, trading off fuel economy against vehicles' other attributes, increasing production costs, and gaming. Different papers have arrived at different conclusions regarding the relative importance of these mechanisms, however. Some of these differences might reflect genuine differences in the economic and regulatory environment across jurisdictions studied. For instance, it may be easier for firms to game the standards in Europe than in the U.S. But some of the differences may be related to differences in models and estimation methods, as suggested in our discussion of Klier and Linn (2012) and Whitefoot et al. (2017). Work that reconciled these differences and used more up-to-date data is needed to help inform fuel economy policy-making.

The existing literature has also generally not studied the extent to which fuel economy standards induce automakers to innovate and expand their technological frontier. Such innovation is likely crucial to meeting long-term decarbonization goals, and spurring such innovation is often a stated goal of fuel economy policy. Research that developed models and credible evidence on fuel economy policy-induced innovation would arguably be the single most valuable contribution IO economists could make to this area. We recognize though that such research is difficult due to challenges in measuring innovative activity, let alone its costs and benefits. This topic may be a setting that could benefit by borrowing ideas and tools from IO research on innovation in other sectors, such as Goettler and Gordon (2011)'s study of innovation in computer microprocessor manufacturing.

3.3 Industrial organization and vehicles' emissions of local air pollutants

In comparison to the body of papers studying consumers' demand for fuel economy and automakers' responses to fuel economy standards, the volume of economic research on regulation of conventional exhaust pollutants, such as PM2.5 and NO$_x$, is relatively modest. Two recent papers, however, have shown that ideas and tools from industrial organization can be informative about the consequences of these regulations. Our hope is that additional IO scholars move into this under-researched area.

First, Miravete et al. (2018) raises the possibility that environmental regulation might be motivated by objectives other than pure reduction of emissions. Specifically, the paper considers European Union (EU) fuel taxes and regulation of conventional pollutants (in particular, NO$_x$) and examines how these policies jointly serve as an effective protectionist policy in the context of international trade. The paper begins by pointing out that EU policies have historically favored diesel vehicles over gasoline vehicles in two ways. First, sales taxes on diesel have been substantially less than taxes on gasoline: 32 versus 46 Euro cents per liter on average during 1991–2013. Second, EU regulation of NO$_x$ pollution is less stringent than in the U.S., again favoring diesels because they produce substantially more NO$_x$ as a byproduct of combustion.

To evaluate the effects of these policies, Miravete et al. (2018) develops an equilibrium model of the EU auto industry, with demand and supply specified in the spirit of Berry et al. (1995). The demand model allows consumers' vehicle valuations to depend both on the vehicle's kilometers traveled per Euro (which is also a function of fuel prices at the time of purchase) and on whether the vehicle is gasoline or diesel-powered (to capture preferences for fuel type that are unrelated to fuel costs). The supply model allows firms' cost of producing a diesel vehicle to differ from the production cost of an otherwise identical gasoline-powered vehicle.

Miravete et al. (2018) then uses its estimated model to show that the EU's fuel taxes and emissions regulations together induced substantial adoption of diesel vehicles by EU consumers. First, the demand estimates indicate that consumers value fuel economy and benefit from lower prices that result from firms not having to install NO$_x$ abatement technology on par with what was required in the U.S.. In counterfactual exercises, Miravete et al. (2018) shows that either equalizing fuel taxes across gasoline and diesel or increasing the stringency of NO$_x$ abatement standards to U.S. levels would decrease diesels' market share in the EU by 5–10%.

Miravete et al. (2018) closes by showing that the EU's differentiated fuel taxes and relatively lax NO$_x$ standards led to substantial profits for EU firms, which are specialized (relative to their counterparts in the U.S.) in diesel technology. These results are consistent with these two policies together acting as a non-tariff barrier to competition from foreign vehicle manufacturers. Miravete et al. (2018) estimates that these policies are equivalent, in terms of effects on domestic versus foreign market shares, to a vehicle import tariff on the order of 20%. The paper therefore highlights that environmental regulations can be used as a tool to help protect incumbent firms, in this case to the detriment of local air quality in the EU.

The second paper we discuss also focuses on diesel vehicle emissions in the EU. Alé-Chilet et al. (2021) studies an alleged collusive agreement between BMW, Daimler, and Volkswagen to under-comply with EU NO_x emissions standards by under-sizing their vehicles' diesel exhaust fluid (DEF) tanks. The question underpinning the paper is why these firms found it necessary to jointly agree to under-comply, rather than under-comply unilaterally. There is no reason to think that consumers would value a large DEF tank, and if anything consumers would seem likely to prefer a small tank since larger tanks reduce available trunk space. A potential answer to the question is that the firms perceived that the expected sanction they faced from colluding was less than that from unilaterally violating the regulation, either because the probability of detection was lower (for instance, if the firms agreed not report one another to the regulator) or because the expected penalty conditional on detection would be lower.

Alé-Chilet et al. (2021) quantifies the three automakers' incentive to collude by specifying and estimating a differentiated product demand and Bertrand competition model similar to those discussed in section 3.2 above. The demand specification includes trunk space as a characteristic, and the marginal cost specification includes DEF tank size as a characteristic. The estimates confirm that consumers value trunk space and that DEF tank size is costly, underscoring the point that firms did not have an incentive to over-comply. Moreover, a natural property of Bertrand competition is that the profits of any given firm are an increasing function of the compliance of its competitors. Thus, absent differences in expected compliance penalties, it is difficult to rationalize why the three firms felt compelled to collude rather than unilaterally under-comply. Alé-Chilet et al. (2021) finds that collusion must have reduced the expected penalty by 188–976 million Euros in order to rationalize the collusive agreement.

The analysis in Alé-Chilet et al. (2021) then leads to two further policy-relevant points. First, the collusion led to increases in both producer and consumer surplus, in contrast to conventional price-fixing collusive agreements that increase firms' profits at the expense of consumers. However, these surplus gains were tied to substantial increases in NO_x emissions, which if evaluated using dose-response functions and a value of a statistical life from the literature, outweigh the collusion's private surplus benefits. Second, the paper highlights that in situations where environmental regulation is weak—in this case, the regulator did not monitor on-road emissions, and firms were unable or had agreed not to monitor and report each other—antitrust enforcement can play a role in buttressing enforcement. In this case, the collusion was detected as a consequence of a European Commission investigation triggered by the Volkswagen emissions scandal in the U.S..

3.4 Consumers' fuel search behavior

In this section, we consider consumer search for gasoline stations. We devote time to this topic for a number of reasons. First, a pressing environmental policy question concerns how vehicle markets will transition away from gasoline fuel towards alternative fuels, and especially towards electric vehicles (EVs). As highlighted in particular by Dorsey et al. (2021), studying how consumers make fueling decisions today can help inform what kinds of EV charging station densities and business models for EV charging are likely to be successful. Second, this literature has successfully drawn a close connection between consumers' search behavior and retail markups charged by gasoline stations. These findings are substantially important on their own given the sheer size of this industry, but they are likely also relevant to other settings (for instance, grocery staples) in which prices change frequently and consumers must undertake some costly search to learn about firms' prices.

We begin by discussing the contributions from Chandra and Tappata (2011), which presents a simple model of consumers' search behavior and uses daily station-level price data to test it. The paper begins by pointing out the importance of considering both consumers' search problem and firms' pricing problem when formulating theoretical predictions that one wants to subsequently test. In particular, if consumers undertake costly search in a rational manner, and if firms set markups rationally in response to one another and to consumers' searching, then the sign of the empirical correlation between equilibrium search activity and price dispersion is not clearly predicted from theory.

To be more precise, consider the following simplified version of the model presented in Chandra and Tappata (2011), which is tailored for application to retail gasoline markets.[24] Consumers have unit demand for fuel up to their valuation v. A share λ of consumers are "shoppers" with zero search cost, who will always buy gasoline from the lowest-price seller. The remaining consumers $1-\lambda$ each have a search cost s drawn from a distribution $G(s)$ with bounded, positive support. Paying this search cost reveals all market prices to the consumer (thus, this model best maps to a situation in which consumers use a price information website rather than driving from station to station). Let $\mu \geq \lambda$ denote the overall share of fully-informed consumers, consisting of both the shoppers and those who choose to undertake costly search.

Firms in Chandra and Tappata (2011) all have an identical marginal cost $c \leq v$. The equilibrium conditions are then that each firm's price p must be a best response to all other firms' prices, given μ, and μ must reflect optimal search given firms' prices.

Following Varian (1980), given a search intensity μ the Nash equilibrium for firms' prices will involve mixed strategies, with firms drawing prices from a distribution $F(p)$ with support on $[p^*(c,v,\mu), v]$. The lower bound $p^*(c,v,\mu) \geq c$ exists because firms always have the alternative of setting the monopoly price v and earning profits from the share $1-\mu$ of consumers who do not search and are evenly distributed across firms. $p^*(c,v,\mu)$ is increasing in marginal cost c and decreasing in search intensity μ. Equilibrium price dispersion, measured as the standard deviation of p, is non-monotonic (specifically, reverse U-shaped) in μ: at $\mu = 0$ (no consumers search) all firms set $p = v$, and at $\mu = 1$ (all consumers search) all firms set $p = c$. But for $\mu \in (0,1)$ price dispersion is non-zero.

On the demand side, consumers' marginal benefit from searching increases with price dispersion and is therefore also a reverse U-shaped function of search intensity. The search cost of the marginal searcher is strictly monotonic in search intensity for $\mu > \lambda$, so there is then a unique equilibrium search intensity μ^* that equates the cost and marginal benefit of searching for the marginal searcher.

Chandra and Tappata (2011) then uses this model to make the point that the effect of a change in con-

---

[24] We omit Chandra and Tappata (2011)'s comparative statics regarding the number of firms and do not present here the full analytic functions that characterize equilibrium.

sumers' search costs on equilibrium price dispersion is ambiguous. Clearly, a decrease in the cost of searching will increase consumers' incentive to search, holding price dispersion fixed. However, price dispersion is not fixed in equilibrium, and depending on the initial search intensity, price dispersion can either increase (if the initial search intensity is very low) or decrease (if the initial search intensity is very high) in response to an increase in search intensity. Thus, a decrease in consumers' search costs can potentially increase or decrease equilibrium price dispersion, depending on the initial search intensity.

In contrast, the effect of decreased search costs on the average price level is unambiguous: average prices decrease. The effects of changes to firms' marginal cost c on the equilibrium are also unambiguous. An increase in c will increase the average price level and compress the price distribution, leading to a decrease in search intensity.

Chandra and Tappata (2011) then examines whether these predictions are borne out in data on daily gasoline station prices, using 18 months of data from four U.S. states. The empirical work begins by providing evidence that gasoline stations appear to use mixed strategies. Chandra and Tappata (2011) organizes the data into pairs of nearby stations and finds that "rank reversals" of which station charges the higher price are a frequent occurence. Of course, such reversals could be caused by other factors, including differences in the time-of-day at which stations change their prices. To isolate the mixed strategy mechanism, Chandra and Tappata (2011) compares rank reversals at stations that share a corner versus reversals at stations that do not. The logic behind this test is that search costs are zero for stations that share a corner, but not for more distant stations. Thus, the search model would predict more rank reversals for distant stations than for corner stations. This prediction is borne out in the data, supporting the idea that gasoline stations are using mixed strategies to set prices.

Next, Chandra and Tappata (2011) tests the comparative static that price dispersion should decrease with firms' marginal cost. Using wholesale gasoline prices as the measure of marginal cost, it finds support for this result using both a pooled regression of market-level dispersion on marginal cost, and a regression that isolates within-market variation over time using market fixed effects.

The empirical results in Chandra and Tappata (2011) therefore find support for the predictions implied by a consumer search model. Its tests are indirect though, in the sense that the paper examines equilibrium price dispersion but does not observe search behavior itself. Lewis and Marvel (2011) and Byrne and de Roos (2017) therefore build on Chandra and Tappata (2011) by studying data that directly measure consumers' search activity on online price platforms.

Lewis and Marvel (2011) examines data from GasBuddy.com, a U.S. website that allows users to find station-level gasoline prices. Using data on the website's traffic, the core empirical finding in Lewis and Marvel (2011) is that search increases when overall gasoline prices are rising (e.g., due to an increase in wholesale fuel prices) and decreases when gasoline prices are falling. Lewis and Marvel (2011) then further shows that gasoline stations' respond to these search tendencies in a way that is consistent with the model in Chandra and Tappata (2011). When prices are rising and search activity increases, retail price dispersion and margins decrease. The opposite occurs when prices are falling.

Thus, the firm-level behavior documented in Lewis and Marvel (2011) is consistent with a standard rational search model, but consumer behavior is not. In the Chandra and Tappata (2011) model, consumers should search less, not more when wholesale gasoline prices are rising and price dispersion is falling. The findings in Lewis and Marvel (2011) therefore suggest that consumers' search behavior is influenced by factors such as salience that are not included in standard search models like that used in Chandra and Tappata (2011).

Byrne and de Roos (2017) studies consumers' gasoline search behavior in Perth, Australia. This market is characterized by regular "Edgeworth cycles" (Maskin and Tirole, 1988), wherein each week stations increase their prices substantially on Thursday, gradually undercut one another over the following days, and then jump up again on the following Thursday. Using data from the "Fuelwatch" price discovery site, Byrne and de Roos (2017) finds three facts. First, search activity peaks on Wednesdays, when prices are at their lowest. This fact is consistent with a form of intertemporal search in which consumers are aware that prices will spike the following day. Second, search activity is higher than average on Thursdays (though not as high as on Wednesdays) when prices are at their highest. This result is consistent with rational search models, since price dispersion is large on Thursdays because not all stations initially jump to the same price level. Third, and finally, even conditioned on day of week effects, search levels are higher when price dispersion is higher. The paper does not pin down the precise mechanism behind this result, but it is consistent with the prediction from the Chandra and Tappata (2011) model that changes in firms' costs will lead to a positive correlation between price dispersion and search.

Finally, Dorsey et al. (2021) connects consumers' fuel search behavior to its potential implications for EVs. As we discuss at greater length in section 3.5 below, an important driver for consumers' EV demand is likely to be the availability of EV charging infrastructure. Dorsey et al. (2021) examines one mechanism by which increased EV charger density might influence preferences for EVs: decreased time needed to search for or drive to a charging station. The paper is especially unique because it is the only paper that we are aware of that studies drivers' on-road (rather than online) fueling station search behavior.

Dorsey et al. (2021) takes advantage of a unique setting in Michigan in which the second-by-second behavior of 108 drivers was tracked over six weeks.[25] The data enable the paper to identify refueling stops, and the paper matches these stops to station-level retail gasoline price data.

The empirical exercise at the center of Dorsey et al. (2021) is an analysis of how drivers decide where to refuel. The idea is that drivers often face a trade-off between choosing a station that does not require extra time to reach (i.e., it is directly on the drivers' route) but has a high price, versus a station that is off-route but has a low price. One new descriptive fact that comes from the paper is that drivers typically do not deviate far from their route to buy gas, and nearly 50% of refuelling stops do not involve any deviation at all.

A seemingly natural way to model drivers' decision would be a discrete choice model in which the utility driver i receives from stopping at station j on day t is a function of the price p_{jt} and travel time t_{ijt}:

$$U_{ijt} = -\alpha p_{jt} - \gamma t_{ijt} + \varepsilon_{ijt} \tag{20}$$

The model in equation (20) assumes that drivers are fully informed about prices at other stations. Be-

---

[25] The data on the 108 drivers studied in Dorsey et al. (2021) were originally collected as part of an engineering study of crash-warning technology. All 108 drivers drove identical cars during the study period, and they had to purchase fuel using their own funds.

cause that assumption is unlikely to hold, Dorsey et al. (2021) instead assumes that consumers respond to a perceived price $\tilde{p}_{jt}$ that is a weighted average of the real time price p_{jt} and the average price $\bar{p}_j$ at station j. This model nests the full information model from equation (20) if the weight on the real time price is 1. This weight is identified in the model by the extent to which drivers' respond to fluctuations in stations' prices over time versus variation in average prices across stations.

The estimates in Dorsey et al. (2021) yield several useful results. First, the estimated price weights indicate that drivers respond substantially more to average price variation across stations than to real time prices. Second, the estimated ratio $\hat{\gamma}/\hat{\alpha}$ represents drivers' value of time while searching for fueling stations.[26] This estimated value of time is about $25 per hour, which is more than double the value of time that would come from a conventional estimate based on one-half of average gross wages (Small, 2012). This estimate—which is closely related to the initial descriptive fact that drivers tend not to go far out of the way to buy cheap gasoline—implies that consumers experience a strong disutility from spending extra time on the road.

Dorsey et al. (2021) closes by noting that this disamenity from extending the duration of a trip poses a challenge to EV adoption: drivers will demand a dense EV charger network. The paper makes this point precise using a simple, but illustrative, Salop (1979) spatial competition model, in which the surplus-maximizing number of charging stations depends on stations' fixed cost, the number of EVs, and the travel time cost. In this model, using the Dorsey et al. (2021) value of time estimate rather than the conventional one implies a roughly 50% increase in the number of charging stations needed to maximize social surplus.

We see Dorsey et al. (2021) as an important first step in using observations of drivers' current fueling behavior to draw implications about future EV and EV charging markets. If anything, we suspect that Dorsey et al. (2021) likely underestimates the overall charging-related challenge to EV adoption, since at least with current technology EV charging requires substantially more refuelling time than does a conventional gasoline station. Future research that further probed the extent to which drivers value minimizing time spent searching for fuel—and time spent fueling once fuel is found—would be valuable in informing future business models for EV charging, including whether charging will be dominated by home and work chargers or whether independent chargers are a viable business model.

3.5 Markets for EVs and EV charging stations

Recent decreases in the production costs of EVs, combined with improvements in EVs' quality and battery range, suggest that EVs may transition from a niche product to the mainstream by the end of this decade. A key challenge, however, to substantially increasing EV penetration is that EVs require the development of a charging network. Because the value of an EV is increasing in the availability of charging, and because the value of a charging station is increasing in the number of circulating EVs, EVs and charging stations should be thought of as indirect network goods. The presence of network externalities then raises at least two questions that have been examined in recent work: (1) given many jurisdictions' policy objective to increase EV penetration, to what extent should governments subsidize EVs versus subsidize charging stations; and (2) what are the economics of charging standards, and should different manufacturers be required to adhere to

---

[26] Dorsey et al. (2021) also uses information on the average amount of fuel purchased per stop to arrive at a value of time estimate.

an interoperability standard? These questions are also closely related to broader questions about the economics of network goods in a variety of industries, and whether governments should enforce interoperability standards between different service providers. The models and estimations strategies used in these papers could potentially serve as jumping-off points for work studying indirect network effects in other settings.

3.5.1 Indirect network effects and EV incentive policies

In EV markets, indirect network effects arise because consumers' demand for EVs depends on the availability of charging stations, and firms' incentive to invest in charging stations depends on the number of EVs in circulation. For a government that seeks to increase EV adoption via subsidies, a question then arises of whether subsidies are better spent on EVs themselves, on charging stations, or on some of both.

This question is addressed in Li et al. (2017) and Springel (forthcoming). Li et al. (2017) studies the EV market in the U.S. from 2011–2013, and Springel (forthcoming) studies the EV market in Norway from 2010–2015. The models studied in these two papers are similar, and they are used to both contribute to our understanding of EV incentives while also providing a research design for estimating and simulating models with indirect network effects. The policy-relevant bottom lines of these papers are also remarkably similar: at the overall subsidy levels in force during the periods studied, subsidies for charging stations were roughly twice as effective at inducing EV adoption than were subsidies for EVs themselves.

We focus our discussion on Springel (forthcoming), given the relative maturity of Norway's EV market. Norway arguably has the most aggressive EV subsidization policy in the world, and EVs have recently accounted for the majority of new vehicle sales there. Norway subsidizes EVs by exempting them from the country's large registration and value-added taxes, and it also provides financial support for the installation of charging stations.

Li et al. (2017) and Springel (forthcoming) ask how counterfactual subsidies that change the amount of government expenditures on EVs versus charging stations would affect EV adoption rates. Answering this question is challenging because it requires understanding how consumers value both charging networks and EVs themselves. Springel (forthcoming) begins by specifying a demand model that incorporates network externalities into EV buyers' utility functions. Specifically, demand comes from a random utility discrete choice model, where the value each consumer i obtains from vehicle model j in market m is given by:

$$u_{ijm} = \beta_i^N \log N_{jm} - \alpha_i p_{jm} + \beta_i^k x_{jm}^k + \xi_{jm} + \varepsilon_{ijm} \tag{21}$$

As usual, equation (21) includes the vehicle price p_{jm}, a k-dimensional vector of observed vehicle characteristics x_{jm}^k, an unobserved quality term ξ_{jm}, and an idiosyncratic model preference ε_{ijm}. The novel term in this model is N_{jm}, which (if model j is an EV) denotes the number of available charging stations in market m.[27] β_i^N then denotes consumer i's preference for charging infrastructure, and the log functional form builds in a declining marginal utility of charging network density.

The inclusion of N_{jm} in the demand model leads to an identification challenge, on top of the usual challenge of price endogeneity: if consumers in market m have an especially strong preference for EVs, that

---

[27] In Springel (forthcoming), all EVs can charge at any charging station.

will lead firms to install more chargers in that market, leading to upward-biased estimates of β_i^N. Springel (forthcoming) addresses this problem by using charging station subsidies, which vary substantially over time, as an instrument for N_{jm}.[28] Ideally, the functional form by which N_{jm} enters equation (21) would allow for multiple parameters, so that the level and decline of marginal utility of station density could vary independently, but credibly estimating such a model was likely infeasible given limited instruments and data.

On the supply side, Springel (forthcoming) focuses on the incentives of charging station owners. Vehicle manufacturers are not explicitly modeled, so vehicle supply is effectively assumed to be perfectly elastic (and all EV subsidies therefore fully pass through to consumers). Because Springel (forthcoming) does not observe charging prices and markups, the paper specifies a reduced-form charging station entry model in which $\log N_{jm}$ is an affine function of $\log Q_{jm}$, the total number of electric vehicles registered in the market, along with controls for charging station incentives. As with the demand model, the presence of network externalities creates an identification challenge, and the paper therefore instruments for $\log Q_{jm}$ using the density of gasoline fueling stations.[29] The logic of this instrument is that fueling stations affect EV sales by affecting the relative fuel cost of gasoline vehicles versus EVs.

Both the demand and supply estimates in Springel (forthcoming) are consistent with strong network effects, which then lead to a series of positive and normative implications. First, an interesting result is that when network effects are accounted for, the cross-price elasticities between EV models are often negative, rather than positive as one would normally expect for substitute goods. This complementarity arises through feedback effects that arise from charging station entry.

Second, the sensitivity of EV demand to charging station density is sufficiently strong that charging station subsidies are actually twice as effective, per Norwegian kroner spent and at current funding levels, at inducing EV adoption than EV subsidies.[30] However, the paper also finds diminishing returns to subsidizing charging stations, so that the gains from additional subsidies to charging stations would not persist indefinitely.

3.5.2 Compatibility between charging networks

Another question for EV deployment concerns the compatibility of charging networks, especially for high-speed (level 3) chargers that can fully charge an EV battery in 30 minutes. Currently, different EV manufacturers have developed at least three distinct, incompatible standards for this charging technology. As a consequence, a Nissan Leaf EV cannot, for example, take advantage of a level 3 charging station designed for a Tesla.

Should the government mandate interoperability? Li (2019) presents a model showing that, despite the obvious appeal to EV consumers of being able to charge at any level 3 station, there is a trade-off

---

[28] For its U.S. context, Li et al. (2017) uses a Bartik-style instrument: an interaction between time-series variation in national public charger investment with spatial variation in the number of grocery stores (which are common sites for public EV chargers).

[29] Li et al. (2017) uses lagged gasoline prices as the instrumental variable for the stock of EVs when estimating its charging station supply equation.

[30] Li et al. (2017)'s intuition for its similar result is that early adopters of EVs are not especially price-sensitive, and they instead greatly value the ability to easily charge their vehicles.

here because interoperability requirements will depress EV manufacturers' incentives to invest in charging networks.

In Li (2019)'s model, EV manufacturers act as Bertrand oligopolists. These firms, in addition to the usual vehicle pricing decision, must also decide how much to invest in their level 3 charging network. Consumer demand is similar to that given in equation (21) from Springel (forthcoming), but in Li (2019) demand depends on both the overall density of level 1 and 2 chargers (which can charge any EV) and the number of level 3 chargers associated with the specific brand.[31] As in Springel (forthcoming), charging station density enters via a log functional form so that there are diminishing returns to adding additional chargers to the network. The fact that demand for a particular EV model j depends on level 3 chargers installed only by the manufacturer of j gives that manufacturer an incentive to install chargers. In contrast, in a model in which interoperability mandates mean that EV owners can use any level 3 charger, manufacturers have an incentive to free-ride on other manufacturers' installations, and the total number of chargers is sub-optimal.

Li (2019) estimates its model by taking advantage of variation in charging station build-out caused by the 2009 American Recovery and Reinvestment Act, which allocated funds for EV chargers across counties in ways that the paper argues are plausibly orthogonal to EV demand. In addition, identification of consumers' price sensitivity makes use of both the standard Berry et al. (1995) instruments and variation in federal and state EV subsidies.

The main counterfactual in Li (2019) then studies market outcomes under an interoperability standard. Despite finding that such a standard would decrease charging station investment by about 3%, the benefits to consumers from interoperability would increase EV sales by 21%. Thus, the paper concludes that the direct benefits from an interoperability standard outweigh its impact on firms' incentive to reduce charging station investment.

3.5.3 Paths for future research on EVs and EV charging

Li et al. (2017), Springel (forthcoming), and Li (2019) together highlight the importance of indirect network effects between EVs and EV charging stations for driving EV adoption. These papers represent important first steps, but much more is needed to understand the rapidly growing market for EVs. Especially as EV market penetration increases in the coming years, we see promise in the following research topics:

- **Fuel prices and EV adoption.** How does variation in the relative prices of electricity and gasoline affect EV adoption? Answers to this question will inform policies such as gasoline taxes or subsidies for purchases of electricity provided through EV chargers.

- **EV charging networks.** To what extent will EV drivers' charging needs be met by home charging, workplace charging, or third party charging stations? How might drivers substitute across these options? Will competition across workplace and third party charging stations resemble current models for retail gasoline station competition, or will the organization of this industry be substantially different?

---

[31] Li (2019)'s model of how demand depends on the charging network is spatially rich, as it accounts both for the density of the charging network around the consumer's home as well as the ability of the network to facilitate inter-city travel.

- **Innovation.** How do EV policies affect innovation in EV technology, especially battery capacity and charging time? How are EV innovation incentives affected by oligopolistic competition among automakers?

4 Electricity markets

Electricity is an input to virtually any economic activity. Electricity is also a unique commodity when it comes to its economic characteristics. While it is a homogeneous good, it is also, differently than in other industries, concurrently produced with a wide range of technologies (e.g., coal, natural gas, hydro power, wind and solar, etc.). Among other unique features, electricity has been extraordinarily difficult to store economically (and still is today, although this is changing). The difficulty of storing electricity, coupled with generally quite inelastic demand in the short- and medium-run, leads to substantial price fluctuations and seasonality, as well as moments of scarcity in which market power can be substantial. The presence of capacity constraints and difficulties in storage is in fact behind the co-existence of such a wide range of generation technologies, depending on their level of utilization and flexibility. The delivery of electricity is also distinct and mainly done via the electric grid, whose operation needs to be centralized and coordinated.

The process of coordinating the second-by-second match between demand and supply while ensuring that the electricity grid can deliver power is done by centralized authorities. Electricity markets have emerged in the last 30 years to help make this process more transparent and cost effective, leading to a large body of research aimed at understanding the efficiency implications of these changes, as well as the potential pitfalls, such as an enhanced ability for firms to exercise market power. These are core questions at the heart of the study of industrial organization. The fact that these markets were traditionally regulated (and some of them (or parts of them) still are) also makes it particularly suited to study canonical questions in the economics of regulation literature. Additionally, the existence of highly granular data with plausible shifters both on the demand and the supply side make electricity markets particularly attractive for IO economists.

Electricity markets are currently again in the midst of enormous transformation, with a rapid decarbonization of electricity generation and an acceleration of the electrification of many major activities, such as transportation. Efficiently divesting from a carbon-intensive to a carbon-free grid will be essential to minimize the costs of the energy transition. Therefore, the analysis of these markets can have crucial impacts to the overall economy as well as the environment and our future prospects. The field of industrial organization can contribute with its tools to analyze market design and market structure to ensure a more efficient and equitable transition.

It would be impossible to cover all of the papers relevant to the study of electricity markets. This section focuses on some of the papers that are most relevant to IO economists while trying to highlight unifying themes across them.

4.1 The restructuring of electricity markets

Since the 1990s, electricity markets have undergone large transformations in market organization with the emergence of liberalized electricity markets in many countries of the world. Traditionally, electricity markets had been operated under rate-of-return regulation and often in the form of a vertically-integrated natural monopoly in all of its segments: generation, transmission and distribution, and retailing. However, several markets transitioned away from this regulated vertical structure and opened up the generation and retail segments of the electricity sector to competition. This tremendous shock to how electricity markets are organized has been a source of a large amount of IO research studying the performance of wholesale and retail markets under imperfect competition, optimal market design, and restructuring's impacts on productivity and efficiency, among other topics.

In the United States, the process of restructuring in wholesale electricity markets has been tumultuous and slow, which has also been an interesting source of variation to researchers. The restructuring efforts started in markets such as California, New England, the Pennsylvania-Jersey-Maryland market (PJM), but the restructuring of other markets was put on hold after the market crash of the California electricity markets in the 2000s, leading to varied forms of market structure and competition. Since then, several regions have transitioned into formal organized markets, but large areas still remain under cost-of-service regulation for the generation segment.

Borenstein and Bushnell (2015) provides a retrospective of the liberalization process in the United States, highlighting its successes and failures. Bushnell et al. (2017) reviews the findings of the literature in more detail than what we can provide here.[32]

4.1.1 Aggregate impacts of restructuring

From an industrial organization perspective, the halt of the restructuring process in the United States provides a unique opportunity to study complex markets that are quite homogeneous in their main purpose (the provision of electricity through the grid) but are organized with a wide range of regulatory tools. A literature has emerged that treats the presence of these organizational forms as a quasi-natural experiment, comparing markets before and after restructuring to similar ones that did not undergo this transformation.[33] Even if it is difficult to ensure a purely apples-to-apples comparison between heterogeneous regions, this large variation in regulatory form is distinct relative to many other markets and worth studying.

Cicala (2020) evaluates the overall impacts of the transition to market-based wholesale electricity operations using the staggered transition to markets from 1999-2012. It assembles a detailed dataset of hourly operations of virtually all power control areas (PCAs) in the United States. The data include hourly demand, hourly generation of a large share of power plants, and engineering estimates of the costs of operating power plants. The paper presents a decomposition of the hourly costs of electricity generation at the PCA level. It compares the actual costs, calculated with the generation and engineering cost data, to the costs that result from minimizing production costs, holding the observed quantity fixed. This is an easy calculation because, by holding quantity fixed, one does not need to worry about how energy flows between PCAs. Additionally, it computes the surplus that arises from trade by comparing the additional opportunity costs of producing at autarkic quantities when compared the observed ones. Because the counterfactual costs of autarkic

---

[32] See also Kwoka (2008) for an earlier assessment and review of the impacts of restructuring in the US electricity sector.

[33] It is important to note that this regulatory transition is not unique to the United States. Other papers have used a similar diff-in-diff strategy to study the impacts of increased restructuring in other regions. See for example Malik et al. (2015), which finds that restructuring in India had limited impacts on productivity of state-owned plants.

quantities are not observed, the paper's method uses the idealized curves instead.

The paper compares these two measures of performance before and after restructuring using a differences-in-differences strategy. It finds that the electricity markets that transitioned to centralized markets had large gains from trade when compared to other regions, consistent with previous papers that had documented large increases in trade after the restructuring of their operations (Mansur and White, 2012). The regions that joined market-based wholesale electricity markets benefited from an increase in gains from trade of 55%. Departures from the idealized merit order curve within a market decrease by 16%.

4.1.2 Plant-level evidence of the impacts of restructuring

Cicala (2020) offers a comprehensive data set of all markets in the United States, but it does not quantify the mechanism behind the improved efficiency, other than the increases in trade. Other papers have explored the mechanisms more directly using plant-level data. Fabrizio et al. (2007) tests the use of labor and other inputs into the production process by using data from utility-owned power plants using a production function approach that models energy inputs as Leontief to motivate their reduced-form regression analysis. The paper finds evidence that plants that are expected to be restructured reduce their labor expenses by 3% and their non-fuel expenses by 5-10%. One limitation, however, is that the plant-level information is limited once power plants are restructured, as the regulation no longer requires the reporting of labor and other materials.

Davis and Wolfram (2012) uses a similar diff-in-diff strategy to quantify the efficiency impacts from restructuring in the operation of nuclear power plants. It documents significant increases in the utilization of nuclear power plants after deregulation, with a central estimate of 10% increased operating performance, which does not seem to come at the expense of decreased safety (Hausman, 2014).

Whereas previous papers examine the impacts of regulation from an ex-post econometric point of view, these papers do not provide structural modelling that can help quantify the impacts of counterfactual regulatory regimes. In light of all the changes that are affecting electricity markets, we aspect the differing performance between regulated and restructured markets to reemerge as an important area of study. IO models of investment under different regulatory benchmarks could help understand the incentives provided by different regulatory regimes. The work in Eisenberg (2020) examining the impacts of regulatory wedges in the performance of the Chinese electricity market is a great step in that direction.

4.1.3 Natural monopoly regulation in distribution

While the generation and retail sector have been restructured in several markets, the distribution of electricity remains a natural monopoly. Therefore, it is subject to the challenges of providing incentives for efficient operations and the provision of quality while regulating prices.

Lim and Yurukoglu (2018) studies the provision of electric distribution services with a comprehensive dataset of measures of quality of service and rates for most of the United States service territories. The paper studies the political economy of the rate-setting process and the potential commitment and moral hazard issues involved. Regulators might be tempted to change their regulation (ratcheting effect) while distribution companies might want to under-provide effort. Regulators have two tools to manage the utility:

provide an incentive to capital via rate-of-return, and provide penalties for the under-provision of quality (e.g., based on standard measures of interruptions of service, which is their data on provision of quality).

The paper shows that the regulator will favor a high rate-of-return and low penalties for the under-provision of quality in environments in which it weighs the utility of the firm more than consumer surplus. The paper then builds a structural model that estimates regulators' preferences as well as the costs of quality provision.[34] In the model, preferences are proxied for using political preferences, and more conservative territories end up favoring higher investment and lower quality. The higher investment is not necessarily detrimental for welfare in the presence of commitment problems, but these preferences also lead to higher costs to consumers and lower quality of service.

Lim and Yurukoglu (2018) presents a stylized model with limited heterogeneity, only via the political orientation of different service territories. The monopoly model is also highly stylized. Given the need to modernize the distribution grid and incorporate new innovations, such as distributed generation and voltage control, understanding the microeconomic underpinning of providing incentives for an efficient transition seems an area of important of study.

Mahadevan (2021) also studies the political economy of electricity distribution, examining India rather than the U.S. Using administrative billing data, Mahadevan (2021) finds evidence that an electric utility substantially under-bills regions that supported the ruling party in a recent election. This under-billing occurs despite the fact that these regions, if anything, appear to consume more electricity than other regions, as measured using satellite nighttime lights. The under-billing result is supported by forensic data work revealing that bills in these regions are excessively likely to be round numbers and violate Benford's Law. The paper closes by estimating consumers' demand (accounting for the fact that some of the billing data are manipulated) and showing that the increase in consumer surplus created by the billing manipulation is outweighed by the decrease in producer surplus (which must ultimately be covered by ratepayers or public funds).

Finally, the distribution of natural gas, also a natural monopoly, will also undergo dramatic changes given the need to decarbonize our economies. In particular, the provision of natural gas via the distribution network is expected to decline, creating a stranded asset problem. More work should go into understanding how this process will unfold, and how to properly incentivize and pay for such exit, a topic that has started to be recently explored in Davis and Hausman (2021).

4.2 Market power in wholesale electricity markets

Several papers have examined the performance of wholesale electricity markets with a focus on the potential exercise of market power. Before getting into the specific papers and contributions, it is useful to consider a simple framework of optimal behavior by a strategic firm participating in the electricity market.

Consider an electricity market at time t with demand given by $D_t(p)$. In the electricity market, several firms $j = 1, \ldots, J$ are competing in supply curves: a schedule specifying their willingness to produce (quantity) at different price levels, $S_{jt}(p)$. One can define the residual demand, left to a given firm i, as

---

[34] For the reader interested in these topics, papers discussed in section 4.5.2, 5.2.2, and section 5.3.1 are closely related to this strand of the literature.

the demand left to that firm at a price p, after taking into account the willingness to produce by the other participants, i.e., $RD_{it}(p) = D_t(p) - \sum_{j \notin i} S_{jt(p)}$. It is crucial to consider the impact of the other firms on the demand left to a given firm, as otherwise the demand function tends to be inelastic, leading to unrealistic predictions about the incentives to exercise market power.

Notice that the presence of a residual demand curve is not unique to electricity markets. One could envision that firms have an implicit supply function of their willingness to supply at different prices for other goods. The key in the electricity market context is that firms explicitly compete in this way in the market. Furthermore, such curves are often observed in the data, and therefore one can be much more explicit about the presence of such objects. In the absence of data on supply curves, bounding assumptions can be made, typically ranging from Bertrand competition to Cournot as suggested by Supply Function Equilibria (SFE) models (Klemperer and Meyer, 1989).

Once the residual demand for a given firm has been defined, the individual problem of the firm, taking the behavior of others *as given*, can be stated as,

$$\max_{S_i(p)} E\left[\, p \times RD_{it}(p, \epsilon) - C\big(S_i(p)\big) \,\right], \text{ s.t. } S_i(p) \leq S_i(p') \;\text{ for }\; p < p', \tag{22}$$

where the introduction of ϵ reflects that the firm might be uncertain about the exact shape or level of the residual demand, and the constraint states that firms need to be willing to supply weakly more quantity at higher prices, as often enforced in electricity auctions.

Abstracting away for now from uncertainty and monotonicity constraints, the program in (22) becomes equivalent to the firm choosing the optimal price to set along the residual demand curve that maximizes net revenue.[35] Assuming that the shape of the residual demand satisfies regularity conditions, the optimization problem leads to the following first-order condition for a profit-maximizing firm:

$$p = C' - \frac{RD_{it}(p)}{RD'_{it}(p)}. \tag{23}$$

This expression is equivalent to the usual Lerner condition in introductory monopoly pricing, $\frac{p-C'}{p} = -\frac{1}{\eta}$, where η is the elasticity of the *residual* demand in this case: the more elastic the residual demand, the lower the markup. The expression also highlights that, holding the slope of residual demand fixed, firms will tend to exercise more market power when they are producing larger quantities; i.e., a high RD_{it}, which in equilibrium equals the quantity produced by the firm, Q_{it}.

4.2.1 Estimating market power

Expression (23) can be easily brought to the data, aided by the quality and availability of cost data in electricity markets. As a consequence, there has been a flourishing IO literature examining markups in the context of electricity markets as well as extending this basic framework.

Wolfram (1999) empirically examines competition in electricity markets in the context of the British electricity spot market, which Green and Newbery (1992) had numerically explored earlier in ex-ante sim-

---

[35] A more explicit derivation based on the bidding strategy of firms will be covered below.

ulations using a similar framework. Wolfram (1999) tests conduct parameters using the above derived first-order condition by directly constructing an analogue to the Lerner index. By using data on prices and marginal costs, she can construct the right-hand side of the equation. By making assumption on the elasticity of demand, one can then test if markups are consistent with the strategic model assumed in (22), which would imply that markups times the elasticity (in absolute terms) should be equal to one. Because Wolfram does not observe the strategies of the firms (or their individual output), she bounds the elasticity of demand using results in the SFE literature (Klemperer and Meyer, 1989; Green and Newbery, 1992). Results suggest that firms are far from exercising market power a la Cournot (fixed quantity strategies) and also for SFE models that imply more competition. When compared to simulations (Green and Newbery, 1992), the estimated markups also fall below those predicted ex-ante. Markups are estimated to be between 20 and 25% during the period of study. She conjectures that such "low" markups–relative to the potential for market power–may be explained by contracting forward positions limiting the inframarginal quantity (see below), regulatory threats, or the threat of new entry.

Several papers have estimated markups using data from other electricity markets. Borenstein et al. (2002) (BBW, henceforth) use data from the California electricity market to examine the crisis suffered in 2000-2001, during which prices skyrocketed. The identification strategy in BBW is to consider market power as the residual explanatory factor behind the increase of electricity prices. There were several shocks affecting the California market during the period of study (1998-2000), such as a severe drought, increased NO_x prices, and increased natural gas costs. Therefore, it is important to account for such cost shifters. BBW calculate markups from market power by comparing observed prices to the prices that would be consistent with a first-best scenario in which firms offer their output at their marginal cost and water is allocated efficiently. They use Monte Carlo methods to simulate potential power plant unavailabilities and allocate the limited water resources to those hours in which they are most valuable. They also estimate the availability of imports (the imports supply curve), which are substantial in California (around 25% of consumption). This procedure provides an estimate of $P^{competitive}$ that they can they use to compute the Lerner index:

$$Lerner_t = \frac{P_t^{observed} - P_t^{competitive}}{P_t^{observed}}.$$

They document that the Lerner index was very modest during 1998 and 1999, but increased dramatically in 2000 as a result of the scarcity conditions in the market. A large part of this increase is explained by increases along the supply curve, moving to areas in which market power is more concerning.

Finally, the paper presents a decomposition of the increased costs to consumers as a function of increased costs, increased competitive rents, and increased market power rents. Given the convex shape of the electricity supply curve, competitive rents can increase substantially when moving along the supply curve, as firms with low cost still get paid the marginal price in the market. Yet though competitive rents can explain part of consumers' cost increase, BBW find that around 59% of increased costs between the summer of 1999 and the summer of 2000 can be attributed to market power.

Given the large increase in profits and prices documented in Borenstein et al. (2002), what can we say about firm conduct? Were firms exercising market power according to a Cournot model during the sample?

Or did they coordinate or collude to sustain even larger profits? Unfortunately, the analysis Borenstein et al. (2002) offers little explicit guidance in answering this question, as markups are estimated directly from the data without a strategic model of competition. To inform this question, Puller (2007) estimates conduct parameters more explicitly in the California electricity market using equation (23). To examine this question, the paper tests whether firm behavior is consistent with Cournot pricing. To do so, Puller estimates the markup equation taking into account the elasticity of the fringe suppliers, which is estimated from the data using forecasted demand as an instrument, as well as the presence of capacity constraints, which might affect marginal costs. He finds that in the years before the crisis, the model is consistent with Cournot behavior. During June-November 2000, the model suggests firms exercise market power above the Cournot levels, but still far from a perfectly collusive equilibrium.

The market structure in California led to large increases in prices during the 2000 and 2001, but the crisis was not concurrent in other restructured markets, such as the New England and PJM markets. Natural gas prices also spiked in the Eastern states, but the effects on price increases were more modest. What are some of the factors explaining such differences? Bushnell et al. (2008) (BMS, henceforth) studies the importance of forward positions to explain this empirical pattern. Forward positions refer to the quantity that the firm has already committed at a pre-established price; e.g., via a physical contract, a financial contract for differences, or a selling position due to vertical integration at regulated (sticky) retail prices. The key is that the price of such forwards should be fixed and not depend on the day-ahead electricity market outcomes.

In such a case, the FOC in (23) becomes:

$$p_t = C'_{it} - \frac{RD_{it}(p) - \theta_{it}}{RD'(p)_{it}}, \tag{24}$$

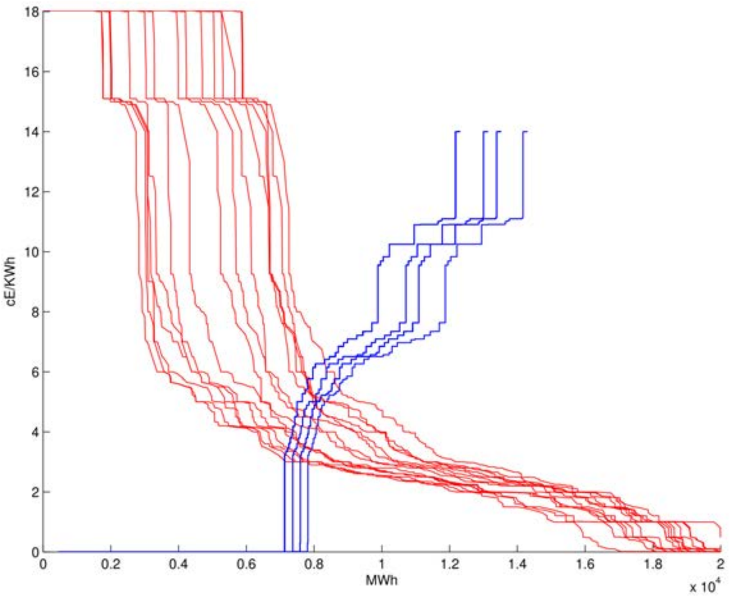
where θ_{it} represents the forward position of firm i at period t. Because this quantity is already committed at a pre-established price, the effect of forward positions is to reduce the infra-marginal quantity of the firm $(RD_{it} - \theta_{it})$ and, therefore, its incentives to exercise market power (Allaz and Vila, 1993). If all output is contracted in advance at a fixed price, the firm will have limited incentives to exercise market power. If the firm is a net buyer in the market $(RD_{it} < \theta_{it})$, the firm might even be willing to exercise monopsony power.

BMS puts together data from three markets (California, PJM, and New England) in which forward contracts are either observed via vertical agreements or zero (as in California). BMS then tests the predictions from three models: a competitive model á la BBW, a Cournot model that ignores the presence of forwards, and a Cournot model that includes forwards in the first-order condition. Its main finding is that a model with forward contracting can rationalize quite well the observed patterns in the data. The competitive one under-predicts prices quite systematically and the Cournot one without vertical agreements leads to an overstatement of market power.[36] The paper is useful to highlight that the absence of forward contracting can in part explain the California crisis.

---

[36] The Cournot assumption contributes in part to the massive market power predicted in the absence of forward contracts, as the strategic firms are assumed to submit inelastic supply curves in the model (quantities), and therefore do not contribute to making the residual demand more elastic.

Figure 2: An example of supply and residual demand draws observed in electricity markets



Note: Electricity markets provide a unique setting in which the supply curve and the residual demand that firms face is observed. *Source:* Authors' construction based on data from the Spanish electricity market. See Reguant (2014).

4.2.2 Taking electricity auctions to heart

The papers we have discussed thus far are not very explicit about the bidding process in the electricity markets they study and do not use bidding data directly. However, electricity markets are typically formally organized as multi-unit uniform auctions. The bidding process and the strategy space in electricity markets are well known and often observed in the data. Indeed, these rich data are a unique feature of electricity markets: the researchers get to observe the supply curves of firms, not just their realized supply in equilibrium, thanks to detailed electricity auctions data. Going back to the maximization problem of the firm in (22), the availability of auction data implies that the ex-post residual demand faced by the firm can be observed. Figure 2 shows an example of bidding data of one firm as well as the residual demand that the firm faces.

Early papers in the literature such as Wolfram (1998) explore how to model the firm's first-order condition as a multi-unit auction, by representing their maximization problem as a function of the distribution of opponent's bids, in a private value framework. The key to understand bidding behavior is to model the trade-off between increasing the price for all accepted units if a bid is accepted, due to the uniform format of the auction, and reducing the probability of having that offer accepted. In equilibrium, firms submit bids with markups above cost that are consistent with the trade-off between these two forces.

Hortaçsu and Puller (2008) (HP, henceforth) shows that the first-order conditions for each bid implied by the multi-unit auction framework can be analogous to expression (24). In particular, if uncertainty only leads to parallel shifts of the residual demand, so that the slope at a given price p remains constant, it is sufficient to know the slope of the residual demand curve at each price to determine the ex-ante optimal supply curve.[37] These ex-ante optimal bids will also be ex-post optimal under such conditions. For the purposes of this Handbook, and given the close link to equation (24) in uniform multi-unit auctions, we will focus on the additional value of bidding data when studying these markets rather than on the fine details between alternative auction models.[38]

How does the presence of bidding data aid the identification of fundamentals in this setting? In electricity markets, the three main potentially unknown factors are typically power plants' costs, the forward position of the firm, and firm behavior (i.e., model of competition). In the traditional IO literature, oftentimes firm behavior needs to be assumed as a source of identification, but this is not necessarily the case in the electricity context, in which firms' supply curves and costs are often observed. Electricity auctions papers differ in which fundamentals are taken from the data, which assumptions are imposed, and which objects are estimated.

Wolak (2000) and Wolak (2003) use a framework of optimal bidding under the assumption of profit maximization to estimate costs. He proposes to smooth out the residual demand curve in order to estimate its slope, $RD'_{it}(p)$, via Kernel smoothing techniques. In his applications, forward contracts are observed for one of the firms, and a rule-of thumb assumption is used for others.

The Kernel representation of the residual demand allows for taking the derivative of the residual demand with respect to a power plant's bid in a simpler fashion than if the lumpiness of step functions where to be considered.[39] In the context of electricity market auctions, in which steps are typically small and the uncertainty around prices is relatively narrow, the probability of setting the price can be difficult to quantify. Kernel smoothing is a practical way of circumventing both of these computational problems. One can express the first-order condition as a function of the bid, where the bid has an indirect impact on the equilibrium price $p(b)$.

With knowledge of the forward position and under expected profit maximization, the firm's first-order condition for a given bid accounting for uncertainty becomes:

$$\sum_{d}\left((p_d - C')\frac{\partial RD_d}{\partial p_d} + RD_d - \theta_{it}\right)\frac{\partial p_d}{\partial b_{ijkt}} = 0, \tag{25}$$

where d represents a given draw in the empirical distribution of the residual demand faced by the firm and b_{ijkt} is the offer made for a step k of plant j. One can estimate the model via GMM (e.g., Wolak (2007)). The equations can also be estimated via a weighted regression approach, in which the weights are given by the a measure of how close the bid is to setting the price (Reguant, 2014). In the limit, when one considers the impact of the bid only when the bid is exactly marginal ($b = p$), this is equivalent to estimating the first-order conditions in (24) for those bids that are exactly marginal.

Hortaçsu and Puller (2008) (HP, henceforth) highlights that, in the presence of bidding data and observed

---

[37] Brown and Eckert (2021) shows the required conditions for this result to hold need to be stronger if one wants to ensure that the supply curve is weakly increasing, a common restriction in these markets.

[38] The interested reader can refer to the auctions chapter in this Handbook for a more detailed derivation of the first-order conditions with bidding data.

[39] See the Handbook chapter on auctions for a detailed treatment of the step-function nature of supply and demand curves in multi-unit auctions.

costs, one can infer the forward position of firms, θ_{it}, by examining the point at which the supply curve and the cost curve cross. That is, in equation (24) the point at which $p = C'$ is the point at which θ_{it} is equal to the offered quantity. This inference allows one to test for behavioral assumptions, instead of assuming profit maximization. Whereas this identification strategy requires firms to place no markup at the point at which their offered quantity matches their contract, HP can still test if the behavior of firms is consistent with profit maximization. HP shows that large firms behave closer to a profit-maximizing bidder, but small firms submit supply bids that are too inelastic, leaving substantial profits on the table and significantly increasing inefficiencies in the market, which are even larger than those induced by market power. Hortaçsu et al. (2019) revisits the results in HP and show that a k-level hierarchical model in which low-sophistication (small) firms bid inelastic curves can rationalize the observed behavior.

The previous papers either take costs as given and estimate conduct and forward positions (Puller, 2007; Hortaçsu and Puller, 2008; Hortaçsu et al., 2019), or observe forward contracts from data and back out costs (Wolak, 2000, 2007). However, with bidding data it is possible to estimate both types of parameters at the same time with minimal assumptions. Intuitively, the forward position is a parameter that is common at the firm level, whereas marginal costs are specific to each plant. As long as more than one bid is observed by some plants or temporal restrictions are considered, it is possible to estimate the cost structure and forward contracts at the same time in a flexible way. Re-writing the first-order condition as a function of the optimal bid helps emphasize the flexible identification of these objects under mild functional form assumptions:[40]

$$b_{ijkt} = c_{jt}(q_j) - \frac{RD_{it}(b) - \theta_{it}}{RD'(b)_{it}}, \tag{26}$$

where k indexes a given price bid for unit j. With mild assumptions on the shape of c_{jt} and several offers per unit j, a flexible estimate of c_{jt} and θ_{it} can be recovered. The identification of the forward contracts is facilitated by the presence of several units per firm, as this is a common parameter. Intuitively, one could identify the two at the same time even with only one plant as long as there are more steps than parameters in the cost function. Empirically, restricting the functional form of the cost function or the forward position based on knowledge about the production function and institutional forward position allows for consistent estimation of the parameters (Reguant, 2014).

4.2.3 FOC approach and pass-through analysis

The data and first-order conditions discussed above lend themselves very well for the study of pass-through of costs. Compared to other applications in the pass-through literature, electricity markets feature a well-known market structure formalized via centralized auctions, detailed high-frequency daily or hourly data, a wide range of cost shocks at a similar level of frequency, and good knowledge of the cost structure of firms.

Fabra and Reguant (2014) uses data from the Spanish electricity market to measure the degree of pass-through of emissions costs. It combines a reduced form analysis of the time series pass-through of emissions costs with a structural analysis of the individual firm's responses. In the time series analysis, and after controlling for the endogeneity of emissions costs using the permit price as an instrument, the paper shows

---

[40] Wolak (2007) derives similar first-order conditions with respect to quantities, as opposed to prices.

very high levels of pass-through. The firm-level analysis tests to which extent firms consider the full cost of emissions in their bidding behavior. It uses a weighted regression estimation similar to equation (26) that takes the following form:

$$b_{ijkt} = \beta c_{jt} + \gamma e_j \tau_t + \theta\, \widehat{markup}_{ikjt} + \epsilon_{ijth}, \tag{27}$$

in which the markup has been estimated following the steps discussed above and is instrumented with demand shifters such as temperature. The authors find that firm behavior is consistent with full internalization of the costs of emissions, i.e., $\gamma \approx 1$. However, they find that the pass-through of marginal costs is greatly attenuated, with β significantly smaller than one. They relate the finding to measurement error, highlighting the limits of using engineering-based marginal cost data in regression analysis. Kim (2021) extends the work in Fabra and Reguant (2014) studying the pass-through of natural gas costs in the New England electricity market. She shows that taking marginal costs as given can lead to the attenuation bias mentioned above, highlighting how heterogeneity in the generation mix can contribute to additional bias.

4.2.4 Market power and dynamics

One limitation of most of the above papers is that the representation of marginal costs is relatively limited and ignores substantial complexities of how power plants operate, such as the presence of ramping costs from increasing or decreasing production, technical minimums that prevent a power plant from reducing its output below certain thresholds, and startup costs. Hortaçsu and Puller (2008) ignores dynamics by focusing only on bidding behavior in the afternoon. Bushnell et al. (2008) highlights that their markups appear to be negative at night, i.e., firms are bidding below their marginal costs, and explains that such biases are likely driven by dynamic considerations. However, the proposed methodology does not directly address these concerns.

Mansur (2008) points out that the approach followed by Borenstein et al. (2002) and others does not take into account the complex dynamics involved in the production of electricity. The paper proposes a novel method to estimate the cost of electricity generation without relying so heavily on marginal cost data. In particular, it presents a flexible model of decision-making under regulation that predicts production choices as a function of observables. The approach uses a dynamic optimization Bellman equation to motivate the covariates that should be included, which are mostly a range of lags and leads of price cost margins. Looking at data from the PJM market, Mansur (2008) shows that the generation patterns are much better explained with this richer model. When looking at the impacts of restructuring, the paper draws to comparison. Using static costs to compute a counterfactual implies that market imperfections resulted in considerable welfare loss, with production costs exceeding the competitive model's estimates by 13%–21%. However, actual costs were only between 3% and 8% above the competitive levels when using predicted prices based on pre-restructuring behavior based on the more flexible proposed methodology. The paper highlights the importance of considering dynamics.

Other papers take a more structural approach. Wolak (2007) expands the estimation of marginal costs by including ramping costs: costs that depend on output in the previous period. This is achieved by expanding the functional form assumption of the marginal cost. The paper models the daily decisions of the firms when they face ramping costs as well as minimum production levels. Reguant (2014) exploits a feature of the electricity auction design in the Spanish electricity market, which enables firms to bid their fixed costs as additional "complementary" bids. It shows how startup costs can be identified from such an auction design. Using a finite horizon model in which firms look at several days ahead, it shows that a model accounting for such costs helps better capture firm behavior.

Reguant (2014) also shows that a model accounting for dynamic costs helps correct the biases identified in Bushnell et al. (2008). Whereas a static model predicts negative markups at night, a dynamic model predicts prices that are below marginal costs for both a competitive and a strategic firm. Importantly, the prices are higher when firms exercise market power (even if below marginal costs). Therefore, market power still increases prices in all hours. When it comes to hours of high demand, a static model tends to underestimate the amount of market power. If demand is only peaking for very few periods, fewer power plants will be in operation, which limits competition.

4.2.5 Sequential markets and arbitrage

Whereas most of the economics literature has focused on the understanding of electricity markets as a daily auction, electricity markets clear several markets for the same product (e.g., electricity delivered at 10 am on May 10, 2011), in the form of sequential centralized markets. Electricity markets typically open the day before delivery, the so-called "day-ahead market," but also closer to delivery, e.g., several hours ahead or in the real-time market. The combination of different horizons allows for better planning of the operations in the market and can potentially enhance efficiency.

What are the strategic incentives in these markets? Ito and Reguant (2016) explores this question in the Iberian electricity market. A simple theoretical model shows that the relatively inelastic demand in the day-ahead market (often based on forecasted demand) combined with limited arbitrage tends to lead to a price premium in the day-ahead market. In the Iberian market, this leads large firms to under-commit in the day-ahead market, whereas renewables producers partially arbitrage the price differences by supplying more power than forecasted in the day-ahead market.

To examine the welfare implications of this behavior, Ito and Reguant (2016) adapts the quantitative model in Bushnell et al. (2008) to include two sequential markets. The model features Cournot competition in two stages. In the day-ahead market, firms commit to their positions. In the real-time market, firms can revise their offers and either buy or sell to change their equilibrium quantities once some uncertainty in the market has been resolved. The first day-ahead market position acts as a forward position in the real-time market, reducing the incentives to exercise market power by the firms. This can lead to a reduced price in the real-time market as long as there is limited arbitrage in the market.[41]

The equilibrium model presented in Ito and Reguant (2016) can replicate well the observed price premium in the market. In counterfactual analysis, the authors consider the impacts of introducing financial

---

[41] Until the introduction of financial participants (see discussion below), the assumption of limited arbitrage is reasonable in these markets. Supply-side participants are limited by their power plant capacities. On the demand-side, the regulatory mechanism imposes that the day-ahead market plans for almost all expected demand, even if it is profitable to delay demand purchases in the presence of a day-ahead price premium.

arbitrageurs in the market. They find that financial arbitrageurs contribute to an increase in consumer surplus, by lowering the prices in the day-ahead market, but do not increase efficiency. The rationale for this result is that arbitrageurs lower the day-ahead price but increase the real-time price, leading to larger inefficiencies in the final settlement. The paper also complements the theoretical literature in Allaz and Vila (1993) by quantitatively showing that, even with strategic distortions, the presence of a secondary market successfully reduces market power concerns when compared to a single primary market. The paper concludes that price convergence, e.g., due to the introduction of financial participants, should not be interpreted as a sign of increased efficiency on its own.

The presence of a gap between the day-ahead and the real-time market persist in many electricity markets. Borenstein et al. (2008) shows that this was the case in California, in which the real-time price market tended to be above the day-ahead market price. In California, large firms were net buyers of power, thus leading to monopsony power and lower day-ahead prices. It posits that the perceived threat of regulatory intervention prevented participants from arbitraging price differences away. However, over time several markets in the United States, including the California one, have been progressively introducing purely financial (virtual) bidders that are actively encouraged to arbitrage these differences. Financial arbitrageurs take a position in the day-ahead market at a given node in the network (long or short), and commit to undoing such position in the real-time market.

Is the introduction of financial arbitrageurs beneficial? Birge et al. (2018) shows that the ability of arbitrageurs to close price gaps is limited by their access to capital and transaction costs. It also highlights that the complex structure of electricity prices and transmissions constraints opens the door to opportunities for increasing consumers' costs. Jha and Wolak (2020) studies the role of financial bidders in the California electricity market and come out with a more positive interpretation of the role of financial traders. It presents evidence that financial arbitrageurs do successfully remove persistent price differences in the market. Using an event study methodology, it also presents event-study evidence that costs in the electricity market were reduced thanks to this policy change. Mercadal (2021) highlights the role of financial bidders in limiting the scope for collusion in the market. Using data from MISO and an event study combined with a structural model of bidding, she finds that the introduction of financial bidders (or rather, the upcoming threat of financial bidders) limited the markups that firms could sustain. Financial bidders also reduced the price gap between markets.

4.2.6 Modeling transmission and market power

Another aspect that is often simplified in the modeling of electricity markets is the transmission electricity grid. Theoretical analysis of the network features of electricity markets show that even models with simple layouts can be extremely difficult to solve, whether only the output market is considered (Borenstein et al., 2000) or also competition over transmission rights is modelled, in which firms can exhibit additional strategic behavior by withholding transmission even if physically feasible (Joskow and Tirole, 2000). Indeed, most papers above treat the electricity market as having a unique price as a simplifying assumption.

Due to the shift to nodal (geographically disaggregated) prices in the United States, and the growth in congestion due to extreme weather and the increase in renewable power, ignoring transmission constraints has become less palatable. Transmission constraints can substantially impact competition by creating isolated areas in which firms face less competition.

In equation (23), transmissions constraints impact the location and slope of the residual demand.[42] Wolak (2015) quantifies the incentives to exercise market power once transmission effects are accounted for in the Alberta electricity market. He finds that the slope of residual demand is substantially more elastic in the presence of no congestion, with the inverse elasticities (which determine the incentives to markup the bids) being two- or three-fold larger in the hours of high demand. To quantify the benefits of transmission due to reduced market power, he corrects observed bids with the ones based on more elastic residual demands and finds competitiveness benefits of *perceived* higher transmission, even in a worse case scenario counterfactual in which transmission does not actually expand.

Ryan (2021) models the impact of transmission constraints on market power in the Indian real-time electricity market. Because the Indian electricity market only has few price regions, the analysis of transmissions constraints is somewhat easier than in nodal markets, which can have thousands of different price points. Ryan documents that congestion leads to significant increases in market power, as measured by the slope of residual demand, as in Wolak (2015). He then proposes an estimator that adapts Wolak (2003) and Reguant (2014) to account for the probability of congestion in the maximization of expected profits. He then constructs a Cournot structural model with transmission constraints and simulates the value of expanding transmission in the Indian electricity market, finding that it would benefit exporting firms in regions with ample supply but that face congestion when selling their power.

In the language of antitrust policy, transmission constraints affect the definition of the relevant market. How to deal with more complicated markets in which there are thousands of price points in the network? Mercadal (2021) contributes to the study of nodal markets by developing a machine learning methodology to define relevant markets. Using data from MISO, she uses a clustering algorithm to find sets of prices that are highly correlated. To discipline the clustering algorithm, which is otherwise unsupervised, she adds a criterion function based on how well the implied relevant markets replicate observed prices. Whereas this approach makes simplifications about the strategic nature of congestion, it is a useful way of making the analysis of nodal markets more tractable and opens the door to analyzing a variety of important question in these markets. We expect that machine learning tools will continue to be useful in simplifying the complexity of electricity markets.

Using a simple prediction model, Davis and Hausman (2016) documents how the sudden closure of a nuclear power plant near Los Angeles affected the supply of electricity. As a first-order effect, it implied the reduction of a substantial amount of low-cost power, increasing costs of production substantially. Among second-order effects, the closure created increased transmission constraints in the California electricity market. The paper the proposes a methodology in the spirit of Mansur (2008). It proposes a flexible regression approach to predict whether a thermal power plant will produce when aggregate thermal production is at certain levels (using bins). This prediction model is based on pre-period data and it can be used to predict what the first-order effect of the nuclear outage should be, under the assumption that the closure increased aggregate thermal production approximately one-to-one. Comparing this to a model of ex-post behavior, the

---

[42] Additionally, equation (23) becomes much more complex if firms do not take transmission bottlenecks as given.

authors find that power plants in the south of California disproportionately increased their production due to congestion. The paper also confirms that two plants in the south exhibited the opposite behavior, constraining their output more than predicted. These power plants were subsequently fined for market manipulation.

4.3 Renewable power and the energy transition

Electricity markets are undergoing substantial transformations with the increased presence of renewable power, mostly wind and solar power. Renewable power has been growing in part thanks to subsidies, credits, and other incentive schemes such as explicit quantity goals in the form of Renewable Portfolio Standards. As of today, their costs have plummeted and they are expected to emerge as a leading source of electricity.

The literature in Industrial Organization has helped understand several aspects of this transition. What are the impacts of growing renewables on electricity market outcomes and the environment? What are the costs of expanding renewables? How do dynamic costs interact with renewable production? What has been the role of learning-by-doing during the transition?

4.3.1 Estimating the environmental impact of renewables

Several papers have estimated the environmental benefits from increased renewables, which tend to displace polluting technologies such as coal or natural gas. Cullen (2013) quantifies the emissions offset by wind production in the Texas electricity market using a simple econometric framework. Cullen regresses the output of thermal generators in the Texas market on wind output, which to a large extent is exogenous and exhibits substantial random variation after controlling for demand, temperature, and congestion. Once the responses of generators are estimated, he quantifies the implications for emissions by using calibrated emissions rates as a function of the technology.

Kaffine et al. (2013) and Novan (2015) complement this analysis by using direct measurements of power plant emissions from the Continuous Emissions Monitoring System (CEMS) when quantifying the impact of wind power on Texas' pollution. The two papers follow a similar approach. Kaffine et al. (2013) regress total aggregate hourly emissions on hourly wind output under the assumption that it is exogenous, whereas Novan (2015) instruments with wind speed and wind installed capacity. Both papers confirms reductions in major pollutants (CO_2, NO_X, SO_2) although the effects are somewhat smaller than in Cullen (2013). Novan (2015) examines the possibility of increased ramping and startup of power plants leads to increases in individual emissions rates. Fell et al. (2021) further quantifies the benefits from improved transmission in ERCOT and find that the marginal value of wind can be up to 30% higher when transmission constraints are relaxed. In another related study, Dorsey-Palmateer (2019) finds that conditional on the total quantity of thermal power generated, increases in wind output are associated with a shift away from coal generation and towards natural gas. The intuition behind this result is that the intermittency of wind increases the value of natural gas generators that, relative to coal, are easier to ramp up and down. This shift away from coal further increases the GHG mitigation benefits of wind power. Callaway et al. (2018) quantifies the value of renewables across a much wider range of locations in the United States and document substantial variation across regions, rather than within regions. This variation highlights the benefits from improved high-voltage transmission. Understanding the bottlenecks in renewable deployment and transmission expansion is an important topic for future research.

4.3.2 Estimating the market impacts of renewables

Renewables have had substantial impact on the operation of electricity markets. There have been concerns about the intermittent nature of renewables, i.e., the fact that their output is not directly controllable and uncertain, in contrast with more traditional technologies that can be used on-demand. When their intermittency is taken into account, the value of renewables both in terms of emissions reductions and profitability might be hampered (Joskow, 2011; Borenstein, 2012; Joskow, 2019).

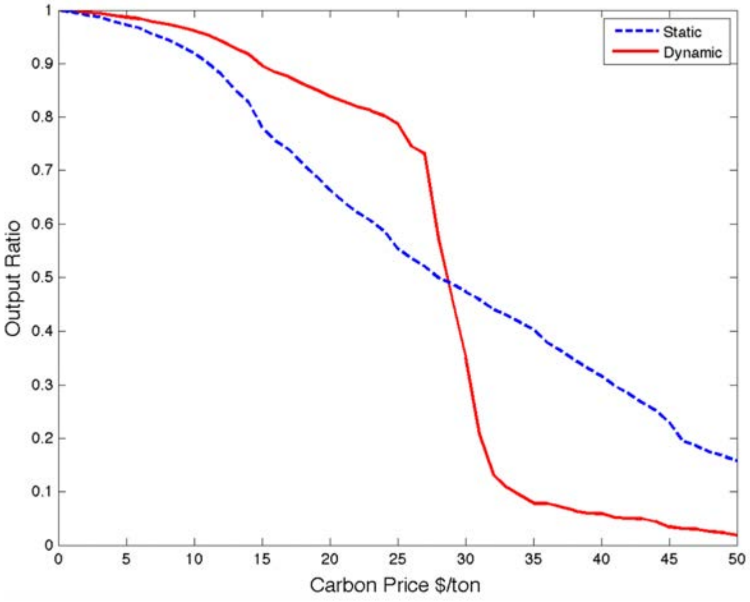
Gowrisankaran et al. (2015) develops a social planner model of electricity dispatch to quantify the cost of intermittency from solar output using data from Arizona. It focuses on understanding the role of reliability constraints in the presence of intermittent output. Electricity markets need to be in constant equilibrium and therefore have certain redundancy built around them to avoid blackouts and other forms of demand curtailment. The model has demand curtailment based on the theoretical work in Joskow and Tirole (2007), which allows for reliability and rationing, and solves for the competitive level of investment. The authors adapt the model to the empirical application of solar production and quantify the amount of investment necessary. The model compares rules of thumb for reliability compared to optimal rules. Through the empirical model of electricity markets operation, it quantifies that the costs of intermittency are overstated when the rules around reliability are not updated to reflect the distinct characteristics of renewable power. Therefore, the paper highlights the need for flexible design of electricity markets in the presence of renewable resources, a topic that needs further attention.

Bushnell and Novan (2018) takes a more reduced-form approach to measure the impacts of renewables on the operation of electricity markets. It uses a retrospective regression analysis to quantify the impact of solar production in the California electricity market. The study highlights that average wholesale prices have decreased thanks to the large increase in utility-scale solar generation, but that the average effect masks heterogeneous impacts. The analysis documents that prices have been raising in the evening, consistent with substantial operational restrictions due to ramping constraints when the sun sets.

Cullen (2015) models explicitly the impact of renewables on the dynamics of electricity markets using a structural model of plant operation. Using detailed hourly plant-level data from ERCOT, he estimates an equilibrium single-agent dynamic model in which thermal power plants (coal and gas) decide whether to run their power plant or not. Power plants are allowed to have minimum production levels, startup costs, and ramping constraints. Due to the presence of startup costs, the introduction of renewables does not necessarily displace coal production, but it can disproportionately displace natural gas, which has smaller startup costs but also smaller emissions rates. Using counterfactual analysis, he concludes that carbon taxes need to be sufficiently large to displace the dirtiest generators, as shown in Figure 3. This exercise highlights the importance of accounting for dynamics when analyzing the energy transition.

Cullen and Mansur (2017) shows that this kinked effect of carbon prices on abatement due to dynamic effects is borne by the data. The emergence of cheap natural gas due to hydraulic fracking can be seen as an analogous cost shock to carbon prices: it makes natural gas relatively cheaper than coal. Using a flexible

Figure 3: Dynamics in electricity markets can substantially affect policy counterfactuals



Note: A static model of electricity generation can be biased. It will overstate the production responses from inflexible generators at low prices, and understate their responses. The dynamic response function is instead kinked. *Source:* Cullen (2015).

regression approach, the paper shows that emissions from fossil fuel power plants have only substantially declined once natural gas prices have made natural gas a more economical technology. Their estimated semi-parametric emissions response to natural gas prices exhibits the same kink as Figure 3.

The presence of these dynamics affects in turn the market price and the profitability of power plants, which impacts their long-run investment strategies. To understand the energy transition, it is crucial to get at these longer run dynamics as well. However, one challenge with the modeling of start up costs in electricity markets is that they make the notion of a competitive equilibrium difficult (O'Neill et al., 2005). Cullen and Reynolds (2015) proposes to adapt a traditional Hopenhayn (1992) model to account for the fact that the presence of startup costs and minimum production limits leads to non-convexities in the production set. The key to their insight is that the model allows firms to have non-convexities, but firms are in themselves assumed to be infinitesimal. This framing of the problem allows to solve for a competitive equilibrium, including investment, while still preserving the main economic impacts of startup costs in profitability.

Linn and McCormack (2019) also develops a computational model to simulate the operational and investment decisions in the electricity market. It uses the model to understand the impact of environmental regulation, renewables, and hydraulic fracturing in the exit of coal power plants. The proposed approach relies on backward induction to solve for the dynamics. To account for startup costs, the dispatch model puts constraints on the number of hours that coal power plants need to operate within a day. One advantage of this approach is that, by simplifying the dynamics to a daily problem, the model can have more detail and arbitrary states, something that is often restricted in dynamic models due to the curse of dimensionality.

Finally Qiu (2020) considers the possibility that ownership of wind assets can convey an informational advantage in wholesale markets. In the U.S. Midwest ISO (MISO), firms owning wind assets historically received higher-quality wind forecasts than did firms without wind assets. Accordingly, Qiu (2020) shows that these firms' bids into wholesale auctions are more responsive to realized wind generation than the bids of firms who are relatively uninformed. The paper then develops a model of MISO auctions in the spirit of Hortaçsu and Puller (2008) to study a counterfactual in which all firms receive the high-quality wind forecasts. Qiu (2020) finds that this information provision induces firms to bid more competitively on average, yielding an average market price reduction of 3%.

In all of these studies, a key ingredient is the fact that renewable power is volatile and, therefore, leads to an increase in the need to startup and shutdown power plants. However, the impact of volatility is expected to be mitigated by the raising emergence of batteries. Optimally using batteries is a strategic dynamic problem and therefore IO tools lend themselves well in this application. How to design and evaluate markets with a large presence of renewables and batteries is an area in which more research is needed. Recent papers in IO have started to explore this dynamic problem in markets in which there is a large presence of solar power, which can lead to important price fluctuations and even negative prices. Batteries act as arbitrageurs to help flatten valleys and peaks.

Karaduman (2020) analyzes the interaction of battery expansion and market power in the Australian national electricity market (NEM). As an interesting observation, it highlights that, in markets with imperfect competition, the incentives of battery operators will be distorted even if they act competitively, as market prices do not send the right signals. He builds a Markov-Perfect Nash Equilibrium model in which each period is a day (with several hours). The dynamic component of the game is to decide the battery level at the end of the day based on the expectations going forward. In the counterfactuals, the paper documents the discrepancy between private and social incentives, which is present when batteries do not internalize their effect on market prices for other participants. This gap between social and private incentives is increased in the presence of market power.

Butters et al. (2020) takes a complementary single-agent dynamic approach to the problem, studying the impact of batteries in California, a market with substantial solar production. It solves for the competitive equilibrium in battery storage, while taking the rest of the market as given with a competitive supply curve. This simpler competitive structure allows them to consider endogenous entry of batteries, which is key to simulate how the market will evolve under alternative subsidy policies. The counterfactuals focus on the understanding the impact of alternative support policies, such as renewable portfolio standards and storage adoption mandates.

4.3.3 Renewables and learning-by-doing

Another dynamic aspect in the adoption of renewable power that speaks to the core of the IO literature is the presence of learning-by-doing in their development and installation. The costs of renewable power have plummeted in recent years and are expected to continue to fall. Early work in the learning-by-doing literature studies the drivers that are behind the cost improvements observed in solar (Nemet, 2006) and wind (Nemet, 2012). These papers take a parametric approach to decomposing the drivers behind the observed marginal cost trends. Anderson et al. (2019) explores to which extent learning is confined to the firm, or it has spillovers to the industry, which can have different implications for optimal subsidy design.

Whether learning-by-doing happens within the firm or there are spillovers, public subsidies can speed progress along the learning curve by incentivizing demand. Gerarden (2019) uses state-of-the art methods in the dynamic games literature in IO to understand the role of demand-induced technological progress in solar photo-voltaic (PV) power. The goal of his empirical model is to quantify the role of public subsidies in stimulating technological improvements. In his model, firms' actions are limited to investing to improve their technological frontier. Subsidies increase the incentives to invest by shifting demand outwards. He estimates the model using detailed firm-level data on the technological efficiency of solar panels, combined with market level information on subsidies, sales (in Watts), and prices. The estimation follows the dynamic games literature and uses conditional choice probability estimation coupled with forward simulation to construct a pseudo maximum likelihood estimator. Using a moment-based Markov perfect equilibrium concept (Ifrach and Weintraub, 2016), the paper then presents two counterfactuals: one in which subsidies only incentivize demand versus one in which increased demand also leads to increased incentives to invest. The main findings of the paper is that subsidies have been essential at expanding demand for solar panels during the study period (2010-2015), leading to substantial environmental benefits (in the order of $15 billion). Additionally, demand-induced technological progress increases the environmental benefits by over 20%.

4.3.4 Renewables and demand-side dynamics

Other papers have focused on the dynamics on the demand side of renewable adoption. De Groote and Verboven (2019) microfounds a dynamic demand system using a single-agent optimal stopping problem in the spirit of Rust (1987). In the model, consumers decide when to optimally invest taking into account the evolution of future costs and subsidies. Additionally, consumers may discount the future benefits of solar panels too much, in line with the literature in fuel efficiency and vehicle adoption discussed in section 3.1.1. This opens the possibility that investment-based subsidies, which occur at the moment of installation, are preferred to production-based subsidies, which are materialized over a longer span of time, typically 20 years.

The study analyzes a program in Belgium between 2006 and 2012 using detailed level at the local market to allow for consumer heterogeneity. The intuition behind the identification of the discount factor, which is typically difficult to identify, lies on the fact that predictable changes in subsidies impact the expected value of waiting but do not impact current utility (Magnac and Thesmar, 2002). Using predictable variation in future solar subsidies, the discount factor is estimated to be of around 15 percent in their application. The authors interpret this discount factor as large when compared to market rates, and quantify that solar programs could be substantially cheaper for the government if they directly subsidized the upfront costs of investment.

Langer and Lemoine (2018) examines residential (rooftop) solar PV subsidies in California by focusing on deriving the optimal subsidy schedule. The paper expand the theoretical literature by considering optimal subsidy decisions when consumers are forward-looking and costs are trending down, highlighting an important trade-off in the setting of optimal subsidies. On the one hand, optimal subsidies could follow a declining path, taking advantage of improved technological progress. On the other hand, optimal subsidies could follow an increasing path, taking advantage of early adopters, who, via selection, have high willingness to pay. Using detailed installation-level data from California Solar Initiative, coupled with electricity rates and demographics, the authors estimate the underlying valuation of consumers for solar panels. Counter to observed programs in practice, the optimal subsidy is increasing.

Feger et al. (2017) also studies the adoption of solar panels using using a unique matched dataset on energy consumption, income, wealth, solar panel installations, and building characteristics for 165,000 households in Switzerland in 2008-2014. The focus of the study is on the distributional implications of solar early adoption, which tends to be correlated with income (e.g., see also Borenstein (2017) for a distributional analysis of the California Solar Initiative). The early adoption of solar panels by wealthy households can lead to what is often called a "death spiral" in utility rates. By investing in solar panels, households are able to reduce their utility bills and contribute less to paying the fixed costs of the electricity grid. The authors examine to which extent a social planner can change the parameters of electricity rates to trade-off efficiency and equity concerns, combining a dynamic adoption model with a regulator's objective to maximize welfare subject to equity concerns.[43]

Feger et al. (2017) uses the data to estimate structural models of energy demand and PV installation. Households are forward-looking and solve a dynamic optimal stopping problem in their solar panel adoption decision. The authors estimate the parameters of their energy demand function using a geographical boundary regression discontinuity design that exploits price variation at spatial discontinuities between electricity providers, to address the endogeneity of energy prices and fees. The combined electricity consumption and PV adoption models allow them to simulate the effect of energy tariffs and subsidies on PV adoption, welfare, and redistribution. In the context of Switzerland and in contrast with results from California (Borenstein, 2017), the paper finds that the impacts of solar adoption are not regressive on average, although they are highly uneven. Whereas a subset of rich households benefit from solar, other rich households do not adopt and pay high electricity costs due to their high levels of consumption.

4.4 Electricity demand

Electricity demand, particularly at the residential level, is known to be substantially inelastic, due to the limited ability or incentive of households to respond to electricity prices. A growing literature is concerned with the estimation of the elasticity of demand as well as experimenting on the best ways to get demand to respond. Increasing the response from demand can become particularly valuable in the presence of growing renewables and extreme weather events. Whereas technological advances in batteries are facilitating the response to such emerging challenges, demand response can contribute to make the transition more efficient.

---

[43] A similar paper is Wolak (2016), which studies the trade-off of setting water rates when considering equity and water balancing goals. Wolak (2016) uses a structural approach to model the optimal non-linear tariff of a utility that needs to constrain water use in the presence of consumer heterogeneity and equity concerns.

4.4.1 Estimating the elasticity of electricity demand

A first step at understanding the scope for demand response is to credibly estimate the elasticity of electricity demand. When thinking about the elasticity of demand, it is important to keep in mind the horizon: short-, medium-, and long-run. The short-run elasticity of electricity demand tends to be very small, in part because short-run retail prices tend to exhibit no variation for many residential consumers, thus preventing any response to wholesale price shocks. The medium- and long-run elasticities of electricity demand can be larger due to the ability of households to respond to high electricity prices via capital investments, habit formation, or adoption of substitutes (e.g., heating or cooking modes). In all cases, the empirical difficulty lies in finding credible and exogenous variation in prices that can enable the estimation of these effects. In this subsection, we review papers that use observational data. In the next, we examine the experimental literature.

Reiss and White (2005) uses a structural model and household-level survey data from the Residential Energy Consumption Survey (RECS) in California to estimate the elasticity of electricity demand. The key to the identification strategy resides in exploiting the presence of non-linear pricing. As is common in many utility settings, consumers face higher marginal prices as their electricity consumption increases. This increasing marginal price schedule implies that an OLS regression of consumption on marginal price cannot consistently estimate households' demand elasticity. By construction, consumers in the higher tiers are those who consume more. Therefore, regressing consumption on prices without taking into account the source of price variation would produce biased results.

Reiss and White (2005) exploits the different tiers as a source of price variation. Variation of behavior within a tier helps identify the role of other covariates in shifting electricity demand. To control for the endogeneity of the tier at which consumers are at, and the presence of bunching, it uses the selection term from the structural model to control for the bias that would otherwise be present in the OLS regression. Intuitively, consumers on the lower tiers have lower unobserved demand shocks that one needs to control for, and these are reconstructed with structural assumptions. The main assumption is that these random shocks (e.g., a month with higher than usual need for electricity) induce consumers to face different prices at the margin. Reiss and White (2005) finds that the elasticity estimated using this approach is -0.39 on average. This elasticity is relatively high for electricity consumption and should be interpreted as medium- to long-run. Interestingly, the paper is also able to identify heterogeneity in the elasticity estimates. Intuitively, it finds that households with either heating or air conditioning are more responsive to the level of prices, with elasticities of -1.02 and -0.64 respectively. It also finds that lower income tends to be associated with more responsive behavior (-0.49 vs. -0.29 for the lowest and highest income quartile).

A key underlying assumption of Reiss and White (2005) is that consumers are aware of the marginal price they are facing when choosing their electricity consumption. Taken literally, the framework predicts that consumers should be bunching around the quantities in which there is a change in the price schedule. However, this bunching is rarely observed in practice. Ito (2014) examines to which extent the lack of bunching in consumption could be due to the fact that consumers respond to average prices, as opposed to marginal prices. To reduce concerns about the comparison between households in different areas, he focuses on studying consumer behavior in Orange County, CA, which is served by two utilities (Southern

California Electric and San Diego Gas & Electric). He then focuses on the consumers surrounding the border of the two regions. Interestingly, the border cuts through several cities, as opposed to coinciding with other administrative borders.

Using detailed monthly electricity billing records, Ito (2014) shows that households at the two sides of the border are comparable, but that they face substantially different non-linear tariffs during the study period. He then uses a differenced panel regression approach with city-by-time fixed effects in which the dependent variable is the log of electricity consumption and the independent variable is the log of the individual price. Because the individual realized price is endogenous to consumption in itself, he instruments the observed price with the predicted price based on that households' lagged consumption. In the context of this regression, he finds that average prices are more negatively correlated with consumer consumption. When both average price and marginal price are included in the regression, the average price seems to be capturing all the explanatory power, thus lending support to the theory that consumers are not fully responding to marginal prices. Based on this estimation, he finds that the elasticity of households to average prices is in the order of -0.08.

Shaffer (2020) uses a similar design using monthly billing data from British Columbia, exploiting a tariff change in which one utility transitioned into a two-tiered tariff but the neighboring utility did not. Different than Ito (2014), Shaffer observes bunching in the distribution of electricity consumption. Using a variety of bunching estimators, he is able to estimate an elasticity of around -0.04. When considering the differenced regression framework as in Ito (2014), the marginal price, as opposed to the average price, is the one with the most explanatory power. How to reconcile these results? Shaffer develops a structural model in which there are three types of consumers: those who respond to marginal prices, those who respond to average prices, and those who believe that the marginal price applies to all consumption. He uses an indirect inference approach to find the share of types by consumption decile that are most consistent with the reduced form regression evidence. He shows that the effects appear to be driven by a small share of consumers who behave as if the marginal price affected all consumption, and thus have very high incentives to bunch.

The previous papers are eminently based on exploiting differences in pricing due to the tariff structure, either as a cross-section or with some panel variation. Can one estimate the elasticity of electricity demand in the absence of non-linear pricing? Deryugina et al. (2020) exploits a time series change with the shift towards municipal choice aggregators in the state of Illinois, which led to sudden and significant reductions in the prices of electricity. A key challenge in the identification of the price effects is that communities that select into community aggregation plans are not exogenous. To alleviate these concerns, they match communities to similar comparison controls based on detailed electricity usage data. Using this flexible difference-in-difference matching approach, they estimate that the price elasticity of demand grows from -0.09 in the first six months to -0.27 two years later. This highlights that consumers can respond to permanent decreases in prices in the medium-run.

The previous papers focus on estimating the medium- to long-run elasticity of residential consumers. Absent explicit pricing programs, most consumers do not face variable prices in the short-run. Fabra et al. (2021) takes advantage of the fact that Spain defaulted a large share of residential consumers into real-time pricing. Using detailed smart meter data together with wholesale electricity prices, the paper estimates individual elasticities using a simple log-log regression, instrumenting the price of electricity with wind production forecasts. The results show that the short-run elasticity of consumers is essentially zero in their setting, in which consumers are often unaware of the price, and price variation in wholesale prices is relatively modest when compared to the final electricity price that consumers face.

Finally, there is a smaller literature estimating the elasticity of commercial and industrial consumers to electricity prices by exploiting the fact that this type of consumers often face variable prices in the form of mandatory time-of-use pricing (Jessoe and Rapson, 2015) or critical peak pricing (Blonz, 2016). Both papers estimate that commercial and industrial consumers are responsive to these programs. In terms of quantifying their elasticity of demand, Blonz (2016) finds, however, just a modest short-run elasticity to peak pricing in the order of -0.11.

4.4.2 Experimental evidence on demand response

A growing literature in demand response exploits the presence of smart meters and the possibility of experimenting with pricing designs that more actively engage consumers. It would be impossible to cover all papers in detail, thus in this section we present a small selection of papers.[44]

Jessoe and Rapson (2014) estimates the demand responses of households to a pricing experiment by partnering with a utility in Connecticut. The pricing experiment features critical peak pricing, i.e., in days with difficult operational conditions, households face much larger prices. In the treatment arms, these scarcity conditions are communicated to households either the day before or 30 minutes in advance. Additionally, a subset of treated households receive an informational display. The main finding is that consumers substantially respond to these pricing events when they have the informational display, and particularly so if they are told a day in advance, with an elasticity of almost -0.20 for the most responsive group. Is this elasticity a truly short-run response? Or does it reflect potential medium-run adaptation? Jessoe and Rapson (2014) finds evidence that consumers responded in the medium-run, by changing their behavior more generally.

Increasing the short-run elasticity of demand can be critical to reduce operational costs and to mitigate market power. Do these pricing events meaningfully engage real-time responses? Or is automatic technology necessary? Bollinger and Hartmann (2020) highlights that peak-pricing experiments have one limitation: the day of the pricing event is not random. Instead, it occurs during periods of high electricity demand. Thus, the estimated effect can be a combination of both short-run and medium-run responses. The paper proposes to use a control function approach to account for the endogeneity of pricing events and better capture the non-linear patterns of demand consumption shocks and their correlation with price. Using experimental data, the findings show that in the absence of technology, the short-run elasticity appears to be close to zero even for households with informational displays. Technology appears to substantially increase the elasticity of consumers. How to enable and evaluate the roll-out of technology-assisted demand response is likely to require more research attention in the near future.

A challenge of many dynamic pricing experiments is that the participation rates are rather modest, thus questioning the ability to scale up such programs. What would happen if consumers were defaulted into such tariffs instead? Fowlie et al. (2020) studies a large randomized controlled trial in the Sacramento Municipal

---

[44] The interested reader can find a recent review of experimental papers in electricity demand in Harding and Sexton (2017).

Utility District (SMUD) between 2011 and 2013. One of the major contributions of the paper is to identify the effect of defaults in the adoption of dynamic pricing by explicitly randomizing the default tariff in the two treatment arms, one with a default flat tariff and one with a default critical peak pricing scheme. In the experiment, over 90% of households decided to keep the dynamic tariff, while only 20% would have actively opted in. Even if the average response of those who were defaulted is lower on average, the aggregate effect can be substantially larger once participation is taken into account. The results suggest that defaulting consumers into explicit dynamic pricing tariffs, when politically feasible, might be a good avenue to increase the scope for demand response. The experiment in itself allows them to quantify the elasticity to critical peak pricing events, which they estimate to be -0.075, in line with the previous literature.

Other papers in the literature have explored the potential of using non-price signals to engage demand response in moments of critical need. Ito et al. (2018) compare critical peak pricing events to messaging encouraging consumers to reduce their electricity demand in the aftermath of the Fukushima accident in Japan. They find that the response to pricing incentives is more pronounced, but that informational campaigns in moment of scarcity can also lead to demand responses. Andersen et al. (2019) also finds evidence consistent with this pattern in a randomized controlled experiment in Denmark. It finds that customers respond to both encouragements to reduce or increase power, by shifting their consumption, and that the patterns are more pronounced in the presence of monetary incentives but still there in the presence of an environmental motivation.

Whereas some of these experiments are encouraging, with the exception of Fowlie et al. (2020), most of them are focused on only a small subset of households who select into the experiments. Given the need to increasingly engage consumers to respond to extreme weather events and extreme operational conditions as well as the growing ability to respond to prices algorithmically, further work in this area will be needed to understand how to scale-up these pricing mechanisms. These responses in demand will be critical to encourage to ease the adaptation to increasingly stressful conditions in the grid due to climate change, e.g., as experienced in California in 2019 and 2020 during the wildfire seasons or due to extreme cold in Texas in 2021. Understanding how demand can be flexibly deployed while ensuring safe conditions and mitigating increased inequalities will be of utmost importance.

4.4.3 Competition in retail electricity markets

Given the low demand elasticity by residential customers, a natural IO question in the context of residential demand is the extent to which firms compete to attract customers. First, it is important to note that competition in the retail side of electricity has lagged behind the liberalization of generation, both in the United States and Europe. This is already an indication that generating competition has been difficult. Second, even conditional on introducing liberalization of retail tariffs, growing competition among firms was modest in its first attempts, due to limited attention and also the difficulty of independent retailers to hedge their contracts, which often leads to attrition in fringe retail competitors.

Because electricity retailing has often been in the hands on the incumbent distribution company, a big hurdle to increase competition is the presence of consumer inertia and switching costs. Hortaçsu et al. (2017) explores the presence of these barriers in the context of the Texas electricity market, ERCOT. The retail market for electricity in ERCOT was liberalized in 2002. Consumers were defaulted into the incumbent company, but they had the option to visit a website to change providers. Even though prices of the incumbent tariff were higher and that savings could be substantial (around 8% per year on average), consumers took a long time to switch. Four years after the policy change, over 60% of consumers were still on the default tariff.

The key contribution of the paper is to separately identify how much of the persistence is driven by search frictions and inattention vs. a preference for the incumbent provider. This separation is important, as it might affect the shape of public policies intended to make the retail market more competitive. They identify these two sources via a structural model of consumer choice with two stages. First, consumers decide whether to consider switching retailers. Second, conditional on going to the website, they can chose among all retailers. The identification strategy combines parametric assumptions on the probability of considering alternative choices and the utility provided by each retailer with unique flow data on the amount of consumers transitioning between retailers. Intuitively, the parameters that inform the probabilities of switching and the households' utilities need to be consistent with the observed flows. The authors estimate the model using a GMM approach in which they match observed flows to predicted flows as a function of parameters. To allow for persistence, the model allows consumers to be more attached to the incumbent the longer they stay with the default provider. The study highlights that small interventions that increase switching and/or reduce the preference for the incumbent would substantially increase consumer welfare.

The previous paper documents the stickiness in consumer behavior in electricity retail markets, but remains silent about the competition implications of such frictions. Duso and Szücs (2017) quantifies the degree of competition in the German retail electricity market by examining the pass-through of wholesale costs to retail tariffs during the period of 2007-2014. It exploits panel variation in Germany's competitive structure and wholesale final prices to examine how cost shocks are passed through to consumers. Using a fixed-effects regression model, the findings show evidence that competitive retailers exhibit larger average pass-through than incumbents (0.70 vs. 0.50), consistent with the presence of switching costs. The results also show evidence consistent with the market becoming more competitive over time, at least as explained by the presence of a pass-through that gets closer to unity in all segments.

Byrne et al. (2019) performs a field experiment to better understand the competition impacts of search frictions in the context of the electricity market of Victoria, Australia. In this market, low-income households are able to receive subsidized electricity rates, but consumer reports suggest that the pass-through of the subsidy is incomplete, leading to concerns about competition. The authors perform a randomized audit study in which actor-consumers pretend to be searching for better electricity rates. They show how these experiments can help identify the degree of discrimination and competition in the market in a setting with negotiated prices and search costs.

Further work is needed to understand how recent changes in the supply side of electricity market are affecting competition in the retail market. It is also critical to understand how the increased used of dynamic pricing schemes will impact competition and consumer surplus. Due to the likely increase in price volatility and costs, the distributional implications from retail competition can also be a substantial, another important area of future work.

4.5 Cap-and-trade regulation in electricity markets

The electricity sector has traditionally been one of the largest emitters of airborne pollutants, either in the form of local pollutants (particulate matter, SO_2, and NO_X) or global pollutants (CO_2 and methane), together with other toxic agents (mercury). Thus, power plants have been subject to numerous environmental regulations, initially due to local pollution concerns and more recently due to climate change concerns. Several tools have been used for the regulation of harmful emissions. Command-and-control approaches and state-level standards have been a common tool, but cap-and-trade markets have gained prominence in the regulation of emissions since the start of the SO_2 trading program in 1995.

Electricity markets provide a unique opportunity to study cap-and-trade regulation because there is a lot of institutional knowledge and data about both the product market (production of electricity) and emissions market. This is not necessarily true in other sectors, in which emissions regulation is well observed but the modeling of the supply side can be challenging due to the presence of varying products and technologies, unobserved heterogeneity, and limited frequency of data. Therefore, one can consider the strategic interactions between firms and markets with a great amount of detail.

There is a vast literature examining the design and performance of these markets. See, for example, a retrospective of the SO_2 trading program in Schmalensee and Stavins (2013), which regulated emissions of SO_2 from power plants in 21 eastern and Midwestern states.[45] There is also a large body of literature studying the properties of cap-and-trade markets. We provide here a necessarily incomplete overview of empirical papers given the breath of IO topics that can be explored using the data from these markets.

4.5.1 Emissions markets in the US electricity sector

The maximization problem faced by electricity generation firms in the presence of cap-and-trade regulation can be summarized as follows:[46]

$$\max_{q_{it},a_{it}} \; p_t q_{it} - C_i(q_{it}, a_{it}) - \tau_t \times (e_i(a_{it}, q_{it}) - \phi_{it}(q_{it})) \,, \tag{28}$$

where i indexes the firm, t a time period (e.g., an hour), p_t is the wholesale electricity price, C are production costs, which can depend on quantity as well as abatement effort, τ_t is the permit price, e_i is the emissions function, which can be reduced with abatement effort, and ϕ_{it} is a free allocation of permits that reduces the environmental costs of firm i. The free allocation can be a fixed amount of permits, what is often called "grandfathering," or potentially an endogenous function of the firm's output q_{it}, what is often called "output-based updating."

There are many IO questions that have been studied in this setting, either theoretically or empirically. Are these markets functioning as expected? Are there market power concerns in both, or inefficiencies spilling from one market to the other? What is the role of the free allocation ϕ? What about in a dynamic sense? What are the interactions between the product and the emissions market?

---

[45] See also Schmalensee et al. (1998) for an early economic review of how the market performed in its first years of operation.

[46] For notational simplicity, the equation here represents a firm with a single power plant, in practice firms maximize profits summing over several power plants.

Carlson et al. (2000) presents a first assessment of the abatement efficiency of the SO_2 market by comparing observed outcomes to an ideal least-cost solution. The study highlights that many of the seemingly large gains from trade from the introduction of the SO_2 market were thanks to the advances in the ability to burn low-sulfur coal at existing generators.[47] It then focuses on computing the efficiency of the market taking holding the low-sulfur technology improvements constant. The paper finds that allowance prices were close to marginal abatement costs, suggesting that the market was sending efficient signals at the margin. However, the gains from trade are estimated to be smaller than the overall potential savings,finding that the performance was in the order of 43% of an idealized least-cost solution. This suggests that there were frictions in the market limiting the amount of trade.

Ellerman and Montero (2007) explores to which extent the SO_2 market functioned efficiently over time and comes with a more positive assessment. Using aggregate data, the paper documents the evolution of Phase I of the SO_2 market, highlighting that firms banked their 30% of the allowances for the more stringent Phase II of the program. It then tests the temporal efficiency of these markets as in the exhaustible resource literature covered in section 2.1.1. The paper calibrates a model of optimal banking and shows that, under reasonable parameters, the observed aggregate trading patterns are not inconsistent with optimal banking. The two findings are not inconsistent, as allowance prices could reach an intertemporal equilibrium consistent with efficient trading but there could still be room in the market for unrealized trades in the presence of frictions.

The presence of transaction costs interacts with a common feature in cap-and-trade markets: the free allowances granted to the participating firms to cover at least part of their expected production, captured by ϕ_{it}. The permit allocation can lead to distortions in the market if there are transaction costs or if the allocation is a function of endogenous factors; e.g., if it is tied to production.[48] In equation (28), the allocation ϕ will clearly distort decisions if it is a function of output (statically or dynamically in a more general framework). It can also lead to distortions if trading permits is costly and thus the effective emissions costs depend on trading positions and not just the permit price τ_t (Stavins, 1995).

Because the initial allocation is by definition almost always endogenous and correlated with output, regressing output on the initial allocation will tend to suffer from reverse causality and lead to positive estimates. Fowlie and Perloff (2013) exploits random allocations due to different allowance allocation timing in the RECLAIM NO_X market in Southern California and finds no evidence that the initial allocation, which was independent of future behavior, affected firms' choices.[49]

In a dynamic setting, a counterfactual approach has been used to explore the impacts of alternative allocation rules. Fowlie et al. (2016), which is covered in more detail in section 5.1.2, examines the role of the allocation in the cement industry. In the electricity market context, Dardati (2016) builds a dynamic

---

[47] Even with the exercise of market power by railroads discussed above that put upward pressure on low-sulfur coal prices (Busse and Keohane, 2008), abatement costs from the program were dramatically reduced thanks to these technological advances.

[48] See Hahn and Stavins (2011) for a review of the distortions that can arise from the initial allocation and a review of the literature and findings in several cap-and-trade markets.

[49] In unpublished work, Reguant and Ellerman (2008) explores a non-linearity in the allocation of permits to coal power plants in Spain to address the endogenity issue, also finding that the allocation did not affect firms' behavior. However, the empirical variation is quite limited. Using a logit model of operation decisions, the paper also shows that electricity firms treated environmental costs in the same way as revenues, lending further support to the lack of distortions.

model in the style of Hopenhayn (1992) to show how the allocation, which in some markets is conditional on staying in the market, can affect dynamic entry and exit decisions. The paper builds an empirical model of the SO_2 market and shows that giving permits to new entrants, as opposed to letting incumbents keep them, can substantially impact productivity by encouraging more entry.

Toyama (2019) explores the extent to which market frictions, such as transaction costs, could have limited trading using a dynamic structural model. The model abstracts away from strategic market power, given the previous evidence, and models firms in a single-agent dynamic competitive equilibrium. In the model, firms are choosing both their abatement in the static problem in equation (28) as well as their decision to bank permits for the next trading period. In the presence of transaction costs, banking can be a substitute for trading. The paper estimates the heterogeneous effect of transaction costs, which depends on trading position: those who buys (sells) permits tend to under- (over-) bank. The paper quantifies the welfare and efficiency implications of this dynamic friction.

Chen (2021) also uses a dynamic single-agent model to explore departures from efficiency in the SO_2 market. In the model, firms have biased beliefs about the permit prices, which can lead to inefficient dynamic decisions about storing allowances and can be a complementary source of trading frictions. Separately identifying belief biases from other unobservables can be challenging in the absence of a metric to define accurate beliefs. Methodologically, the paper shows how to identify the bias in the beliefs by taking advantage of the fact that decisions of firms, other than their banking of permits, also affect compliance costs. In particular, firms can use lower sulfur coal, and the value of using a permit today vs. reducing sulfur content should equalize in equilibrium. Failure of treating these two choices analogously can provide evidence of lack of optimization. Empirically, the results suggest that firms under-reacted to price fluctuations, and therefore it is consistent with firms' expectations being biased towards lower price volatility.

Other papers have examined strategic distortions arising from the interaction between the permit and the electricity market. Permit prices are an important cost shifter to producers in the electricity market, and therefore they can have substantial impacts on product prices. Kolstad and Wolak (2008) explores these strategic interactions in the South Coast Air Quality Management District (SCAQMD), which regulated NO_X emissions in Los Angeles. It finds that some power plants claimed to face higher than average prices in order to circumvent regulatory scrutiny in the power market.

In recent years, cap-and-trade markets have also expanded to cover emissions related to climate change, for example with the emergence of the EU Emissions Trading Scheme in Europe or the AB-32 cap-and-trade market in California. Given the consolidation of these existing cap-and-trade markets for greenhouse gas emissions, and the likely expansion of these markets to other territories, we expect the work in this area to continue to grow.

4.5.2 Interactions between cap-and-trade regulation and regulatory regime

As explained in section 4.1, the United States is unique in its heterogeneous regulatory treatment of electricity generation. Several papers have studied how this regulatory heterogeneity interacts with the functioning of environmental markets.

Fowlie (2010) studies heterogeneity in compliance choice in the context of the NO_X Budget Trading Program, which set a cap of NO_X emissions across several states in the Eastern interconnection. The paper begins by documenting the technological frontier that establishes a trade-off between various capital investments, with NO_X control technologies of varying effectiveness, and the purchase of emissions permits. Using a discrete choice logit model, the paper then shows that generators in areas governed by traditional cost-of-service regulation were likely to comply with the NO_X regulations by investing in high-capital options that aggressively reduced pollution. In contrast, plants in restructured electricity markets were more likely to comply by using less effective but cheaper technologies or by just purchasing permits in the NO_X market.[50] Because the damages from NO_x are not evenly distributed in space, the paper documents that the disparity in capital investment choices between states with traditional utility regulation versus restructured markets led to emissions reductions being relatively less ambitious in the high damage areas.[51]

Abito (2019) develops a structural model of the regulatory distortions arising from natural monopoly rate-of-return regulation. Compared to Fowlie (2010) and Cicala (2015), the paper studies these interaction in the context of the Phase I of the SO_2 market, in which capital investments were more limited. The paper focuses instead on the distortions in effort to reduce costs (Laffont and Tirole, 1993), showing that firms are more inefficient in periods in which rates are being set. The distortions in effort due to asymmetric information result in marginal abatement costs that are larger than in a first-best setting, therefore interacting with the cap-and-trade market. A cap-and-trade market designed to achieve the first-best will fail to do so in the presence of these additional distortions. The paper exploits the timing of rate cases to identify the cost and effort function and infer the marginal abatement cost curve under optimal effort vs. no effort. It then derives the optimal contract to achieve the second-best and compares it to a limited contract that allows firms to either select into a fixed price contract or choose to be reimbursed for all of their costs, finding that this more intuitive contract can deliver a substantial amount (65%) of the welfare gains achieved by the optimal contract.

5 Environmental regulation of energy-intensive industries and natural resources

Energy markets and resource-intensive activities are heavily regulated due to the large presence of environmental externalities. We have already covered many of the regulations aimed at the energy sector that have the purpose of mitigating pollution. In this section, we explore applications that are not confined to energy markets.

A large literature in environmental economics documents the health impacts of environmental externalities, how regulation has lead to a reduction in air pollution, and the economic value of such reductions.[52]

---

[50] This pattern is also documented in the context of sulfur dioxide abatement in Cicala (2015), which highlights how the distortion fits a model in which utilities have an incentive to over-invest in capital (Averch and Johnson, 1962).

[51] Regulated and unregulated areas affect each other through the equilibrium cap-and-trade price. Fowlie and Muller (2019) further documents the environmental consequences of the disparities in marginal abatement cost curves for regulated and unregulated states. The paper studies the benefits of moving towards alternative cap-and-trade designs that account for the geographical nature of marginal damages using a highly detailed geographical model of emissions damages based on Muller and Mendelsohn (2009).

[52] A leading example of such regulations is the Clean Air Act; see Currie and Walker (2019) and Aldy et al. (2021) for a more

Here we will focus the discussion on papers that use IO tools to study the impacts of regulation on manufacturing firms and their production choices. Due to the importance of trade in the discussions of how to regulate environmental externalities in the manufacturing sector, a large related literature in trade has also examined these questions (Copeland and Taylor, 2004; Cherniwchan et al., 2017).

5.1 Environmental regulation in manufacturing and resource-intensive sectors

The regulation of manufacturing industries is often uneven across countries. Researchers and policymakers have long been interested in understanding the impacts of environmental regulation in the outsourcing of polluting activities (Ederington et al., 2005; Levinson and Taylor, 2008). Such displacement of polluting manufacturing sectors is often rationalized via an environmental Kuznets curve, in which lower income countries are willing to take the more polluting activities.

Recent work in this area has focused on exploiting energy price variation to understand the potential responses to stronger climate change regulation; e.g., in the form of carbon taxes. Ganapati et al. (2020) uses confidential Census data between 1972-1997 to show the response of manufacturing output and prices to energy costs. The paper's goal is to understand whether consumers or producers bear the costs of such shocks. It examines seven industries that reported quantities to the Census: boxes, bread, cement, concrete, gasoline, refining, and plywood. It first estimates the marginal costs of these industries by estimating markups (Hall, 1986; De Loecker and Warzynski, 2012). It then performs regression analysis by looking at the pass-through of marginal costs to prices. Because part of the marginal costs are due to energy inputs, the paper uses a shift-share instrument based on the generation mix of electricity of different regions of the US. The paper finds that producers do not fully shift the costs to consumers, which could either be a sign of market power or a sign that firms are exposed to international competition and facing asymmetric cost shocks.

Muehlegger and Sweeney (2017) examines pass-through in one of the sectors of study of Ganapati et al. (2020), refineries. It exploits the fact that the fracking boom created differentiated shocks depending on whether refineries were close to the new areas of exploitation. It finds evidence that the degree of pass-through is largely affected by whether cost shocks are regional or common to all firms. This heterogeneity in cost shocks is at the heart of the discussions surrounding leakage concerns.

5.1.1 Climate regulation and leakage

In the presence of global pollutants, such as greenhouse gases contributing to climate change, the outsourcing of polluting activities not only has the problem of shifting economic activity and pollution overseas, but it additionally has the challenge that pollution overseas is equally costly to those facing the increased regulation costs. Therefore, the shifting of emissions outside of a jurisdiction does not imply reduced damages to the regulating area. This problem can make unilateral environmental regulation quite ineffective.

"Leakage," the term used for emissions going to unregulated jurisdictions, can be a serious barrier when enacting climate legislation. International trade organizations and countries are recognizing more and more

---

comprehensive retrospective of the literature and its findings than what we can offer here.

the need for carbon adjustments that would tax emissions at the border, but the details on a transparent and wide implementation are still not in place. For example, these regulations can be threatened by the difficulty of verifying emissions from manufacturers in other jurisdictions, leading to potentially inaccurate border adjustments that are difficult to measure. For this reason, the most common regulation to address leakage is via output-based subsidies. Even in such cases, the measurement challenges to set individualized efficient tariffs can be challenging (Fowlie and Reguant, 2018; Lyubich et al., 2018; Fowlie and Reguant, 2021).

Fowlie and Reguant (2021) seeks to inform optimal climate policy that takes into account leakage. It derives simple formulas for optimal output-based emissions permit allocations under leakage, which depend on the elasticity of domestic production, imports, and export, to carbon taxes. To estimate these elasticities, it uses regression analysis to measure the impact of energy costs on manufacturing output using aggregate publicly available data. The cost variation used in the analysis is related to Ganapati et al. (2020) and Muehlegger and Sweeney (2017), exploiting the fact that different industries located in different areas face different cost shocks with the fracking boom. Crucially, the fracking boom also generates differences between domestic and foreign energy costs, which are needed to identify the elasticities of output, imports, and exports needed to inform the optimal output-based subsidies. This paper is more ambitious in its goal, as it allows for a quantification of optimal subsidy policies in cap-and-trade regulation. However, the empirical findings are limited by the use of aggregate data.

Current discussions on how to properly tax emissions at the border involve tracking of the inputs and outputs that go into the production process that are generally abstracted away in IO modeling. To inform these questions, our tools will need to double down on the "IO" and bring back more comprehensive input-output analysis to the table.

5.1.2 Leakage in dynamic models

The regulation of emissions can also have dynamic consequences by affecting firms' entry and exit decisions and location choices. Mechanisms that are equivalent in a static model, such a grandfathering regime or a carbon tax (Coase, 1960), can have substantially different dynamic consequences.

Fowlie et al. (2016) examines the dynamic leakage implications of alternative climate policies in the Portland cement industry, which is responsible for 5% of GHGs emissions worldwide. The paper extends the dynamic game model in Ryan (2012) to incorporate different mechanisms to tax firms for their carbon emissions.[53] The paper is focused on the comparison between a carbon tax, a carbon tax with grandfathered permits, an output-based adjustment mechanism, and a border tax adjustment, in line with the policy tools that are being considered to mitigate leakage in several markets, such as the EU-ETS or AB-32 in California.

Cement has the distinctive feature of being more easily transported by ship than by road. Therefore, there is variation in the degree of international exposure to leakage. The paper estimates cement demand and supply in several U.S. markets and performs counterfactual simulations of coastal vs. landlocked markets. In coastal areas, even though they are more competitive, the social surplus-maximizing carbon price is substantially below the standard Pigouvian tax due to the presence of leakage under output-based regulation.

---

[53] Given the use of dynamic estimation techniques in both Ryan (2012) and Fowlie et al. (2016), the technical aspects of these papers are already covered in more detail in the chapter "Dynamic Games in Empirical Industrial Organization."

The most cost-effective policy for these markets is a border tax adjustment. For inland markets that are not exposed to leakage, surplus-maximizing carbon prices can still be below the standard Pigouvian tax due to the high degree of concentration. The paper also shows that for these concentrated, non-trade-exposed markets, initial levels of taxes can transfer monopoly rents from producers to consumers while having limited impacts on production.

Hsiao (2020) examines the scope for leakage in the palm oil industry, another major contributor to climate change. Palm oil production drives widespread deforestation in Indonesia and Malaysia, including in carbon-dense peatland regions. The paper explores the general equilibrium effects of border carbon adjustments on palm oil imports when not all importing countries participate. Under leakage, lower demand from countries that establish a carbon tax leads to a reduction in the market price of palm oil, which can incentivize demand in other areas. The paper assesses the importance of international coordination and commitment in addressing this leakage.

To get at general equilibrium demand effects, the paper builds a model of world demand for palm oil with a country-level almost-ideal demand system for vegetable oils. On the supply side, it builds a dynamic equilibrium model in which farmers make individual decisions about whether to deforest and, conditional on deforestation, how much to plant. The paper uses Euler techniques to estimate the supply-side parameters (Arcidiacono and Miller, 2011; Kalouptsidi et al., 2021). In the counterfactuals, the paper shows that both commitment and coordination are crucial. Only when combined are these tools moderately successful at reducing deforestation.

Recent decades have seen little success in curbing the speed and extent of deforestation, and further work is needed on this important topic. IO tools are particularly well-suited for studying how deforestation and land use choices respond to environmental regulation. While this handbook is necessarily incomplete, papers like Scott (2013), Souza-Rodrigues (2019), Assunção et al. (2021), and Cole et al. (2021) contribute to our understanding of land use and deforestation and are great examples of how to apply IO tools to address these pressing questions.

5.2 Imperfect monitoring and the enforcement of regulations

A growing literature examines how firms respond to environmental regulation and monitoring using IO tools. Properly enforcing environmental rules is critical to ensure that the expected gains and losses from a given policy are realized. Whereas many of the previous papers assume that compliance with regulations in the jurisdiction of study can be ensured, compliance is not always the norm in practice. In section 3.3, we discussed examples of firms violating fuel economy standards by mis-reporting their vehicles' performance. Here, we highlight other recent work examining non-compliance with environmental regulations.

5.2.1 Evidence of cheating in environmental settings

Properly understanding firms' strategic responses to enforcement and monitoring of environmental regulations can be difficult due endogeneity concerns, such as higher inspections in days with high levels of pollution. To address this challenge, Zou (2021) exploits quasi-experimental variation in the auditing schedule of emissions monitors to show reduced-form evidence that manufacturing plants appear to endogenously respond by shifting pollution away from the days in which they expect to be monitored. It shows that, for monitoring stations in which the day of auditing is quasi-random (determined on a six-day schedule), satellite measures in the surroundings of the monitors appear to record lower pollution levels during the days in which monitoring occurs. The paper then addresses the question of what the mechanisms are that could lead to lower pollution in days in which emissions are officially recorded. It shows that regulators themselves are 10 percent more likely to issue an advisory on monitoring days, which implies that local authorities themselves are being strategic to avoid falling out of compliance.

Oliva (2015) examines cheating at car smog checks in Mexico and finds evidence of bribing at the test centers. The test exploits the institutional details of cheating, which rely on using the measurements from the previous car in the testing lane. Serial correlation with testing outcomes from the previous car can be used as a reduced-form test for cheating. In addition, the paper constructs a maximum likelihood estimator based on the transition probabilities of passing the test. Exploiting the fact that there is substantial variance in the outcomes of the test, even conditional on the same car, the identification assumes that drivers do not learn about their car's pollution after failing a test. At that point, they are faced with the decision of paying a bribe to pass, or re-taking the test, which they can do only once for free. The fact that the second test is free, but not the third one, is helpful to identify the parameters in a two-period model in which drivers can decide to test the car, with or without bribing, and re-test if they fail. The estimation is based on matching the observed probability of passing the test at different observable stages of the process (passing/no passing, retesting, etc.). The estimation finds that 10% of cars pass the test via cheating, an estimate that is confirmed using a simpler approach that compares centers with and without evidence of cheating in their tests. The paper then compares situations with and without bribing to quantify the welfare implications of cheating.

5.2.2 Structural models of enforcement

The previous two papers show evidence of cheating but do not model explicitly the regulatory costs of designing, monitoring, and enforcing regulations. Recent work has explored that interaction structurally.

Kang and Silveira (2018) builds a structural model to understand the regulatory game between the regulator and the manufacturers when it comes to environmental enforcement. The paper focuses on the trade-off between transparency and discretion faced by regulators when setting their rules. Whereas discretion might allow regulators to fill in gaps in otherwise incomplete enforcement policies, it has the risk of leading to regulatory capture and disparate outcomes at otherwise similar sites. The paper's application is enforcement of the Clean Water Act in California. It first documents that penalties for violations of standards appear to exhibit discretion. The paper then builds a model of optimal environmental regulation in which firms are heterogeneous in their costs of abatement but their type is unknown to the regulator. It then sets up an optimal regulation problem a la Laffont and Tirole (1993) that it brings to the data. To estimate the regulator's cost of enforcement, it exploits a regulatory change in 2006 that made monitoring easier and increased the regulator's resources. The empirical findings suggest that discretion provides significant value in this setting.

Blundell et al. (2020) examines the dynamic implications of monitoring and enforcement in the context of air pollution regulation by the U.S. EPA. It starts by noting that the regulation currently in place has a dynamic component, with repeat offenders being subject to larger fines and higher scrutiny. Thus, the mechanism has an escalation component, by which offenders placed under the high priority violator category end up paying much larger fines if they are again out of compliance in a future period. The authors build a dynamic model in which plants need to chose the optimal level of abatement investment given the predetermined schedule of fines set out by the regulator and the presence of dynamic enforcement. The dynamic model is used to estimate the costs of investment and the disutility of paying fines and being placed under high violator status. To allow for flexible heterogeneity, the authors use the estimator in Fox et al. (2011) that allows for discrete types. They then consider counterfactuals in which the regulator does not use dynamic penalties. Their estimates imply that the dynamic structure is effective at preventing offenses while reducing the need for collecting fines. A static policy would require much larger fines to ensure the same level of enforcement.

The enforcement of environmental regulations can potentially be more critical in developing countries, in which the baseline levels of pollution tend to be larger and the ability to monitor could be harder. Using a dynamic model combined with experimental results, Duflo et al. (2018) explores the value of regulatory targeting in the state of Gujarat, India. The state has several major cities in violation of air quality standards, and therefore the value of enforcement can be much larger. Yet in practice, inspections appear to be performed less often than mandated. The paper implements an experimental treatment in which inspections to polluting manufacturing plants are scheduled according to the regulation.[54] However, the paper finds that the reduced-form impacts of the experimental treatment are rather muted. To reconcile the findings, the authors observe that regulators in the control group appear to be targeting their inspections. Even if regulators inspect less often, they often appear to inspect the most suspect facilities. They build a dynamic model in which regulators can use lagged values of pollution to target investment. They estimate the plant's abatement costs using a Rust-style estimator, taking the policy function of the regulator as given. Additionally, they estimate the regulator's preferences by assuming the observed policy is optimal. Using the model, they confirm that increased random inspections are not particularly valuable in this setting.

Continuing to understand how to improve the monitoring and enforcement of environmental regulations will be essential in the following decades, in which fossil fuel emissions need to decline dramatically. Work on reliable and strategy-proof mechanisms to monitor emissions would be valuable. For example, more work could be done in IO to study how to use satellite data to closely monitor worldwide emissions in a way that can circumvent the major challenges of international cooperation and compliance. More work about how to more reliably use satellite measurements will be highly valuable (see, for example, Torchiana et al. (2020) in the context of land use changes in satellite data).

5.3 Regulation of water markets

The over-exploitation of environmental resources is another area that has gotten the attention of the energy and environmental literature. IO tools are well-suited for the study of resource extraction problems, as

---

[54] See also Duflo et al. (2013) for complementary work on the incentives faced by auditors, as this inspection treatment was cross-randomized with an audit reform experiment, with auditors in the treatment being randomly assigned vs. chosen by the plants.

explained in section 2.1. In non-energy settings, water and fisheries are two areas in which IO tools have been used to understand optimal extraction and evaluate or propose solutions for the allocation of these scarce resources.

In this section, we focus on water markets, an area that is getting increasing attention. Water is a resource that has been traditionally under-priced or even not priced at all. With the growing pressures on water demand and the increased variance in rainfall, changes in the institutional treatment of water allocation are likely to emerge in many areas (Libecap, 2011).

Water markets have historically been present in areas with substantial scarcity, such as Spain and Australia, and are gaining prominence in the discussions surrounding other areas facing increasing scarcity. Given the increasing pressures on earth's natural resources, we expect a growing need for economics researchers to help in the design and evaluation of emerging water markets.

5.3.1 Water use and adaptation

Timmins (2002) studies the pricing of groundwater in California in the 1990's. He first confirms the basic fact that water is vastly under-priced by comparing the marginal extraction costs of water (e.g., due to pumping costs) with the price charged to consumers. Prices charged to consumers are always below cost in his sample, and they tend to be much lower. Importantly, this cost comparison ignores the opportunity cost of over-extracting water and depleting aquifers, along with the environmental costs of doing so. Timmins (2002) develops a dynamic single-agent model to quantify the opportunity cost of water. The cost only includes the increased pumping costs of declining aquifers, so it should be considered a lower bound. The paper finds that the opportunity cost of depleting aquifers from a social point of view, yet, one observes prices that are too low. To rationalize these observed patterns, the paper posits that water managers are optimizing a welfare function with different weights for producer and consumer surplus. It then estimates the model and finds that lower prices can be rationalized as the water manager catering to the constituents, as opposed to minimizing costs.

From a methodological point of view, Timmins (2002) develops a single-agent model with persistent unobserved shocks. His proposed approach uses a maximum likelihood approach to deal with the initial conditions problem. The counterfactuals in Timmins (2002) assume that either managers internalized the opportunity cost of water or not, which enables the quantification of the inefficient intertemporal allocation of water due to myopic behavior. In addition to intertemporal mis-allocation, cross-sectional mis-allocation of water is a concern in the absence of markets, an issue that is abstracted away in the paper.

Bruno and Jessoe (2021) and Burlig et al. (2021) study the demand for groundwater in California.[55] Bruno and Jessoe (2021) uses data from the Coachella Valley Water District in the southern part of the state, while Burlig et al. (2021) uses data covering all agricultural users in Pacific Gas & Electric's service territory, which includes most of California's central valley. To estimate agricultural users' demand elasticity, Bruno and Jessoe (2021) uses variation in the price of groundwater itself, and Burlig et al. (2021) uses variation in pumping costs induced by the fact that different farms are on different electricity tariff plans.

---

[55] Bruno and Jessoe (2021) and Burlig et al. (2021) also include reviews of the literature estimating the elasticity of water demand.

Both Bruno and Jessoe (2021) and Burlig et al. (2021) use panel fixed effects regressions to identify the water demand elasticity but arrive at different results. Bruno and Jessoe (2021) reports elasticity estimates from -0.16 to -0.2, but the main estimate in Burlig et al. (2021) is -1.12. This large difference in results could reflect differences in location (e.g. different types of crops grown in Northern versus Southern California) or differences in the frequency and magnitude of the price changes in their data, conditional on the fixed effects. Both papers use data at the monthly frequency, though the price changes observed in Burlig et al. (2021) are long duration, and the paper finds similar elasticities when the data are collapsed to the annual level. When Bruno and Jessoe (2021) uses annual data it estimates a larger elasticity of -0.37, though this is still substantially smaller than in Burlig et al. (2021). Burlig et al. (2021) also finds evidence that crop switching is the main mechanism behind farmers' price responsiveness. Additional research is needed to more fully understand the differences between these two papers' results.

Hagerty (2017) focuses explicitly on long-run mis-allocation of surface water in California. The paper's regression approach takes advantage of differing approaches to the management of water across districts to estimate the long-run responses in terms of revenue and crop choice to surface water scarcity. The paper finds that crop revenues are significantly reduced even in the long-run, suggesting that adaptation cannot mitigate many of the impacts of more scarce and volatile water availability.

5.3.2 Studying formal water markets

Several papers have also analyzed the functioning of institutionalized water markets, in which participants have allowances to use water and can trade with others. We focus here on recent studies that use IO tools for their analysis.

Rafey (2021) studies the southern Murray-Darling Basin (sMDB) in Australia during 2007-2015. The paper proceeds in two steps. First, it uses production function tools (Olley and Pakes, 1996; Ackerberg et al., 2015) to estimate unobserved productivity at agricultural plots. The estimation exploits detailed rotating panel survey data together with meteorological data that together collect information on land, irrigation, rainfall, and other flexible factors (labor and materials). Importantly, the approach combines the standard strategy to control for drivers of productivity with lagged instruments based on the allocation rules for water rights. Second, once productivity measures have been estimated, the paper uses detailed trading data to examine the gains from trade in this market. The paper finds gains from trade in the order of 4-6%.

In a historical context, Donna and Espin-Sanchez (2018a) and Donna and Espin-Sanchez (2018b) study the performance of water auctions in Mula (Spain) during the mid-twentieth century. Donna and Espin-Sanchez (2018a) develops a model of auctions with complementarities to explore to what extent consecutive blocks of irrigated water are complements or substitutes, finding that accounting for complementarities in water use is important to obtain unbiased estimates. The results indicate complementarities are most important in the driest summer months, consistent with the need to irrigate the dry canals for a longer time before water can be productively used. In the same historical context, Donna and Espin-Sanchez (2018b) compares an auction versus a quota mechanism to allocate water, exploiting a historical change in the way water was allocated in Mula. The structural model takes into account market frictions, putting emphasis on liquidity constraints faced by poorer farmers. In the presence of these liquidity constraints, the paper shows that the elasticity of demand from a model ignoring these constraints is biased, overstating the elasticity of demand, and that the quota system can be better as it does not preclude poorer farmers from cultivating productive areas.

As stated above, we expect the work studying water markets to grow substantially in the next decade. More work using IO tools to study the ex-ante ability and ex-post performance of markets to allocate water in an efficient and equitable manner will be needed.

6 Concluding remarks

This chapter has discussed the variety of insights that industrial organization economists have made into energy markets, market regulation, and environmental policy. These insights have stemmed from application of nearly every tool in IO economists' toolkit, ranging from models of differentiated product competition to auction models to models of dynamic oligopoly. In many cases, IO economists' work on energy and environmental problems have enriched these models and their associated empirical methods in ways that are applicable to economic and policy questions in other sectors.

Rather than recap the preceding pages, we conclude this paper by highlighting what we see as the most important energy and environmental questions in need of future contributions from IO economists. Policy interest in climate mitigation is especially intensifying, and insights from IO economists are needed to help design and assess the likely impacts of policy options. In particular, we see the following areas as being especially promising for future IO research:

- **Productivity and innovation in energy technologies.** Productivity growth and innovation have been central in governing the world's dominant sources of energy supply to date, and the future of zero-emissions energy supply will arguably be governed more by the pace of innovation in clean technologies than by any other factor. Our chapter has highlighted the many contributions IO economists have already made to understanding sources of productivity growth and innovation for both fossil and clean energy sources. More work is needed to better understand factors that influence productivity growth, how innovation is affected by market structure, and how policy can accelerate innovative activity in green technologies.

- **Industrial organization of zero-emissions energy sources.** IO economists have delivered deep insights into the economics of fossil fuel energy supply, and especially of the oil and gas industry. Given the ascension of wind and solar power, and the possibilities for geothermal power and carbon sequestration technologies in the future, there is an opportunity for IO economists to apply models and tools from work on fossil fuels—for instance, models for leasing, auctions, productivity growth, and investment dynamics—to these emerging green technologies.

- **Electric and autonomous vehicles.** EVs are beginning to achieve a non-trivial share of new vehicle sales in many countries, so in the next few years we anticipate that a wealth of data on EV markets and EV charging will become available. These new data will present IO economists with opportunities to contribute to our understanding of EV markets and EV policy, well above and beyond the small set of existing papers we have discussed in this chapter. Looking further ahead, the autonomous vehicle (AV) industry is even more nascent than the EV industry, but as momentum gradually builds behind autonomy, research will be needed to understand potential business models for AVs and the impacts of policies proposed to govern this industry on consumers, energy use, and the environment. Ostrovsky and Schwarz (2018) is an early theoretical contribution in this area that provides a framework for jointly thinking about AVs, congestion pricing, and ride sharing. Additional work that further develops this framework, with potential empirical applications, is needed.

- **The future of electricity distribution.** Improvements in technology are enabling households to invest in distributed generation resources (especially rooftop solar), switch from natural gas to electricity for home heating, and charge their EVs at home. Consumers' willingness to choose these technologies is likely to depend on electric utilities' retail pricing policies. As noted in Borenstein and Kellogg (2021), distribution utilities' standard practice of incorporating their fixed costs into volumetric rates will distort households' incentives, since as the grid becomes cleaner retail prices are likely to substantially exceed social marginal cost. This standard practice may also lead to substantial inequities in the presence of wide-scale adoption of distributed generation, which is likely to be disproportionately taken up by high-income households (Borenstein and Davis, 2016). IO economists can contribute to understanding the implications of potential alternative business models and regulatory frameworks for electricity distribution and retailing that address both the efficiency and equity shortcomings of the status quo.

- **Integration of large-scale storage into wholesale electricity markets.** The intermittency of wind and solar energy resources implies that battery storage can potentially play an important role in creating an electricity grid that can supply zero-emission electricity on demand. It is not yet clear how wholesale power markets will operate in the presence of such technologies and how they will provide incentives for the investment in and use of battery storage. Input from IO economists will be vital in evaluating the merits of alternative market designs, see for example Joskow (2019) and Wolak (2021).

- **Stranded fossil fuel assets.** In the event that climate policy leads to a rapid transformation of energy supply, fossil fuel physical and human capital are likely to be abandoned before the end of their originally foreseen useful life. This premature abandonment will have substantial distributional implications, and it may also lead to environmental hazards in the event that physical capital is not properly decommissioned. IO economists' expertise in (dis)investment problems and utility regulation can help inform policies aimed at accelerating asset decommissioning while equitably distributing the costs of doing so.

References

Abito, Jose Miguel, "Measuring the Welfare Gains from Optimal Incentive Regulation," *The Review of Economic Studies*, 09 2019. rdz050.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Adams, William James and Janet L. Yellen, "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 1976, *90* (3), 475 – 498.

Agerton, Mark, "Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale," 2020. working paper.

Aldy, Joseph E., Maximilian Auffhammer, Maureen Cropper, Arthur Fraas, and Richard Morgenstern, "Looking Back at Fifty Years of the Clean Air Act," *Journal of Economic Literature*, 2021.

Alé-Chilet, Jorge, Cuicui Chen, Jing Li, and Mathias Reynaert, "Colluding Against Environmental Regulation," 2021. working paper.

Allaz, Blaise and Jean-Luc Vila, "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, feb 1993, *59* (1), 1–16.

Allcott, Hunt and Nathan Wozny, "Gasoline Prices, Fuel Economy, and the Energy Paradox," *Review of Economics and Statistics*, 2014, *96* (5), 779 – 795.

Andersen, Laura Mørch, Lars Gårn Hansen, Carsten Lynge Jensen, and Frank A. Wolak, "Can Incentives to Increase Electricity Use Reduce the Cost of Integrating Renewable Resources," *National Bureau of Economic Research*, mar 2019.

Anderson, John W., Gordon W. Leslie, and Frank A. Wolak, "Measuring the Impact of Own and Others' Experience on Project Costs in the U.S. Wind Generation Industry," NBER working paper 26114 2019.

Anderson, Soren T. and James M. Sallee, "Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards," *American Economic Review*, 2011, *101* (4), 1375 – 1409.

_ and _ , "Designing Policies to Make Cars Greener: A Review of the Literature," *Annual Review of Resource Economics*, 2016, *8*, 157 – 180.

_ , Ryan Kellogg, and James M. Sallee, "What Do Consumers Believe About Future Fuel Prices?," *Journal of Environmental Economics and Management*, 2013, *66*, 383 – 403.

_ , _ , and Stephen W. Salant, "Hotelling Under Pressure," *Journal of Political Economy*, 2018, *126*, 984–1026.

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

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "(Mis)Allocation, Market Power, and Global Oil Extraction," *American Economic Review*, 2019, *109* (4), 1568 – 1615.

Assunção, Juliano, Robert Mcmillan, Joshua Murphy, Eduardo Souza-Rodrigues, Fernanda Brollo, Sylvain Chabe-Ferret, Francisco Costa, Clarissa Gandour, Kelsey Jack, Charles Manski, Ismael Mourifié, Marcel Oestreich, Martino Pelli, Alex Pfaff, Alberto Salvo, Paul Scott, Edson Severini, Kate Sims, and Aloysius Siow, "Optimal Environmental Targeting in the Amazon Rainforest *," 2021.

Athey, Susan and Philip A. Haile, "Identification of Standard Auction Models," *Econometrica*, 2002, *70* (6), 2107 – 2140.

Averch, Harvey and Leland L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *The American Economic Review*, 1962, *52* (5), 1052–1069.

Bajari, Patrick, Lanier Benkard, and Jonathan Levin, "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 2007, *75* (5), 1331 – 1370.

Baumeister, Christiane and James D. Hamilton, "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 2019, *109* (5), 1873 – 1910.

Bento, Antonio M., Lawrence H. Goudler, Mark R. Jacobsen, and Roger H. von Haefen, "Distributional and Efficiency Impacts of Increased US Gasoline Taxes," *American Economic Review*, 2009, *99* (3), 667 – 699.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841 – 890.

Bhattacharya, Vivek, Andrey Ordin, and James W. Roberts, "Bidding and Drilling Under Uncertainty: An Empirical Analysis of Contingent Payment Auctions," 2020. working paper.

Birge, John R., Ali Hortaçsu, Ignacia Mercadal, and J. Michael Pavlin, "Limits to arbitrage in electricity markets: A case study of MISO," *Energy Economics*, 2018.

Black, Fischer and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 1973, *81* (3), 637 – 654.

Blonz, Joshua A, "Making the Best of the Second-Best: Welfare Consequences of Time-Varying Electricity Prices," *Energy Institute at Haas Working Paper*, 2016.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer, "Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations," *American Economic Review, forthcoming.*, 2020.

Board, Simon, "Selling Options," *Journal of Economic Theory*, 2007, *136*, 324 – 340.

Bollinger, Bryan K. and Wesley R. Hartmann, "Information vs. Automation and implications for dynamic pricing," *Management Science*, 2020.

Bolton, Patrick and Christopher Harris, "Strategic Experimentation," *Econometrica*, 1999, *67* (2), 349 – 374.

Boomhower, Judson, “Drilling Like There’s No Tomorrow: Bankruptcy, Insurance, and Environmental Risk,” *American Economic Review*, 2019, *109* (2), 391 – 426.

Borenstein, Severin, “The Private and Public Economics of Renewable Electricity Generation,” *Journal of Economic Perspectives*, 2012, *26* (1), 67 – 92.

_ , “Private net benefits of residential solar pv: The role of electricity tariffs, tax incentives, and rebates,” *Journal of the Association of Environmental and Resource Economists*, 2017.

_ **and James Bushnell**, “The US Electricity Industry after 20 Years of Restructuring,” *Annual Review of Economics*, aug 2015, *7* (1), 437–463.

_ **and Lucas W. Davis**, “The Distributional Effects of U.S. Clean Energy Tax Credits,” *NBER Tax Policy and the Economy*, 2016, *30* (1), 191–234.

_ **and Ryan Kellogg**, “Challenges of a Clean Energy Transition and Implications for Energy Infrastructure Policy,” 2021. Aspen Economic Strategy Group paper.

_ **, James B. Bushnell, and Frank A Wolak**, “Measuring market inefficiencies in California’s restructured wholesale electricity market,” *American Economic Review*, 2002, *92* (5), 1376–1405.

_ **, James Bushnell, and Steven Stoft**, “The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry,” *The RAND Journal of Economics*, 2000, *31* (2), 294.

_ **,** _ **, Christopher R Knittel, and Catherine Wolfram**, “Inefficiencies and Market Power in Financial Arbitrage: A Study of California’s Electricity Markets,” *The Journal of Industrial Economics*, 2008, *56* (2), 347–378.

BP, “Statistical Review of World Energy, 69th edition,” 2020. `https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf`.

Brehm, Paul A. and Eric Lewis, “To Trade or Not to Trade: Oil Leases, Information Asymmetry, and Coase,” *RAND Journal of Economics*, forthcoming.

Bresnahan, Timothy F. and Peter C. Reiss, “Dealer and Manufacturer Margins,” *RAND Journal of Economics*, 1985, *16* (2), 253 – 268.

Brown, David P. and Andrew Eckert, “Analyzing firm behavior in restructured electricity markets: Empirical challenges with a residual demand analysis,” *International Journal of Industrial Organization*, jan 2021, *74*, 102676.

Bruno, Ellen M. and Katrina Jessoe, “Missing markets: Evidence on agricultural groundwater demand from volumetric pricing,” *Journal of Public Economics*, apr 2021, *196*, 104374.

Burlig, Fiona, Louis Preonas, and Matt Woerman, “Energy, Groundwater, and Crop Choice,” 2021. working paper.

Bushnell, J, E Mansur, and C Saravia, "Vertical Arrangments, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets," *American Economic Review*, 2008, *98* (1), 237–266.

Bushnell, James and Kevin Novan, "Setting with the Sun: The Impacts of Renewable Energy on Wholesale Power Markets," *National Bureau of Economic Research Working Paper Series*, 2018.

Bushnell, Jim, Erin T. Mansur, and Kevin Novan, "Review of the Economics Literature on US Electricity Restructuring," 2017. UC Davis.

Busse, Meghan R. and Nathaniel O. Keohane, "Market Effects of Environmental Regulation: Coal, Railroads, and the 1990 Clean Air Act," *RAND Journal of Economics*, 2008, *38* (4), 1159 – 1179.

_ , Christopher R. Knittel, and Florian Zettelmeyer, "Are Consumers Myopic? Evidence from New and Used Car Purchases," *American Economic Review*, 2013, *103* (1), 220 – 256.

Butters, R. Andrew, Jackson Dorsey, and Gautam Gowrisankaran, "Soaking Up the Sun: Battery Investment, Renewable Energy, and Market Equilibrium," Technical Report aug 2020.

Byrne, David P. and Nicolas de Roos, "Consumer Search in Retail Gasoline Markets," *Journal of Industrial Economics*, 2017, *65* (1), 183 – 193.

_ , Leslie A. Martin, and Jia Sheen Nah, "Price Discrimination, Search, and Negotiation in an Oligopoly: A Field Experiment in Retail Electricity," *SSRN Electronic Journal*, 2019.

Callaway, Duncan S., Meredith Fowlie, and Gavin McCormick, "Location, location, location: The variable value of renewable energy and demand-side efficiency resources," *Journal of the Association of Environmental and Resource Economists*, 2018.

Carlson, Curtis, Dallas Burtraw, Maureen Cropper, and K. Palmer, "Sulfur dioxide control by electric utilities: What are the gains from trade?," *Journal of Political Economy*, jul 2000, *108* (6), 1292–1326.

Chandra, Ambarish and Mariano Tappata, "Consumer Search and Dynamic Price Dispersion: an Application to Gasoline Markets," *RAND Journal of Economics*, 2011, *42* (4), 681 – 704.

Chen, Cuicui, "Identifying Beliefs in Continuous-Action Dynamic Models: An Application to the US SO2 Allowance Market," *Mimeo*, 2021.

Cherniwchan, Jevan, Brian R. Copeland, and M. Scott Taylor, "Trade and the Environment: New Methods, Measurements, and Results," *https://doi.org/10.1146/annurev-economics-063016-103756*, aug 2017, *9*, 59–85.

Cicala, Steve, "When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation," *American Economic Review*, January 2015, *105* (1), 411–44.

_ , "Imperfect Markets versus Imperfect Regulation in U.S. Electricity Generation," *American Economic Review*, 2020.

Clay, Karen, Akshaya Jha, Nicholas Muller, , and Randall Walsh, "External Costs of Transporting Petroleum Products: Evidence from Shipments of Crude OIl from North Dakota by Pipelines and Rail," *Energy Journal*, 2019, *40* (1), 55 – 72.

Coase, R. H., "The Problem of Social Cost," *The Journal of Law and Economics*, 1960, *3*, 1–44.

Cole, Matthew T., Jacqueline M. Doremus, and Stephen F. Hamilton, "Import Restrictions by Eco-Certification: Quantity Effects on Tropical Timber Production," *Journal of Environmental Economics and Management*, 2021, *107*, 102423.

Copeland, Brian R. and M. Scott Taylor, "Trade, Growth, and the Environment," *Journal of Economic Literature*, 3 2004, *42*, 7–71.

Covert, Thomas R., "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale," 2015. working paper.

_ **and Richard L. Sweeney**, "Relinquishing Riches: Auctions vs Informal Negotiations in Texas Oil and Gas Leasing," NBER working paper 25712 2019.

_ **and Ryan Kellogg**, "Crude by Rail, Option Value, and Pipeline Investment," NBER working paper 23855 2018.

Cullen, Joseph, "Measuring the environmental benefits of wind-generated electricity," *American Economic Journal: Economic Policy*, 2013, *5* (4), 107–133.

_ , "Dynamic Response to Environmental Regulation in the Electricity Industry," *Working Paper*, 2015.

Cullen, Joseph A. and Erin T. Mansur, "Inferring carbon abatement costs in electricity markets: A revealed preference approach using the shale revolution," *American Economic Journal: Economic Policy*, 2017.

Cullen, Joseph and Stanley Reynolds, "Market Dynamics and Investment in the Electricity Sector," *Working Paper*, 2015.

Currie, Janet and Reed Walker, "What Do Economists Have to Say about the Clean Air Act 50 Years after the Establishment of the Environmental Protection Agency?," *Journal of Economic Perspectives*, 2019, *33* (4), 3–26.

Dardati, Evangelina, "Pollution permit systems and firm dynamics: How does the allocation scheme matter?*," *International Economic Review*, feb 2016, *57* (1), 305–328.

Davis, Lucas and Catherine Hausman, "Market impacts of a nuclear power plant closure," *American Economic Journal: Applied Economics*, 2016, *8* (2), 92–122.

Davis, Lucas W and Catherine Hausman, "Who will pay for legacy utility costs?," 2021.

Davis, Lucas W. and Catherine Wolfram, "Deregulation, Consolidation, and Efficiency: Evidence from US Nuclear Power," *American Economic Journal: Applied Economics*, July 2012, *4* (4), 194–225.

De Loecker, Jan and Frederic Warzynski, "Markups and Firm-Level Export Status," *American Economic Review*, oct 2012, *102* (6), 2437–71.

DeMarzo, Peter M., Ilan Kremer, and Andrzej Skrzypacz, "Bidding with Securities: Auctions and Security Design," *American Economic Review*, 2005, *95* (4), 936 – 959.

Deryugina, Tatyana, Alexander MacKay, and Julian Reif, "The long-run dynamics of electricity demand: Evidence from municipal aggregation," *American Economic Journal: Applied Economics*, 2020.

Dixit, Avinash K. and Robert S. Pindyck, *Investment Under Uncertainty*, Princeton, NJ:: Princeton University Press, 1994.

Donna, Javier and Jose-Antonio Espin-Sanchez, "Complements and Substitutes in Sequential Auctions: The Case of Water Auctions," *RAND Journal of Economics*, feb 2018, *49* (1).

_ and _ , "The Illiquidity of Water Markets: Efficient Institutions for Water Allocation in Southeastern Spain," *SSRN Electronic Journal*, aug 2018.

Dorsey, Jackson, Ashley Langer, and Shaun McRae, "Fueling Alternatives: Gas Station Choice, the Value of Time, and Implications for Electric Charging," 2021. working paper.

Dorsey-Palmateer, Reid, "Effects of Wind Power Intermittency on Generation and Emissions," *The Electricity Journal*, 2019, *32* (3), 25–30.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india," *Quarterly Journal of Economics*, 2013.

_ , _ , _ , and _ , "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India," *Econometrica*, 2018.

Duso, Tomaso and Florian Szücs, "Market power and heterogeneous pass-through in German electricity retail," *European Economic Review*, 2017, *98* (C), 354–372.

Ederington, Josh, Arik Levinson, and Jenny Minier, "Footloose and pollution-free," *Review of Economics and Statistics*, 2005, *87* (1), 92–99.

Eisenberg, Thomas, "Regulatory Distortions and Capacity Investment: The Case of China's Coal Power Industry," *Mimeo*, 2020.

Ellerman, A. Denny and Juan Pablo Montero, "The efficiency and robustness of allowance banking in the U.S. Acid Rain Program," *Energy Journal*, 2007, *28* (4), 47–71.

Environmental Protection Agency, "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2019," 2021. Washington, D.C. https://www.epa.gov/sites/production/files/2021-02/documents/us-ghg-inventory-2021-main-text.pdf.

Fabra, N. and M. Reguant, "Pass-through of emissions costs in electricity markets," *American Economic Review*, 2014, *104* (9).

Fabra, Natalia, David Rapson, Mar Reguant, and Jingyuan Wang, "Estimating the Elasticity to Real Time Pricing: Evidence from the Spanish Electricity Market," *American Economic Review Papers and Proceedings*, 2021.

Fabrizio, Kira R., Nancy L. Rose, and Catherine D. Wolfram, "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency," *American Economic Review*, September 2007, *97* (4), 1250–1277.

Fan, Ying, "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 2013, *103* (5), 1598 – 1628.

Feger, Fabian, Nicola Pavanini, and Doina Radulescu, "Welfare and Redistribution in Residential Electricity Markets with Solar Power," *Working paper*, 2017.

Fell, Harrison, Daniel T. Kaffine, and Kevin Novan, "Emissions, Transmission, and the Environmental Value of Renewable Energy," *American Economic Journal: Economic Policy*, 2021.

Fetter, T. Robert, Andrew L. Steck, Christopher Timmins, and Douglas H. Wrenn, "Learning by Viewing? Social Learning, Regulatory Disclosure, and Firm Productivity in Shale Gas," 2020. working paper.

Fowlie, M., M. Reguant, and S.P. Ryan, "Market-based emissions regulation and industry dynamics," *Journal of Political Economy*, 2016, *124* (1).

Fowlie, Meredith, "Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement," *American Economic Review*, June 2010, *100* (3), 837–69.

_ and Jeffrey M. Perloff, "Distributing Pollution Rights in Cap-and-Trade Programs: Are Outcomes Independent of Allocation?," *Review of Economics and Statistics*, dec 2013, *95* (5), 1640–1652.

_ and Mar Reguant, "Challenges in the Measurement of Leakage Risk," *AEA Papers and Proceedings*, may 2018, *108*, 124–129.

_ and Nicholas Muller, "Market-Based Emissions Regulation When Damages Vary across Sources: What Are the Gains from Differentiation?," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (3), 593–632.

_ , Catherine Wolfram, Annika Todd, Patrick Baylis, and Peter Cappers, "Default Effects and Follow-On Behavior: Evidence from an Electricity Pricing Program," *Review of Economic Studies*, 2020.

Fowlie, Meredith L. and Mar Reguant, “Mitigating Emissions Leakage in Incomplete Carbon Markets,” *Journal of the Association of Environmental and Resource Economists*, 2021, *forthcomin.*

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari, “A simple estimator for the distribution of random coefficients,” *Quantitative Economics*, 2011.

Ganapati, Sharat, Joseph S. Shapiro, and Reed Walker, “Energy Cost Pass-Through in US Manufacturing: Estimates and Implications for Carbon Taxes,” *American Economic Journal: Applied Economics*, apr 2020, *12* (2), 303–42.

Gerarden, Todd, “Demanding Innovation: The Impact of Consumer Subsidies on Solar Panel Production Costs,” Technical Report, Cornell University 2019.

Gillingham, Kenneth T., Sébastien Houde, and Arthur A. Van Benthem, “Consumer Myopia in Vehicle Purchases: Evidence from a Natural Experiment,” *American Economic Journal: Economic Policy*, forthcoming.

Goettler, Ronald L. and Brett R. Gordon, “Does AMD Spur Intel to Innovate More?,” *Journal of Political Economy*, 2011, *119*, 1141 – 1200.

Goldberg, Pinelopi Koujianou, “The Effects of the Corporate Average Fuel Efficiency Standards in the US,” *Journal of Industrial Economics*, 1998, *46* (1), 1 – 33.

Gowrisankaran, Gautam, Stanley S Reynolds, and Mario Samano, “Intermittency and the value of renewable energy,” *Journal of Political Economy*, 2015, pp. 1–52.

Green, R. J. and D. M. Newbery, “Competition in the British electricity spot market,” *Journal of Political Economy*, 1992.

Grigolon, Laura, Mathias Reynaert, and Frank Verboven, “Consumer Valuation of Fuel Costs and Tax Policy: Evidence from the European Car Market,” *American Economic Journal: Economic Policy*, 2018, *10* (3), 193 – 225.

Groote, Olivier De and Frank Verboven, “Subsidies and Time Discounting in New Technology Adoption: Evidence from Solar Photovoltaic Systems,” *American Economic Review*, June 2019, *109* (6), 2137–72.

Hagerty, Nick, “Liquid Constrained in California: Estimating the Potential Gains from Water Markets,” *Unpublished Ph.D thesis*, 2017, pp. 1–75.

Hahn, Robert W. and Robert N. Stavins, “The Effect of Allowance Allocations on Cap-and-Trade System Performance,” *The Journal of Law and Economics*, nov 2011, *54* (S4), S267–S294.

Haile, Philip, Kenneth Hendricks, and Robert Porter, “Recent U.S. Offshore Oil and Gas Lease Bidding: A Progress Report,” *International Journal of Industrial Organization*, 2010, *28*, 390–396.

Hall, Robert, “Market Structure and Macroeconomic Fluctuations,” *Brookings Papers on Economic Activity*, 1986, *17* (2), 285–338.

Hamilton, James D., "Historical Oil Shocks," in Randall E. Parker and Robert Whaples, eds., *Routledge Handbook of Major Events in Economic History*, New York: Routledge Taylor and Francis Group, 2013.

Harding, Matthew and Steven Sexton, "Household Response to Time-Varying Electricity Prices," *Annual Review of Economics*, October 2017, *9* (1), 337–359.

Hausman, Catherine, "Corporate Incentives and Nuclear Safety," *American Economic Journal: Economic Policy*, 2014, *6*, 178–206.

Heckman, James J., "Heterogeneity and State Dependence," in Sherwin Rosen, ed., *Studies in Labor Markets*, University of Chicago Press, 1981.

Hendricks, Ken and Robert H. Porter, "An Empirical Perspective on Auctions," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, North Holland, 2007.

_ , _ , and Bryan Boudreau, "Information, Returns, and Bidding Behavior in Ocs Auctions, 1954–1969," *Journal of Industrial Economics*, 1987, *35* (4), 517 – 542.

_ , _ , and Guofo Tan, "Optimal Selling Strategies for Oil and Gas Leases with an Informed Buyer," *American Economic Review Papers and Proceedings*, 1993, *83* (2), 234–239.

Hendricks, Kenneth and Dan Kovenock, "Asymmetric Information, Information Externalities, and Efficiency: the Case of Oil Exploration," *RAND Journal of Economics*, 1989, *20* (2), 164 – 182.

_ and Robert H. Porter, "The Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts," *American Economic Review*, 1996, *86* (3), 388 – 407.

Herfindahl, Orris C., "Depletion and Economic Theory," in Mason Gaffney, ed., *Extractive Resources and Taxation*, University of Wisconsin Press, 1967.

Herrnstadt, Evan M., Ryan Kellogg, and Eric Lewis, "The Economics of Time-Limited Development Options: The Case of Oil and Gas Leases," NBER working paper 27165 2020.

Ho, Jacqueline S., Jhih-Shyang Shih, Lucija Muehlenbachs, Clayton Munnings, and Alan J. Krupnick, "Managing Environmental Liability: An Evaluation of Bonding Requirements for Oil and Gas Wells in the United States," *Environmental Science and Technology*, 2018, *52*, 3908 – 3916.

Hodgson, Charles, "Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry," 2018. working paper.

Hopenhayn, Hugo A., "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992.

Hortaçsu, Ali and Steven L. Puller, "Understanding strategic bidding in multi-unit auctions: A case study of the Texas electricity spot market," *RAND Journal of Economics*, 2008, *39* (1), 86–114.

_ , Fernando Luco, Steven L. Puller, and Dongni Zhu, "Does strategic ability affect efficiency? Evidence from electricity markets," *American Economic Review*, 2019.

_ , Seyed Ali Madanizadeh, and Steven L. Puller, “Power to choose? An analysis of consumer inertia in the residential electricity market,” *American Economic Journal: Economic Policy*, 2017.

Hotelling, Harold, “The Economics of Exhaustible Resources,” *Journal of Political Economy*, 1931, *39* (2), 137 – 175.

Hsiao, Allan, “Coordination and Commitment in International Climate Action: Evidence from Palm Oil,” *Working Paper*, 2020.

Hughes, Jonathan E., “The Higher Price of Cleaner Fuels: Market Power in the Rail Transport of Fuel Ethanol,” *Journal of Environmental Economics and Management*, 2011, *62* (2), 123 – 139.

_ and Ian Lange, “Who (Else) Benefits from Electricity Deregulation? Coal Prices, Natural Gas, and Price Discrimination,” *Economic Inquiry*, 2020, *58* (3), 1053 – 1075.

Hyne, Norman J., *Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production*, 2nd ed., Tulsa, Oklahoma: PennWell Corporation, 2001.

Ifrach, Bar and Gabriel Y. Weintraub, “A Framework for Dynamic Oligopoly in Concentrated Industries,” *The Review of Economic Studies*, 09 2016, *84* (3), 1106–1150.

Ito, K. and M. Reguant, “Sequential markets, market power, and arbitrage,” *American Economic Review*, 2016, *106* (7), 1921–1957.

Ito, Koichiro, “Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing,” *American Economic Review*, 2014, *104* (2), 537–563.

_ and James M. Sallee, “The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards,” *Review of Economics and Statistics*, 2018, *100* (2), 319 – 336.

_ , Takanori Ida, and Makoto Tanaka, “Moral Suasion and economic incentives: Field experimental evidence from energy demand,” *American Economic Journal: Economic Policy*, 2018.

Jacobsen, Mark R., “Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity,” *American Economic Journal: Economic Policy*, 2013, *5* (2), 148 – 187.

_ and Arthur A. van Benthem, “Vehicle Scrappage and Gasoline Policy,” *American Economic Review*, 2015, *105* (3), 1312 – 1338.

Jessoe, Katrina and David Rapson, “Knowledge is (Less) power: Experimental evidence from residential energy use,” *American Economic Review*, 2014.

_ and _ , “Commercial and Industrial Demand Response Under Mandatory Time-of-Use Electricity Pricing,” *Journal of Industrial Economics*, 2015.

Jha, Akshaya and Frank A. Wolak, “Can Forward Commodity Markets Improve Spot Market Performance? Evidence from Wholesale Electricity,” *SSRN Electronic Journal*, 2020.

Joskow, Paul and Jean Tirole, "Reliability and competitive electricity markets," *RAND Journal of Economics*, 2007.

Joskow, Paul L., "Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies," *American Economic Review*, 2011, *101* (3), 238 – 41.

_ , "Challenges for wholesale electricity markets with intermittent renewable generation at scale: The US experience," *Oxford Review of Economic Policy*, 2019.

_ **and Jean Tirole**, "Transmission Rights and Market Power on Electric Power Networks," *The RAND Journal of Economics*, 2000, *31* (3), 450.

Kaffine, Daniel T., Brannin J. McBee, and Jozef Lieskovsky, "Emissions savings from wind power generation in Texas," *Energy Journal*, 2013.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues, "Linear IV regression estimators for structural dynamic discrete choice models," *Journal of Econometrics*, may 2021, *222* (1), 778–804.

Kang, Karam and Bernardo S Silveira, "Understanding Disparities in Punishment: Regulator Preferences and Expertise," *Working Paper*, 2018.

Karaduman, Omer, "Economics of Grid-Scale Energy Storage," *Mimeo*, 2020.

Kellogg, Ryan, "Learning by Drilling: Inter-Firm Learning and Relationship Persistence in the Texas Oilpatch," *Quarterly Journal of Economics*, 2011, *126* (4), 1961 – 2004.

_ , "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review*, 2014, *104* (6), 1698 – 1734.

Kilian, Lutz and Robert J. Vigfusson, "The Role of Oil Price Shocks in Causing U.S. Recessions," *Journal of Money, Credit, and Banking*, 2017, *49* (8), 1747 – 1776.

Kim, Harim, "Heterogeneous Impacts of Cost Shocks, Strategic Bidding, and Pass-Through: Evidence from the New England Electricity Market," *American Economic Journal: Microeconomics*, 2021.

Klemperer, Paul D. and Margaret A. Meyer, "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, 1989.

Kleven, Henrik Jacobsen, "Bunching," *Annual Review of Economics*, 2016, *8*, 435 – 464.

Klier, Thomas and Joshua Linn, "New-vehicle Characteristics and the Cost of the Corporate Average Fuel Economy Standard," *RAND Journal of Economics*, 2012, *43* (1), 186 – 213.

Knittel, Christopher R., "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector," *American Economic Review*, 2011, *101* (7), 3368 – 3399.

Kolstad, Jonathan and Frank Wolak, "Using Environmental Emissions Permit Prices to Raise Electricity Prices : Evidence from the California Electricity Market," *Center for the Study of Energy Markets*, 2008, (August), 1–62.

Kong, Yunmi, "Not Knowing the Competition: Evidence and Implications for Auction Design," *RAND Journal of Economics*, 2020, *51* (3), 840 – 867.

_ , "Sequential Auctions with Synergy and Affiliation Across Auctions," *Journal of Political Economy*, forthcoming.

_ , Isabelle Perrigne, and Quang Vuong, "Multidimensional Auctions of Contracts: An Empirical Analysis," 2020. working paper.

Krautkraemer, Jeffrey A., "Nonrenewable Resource Scarcity," *Journal of Economic Literature*, 1998, *36* (4), 2065 – 2107.

Kwoka, John, "Restructuring the U.S. Electric Power Sector: A Review of Recent Studies," *Review of Industrial Organization*, 2008, *32* (3/4), 165–196.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94* (3), 614 – 641.

_ and _ , *A Theory of Incentives in Procurement and Regulation* number 0262121743. In 'MIT Press Books.', The MIT Press, August 1993.

Langer, Ashley and Derek Lemoine, "Designing Dynamic Subsidies to Spur Adoption of New Technologies," Technical Report, University of Arizona 2018.

Levinson, Arik and M. Scott Taylor, "Unmasking the pollution haven effect," *International Economic Review*, feb 2008, *49* (1), 223–254.

Levitt, Clinton J., "Information Spillovers in Onshore Oil and Gas Exploration," *Resource and Energy Economics*, 2016, *45*, 80 – 98.

Lewis, Eric, "Patchwork Policies, Spillovers, and the Search for Oil and Gas," *American Economic Journal: Economic Policy*, 2019, *11* (1), 380 – 405.

Lewis, Matthew S. and Howard P. Marvel, "When Do Consumer Search?," *Journal of Industrial Economics*, 2011, *59* (3), 457 – 483.

Li, Jing, "Compatibility and Investment in the U.S. Electric Vehicle Market," 2019. working paper.

Li, Shanjun, Lang Ton, Jianwei Xing, and Yiyi Zhou, "The Market for Electric Vehicles: Indirect Network Effects and Policy Design," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (1), 89 – 133.

Libecap, Gary D., "Institutional Path Dependence in Climate Adaptation: Coman's "Some Unsettled Problems of Irrigation"," *American Economic Review*, feb 2011, *101* (1), 64–80.

Lim, Claire S. H. and Ali Yurukoglu, "Dynamic Natural Monopoly Regulation: Time Inconsistency, Moral Hazard, and Political Environments," *Journal of Political Economy*, 2018, *126* (1), 263–312.

Lin, Cynthia C.-Y., "Strategic Decision-Making with Information and Extraction Externalities: a Structural Model of the Multistage Investment Timing Game in Offshore Petroleum Production," *Review of Economics and Statistics*, 2013, *95* (5), 1601 – 1621.

Linn, Joshua and Kristen McCormack, "The roles of energy markets and environmental regulation in reducing coal-fired plant profits and electricity sector emissions," *RAND Journal of Economics*, 2019.

Lyubich, Eva, Joseph Shapiro, and Reed Walker, "Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Environmental Policy," *AEA Papers and Proceedings*, may 2018, *108*, 136–42.

Magnac, Thierry and David Thesmar, "Identifying dynamic discrete decision processes," *Econometrica*, 2002.

Mahadevan, Meera, "The Price of Power: Costs of Political Corruption in Indian Electricity," 2021. working paper.

Malik, Kabir, Maureen Cropper, Alexander Limonov, and Anoop Singh, "The Impact of Electricity Sector Restructuring on Coal-fired Power Plants in India," *The Energy Journal*, 2015, *36* (4), 287–312.

Mansur, Erin T, "Measuring Welfare in Restructured Electricity Markets," *Review of Economics and Statistics*, 2008, *90* (2), 369–386.

Mansur, Erin T. and Matthew W. White, "Market Organization and Efficiency in Electricity Markets," *Mimeo*, 2012, pp. 1–56.

Maskin, Eric and Jean Tirole, "A Theory of Dynamic Oligopoly II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 1988, *56* (3), 571 – 599.

Mercadal, Ignacia, "Dynamic competition and arbitrage in electricity markets: The role of financial players," *American Economic Journal: Microeconomics*, 2021.

Milgrom, Paul R. and Robert J. Weber, "A Theory of Auctions and Competitive Bidding," *Econometrica*, 1982, *50* (5), 1089–1122.

Miravete, Eugenio J., María J. Moral, and Jeff Thurk, "Fuel Taxation, Emissions Policy, and Competitive Advantage in the Diffusion of European Diesel Automobiles," *RAND Journal of Economics*, 2018, *49* (3), 504 – 540.

Muehlegger, Erich and Richard L. Sweeney, "Pass-Through of Own and Rival Cost Shocks: Evidence from the U.S. Fracking Boom," *National Bureau of Economic Research*, nov 2017.

Muehlenbachs, Lucija, “A Dynamic Model of Cleanup: Estimating Sunk Costs in Oil and Gas Production,” *International Economic Review*, 2015, *56* (1), 155 – 185.

Muller, Nicholas Z. and Robert Mendelsohn, “Efficient Pollution Regulation: Getting the Prices Right,” *American Economic Review*, dec 2009, *99* (5), 1714–39.

Nemet, Gregory F., “Beyond the learning curve: factors influencing cost reductions in photovoltaics,” *Energy Policy*, 2006.

_ , “Subsidies for New Technologies and Knowledge Spillovers from Learning by Doing,” *Journal of Policy Analysis and Management*, 2012.

Newell, Richard G. and Brian C. Prest, “The Unconventional Oil Supply Boom: Aggregate Price Response from Microdata,” *Energy Journal*, 2019, *40* (3), 1 – 30.

_ , _ , and Ashley B. Vissing, “Trophy Hunting versus Manufacuring Energy: The Price Responsiveness of Shale Gas,” *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (2), 391 – 431.

Nguyen, Anh, “Information Design in Common Value Auction with Moral Hazard: Application to OCS Leasing Auctions,” 2021. working paper.

Novan, Kevin, “Valuing the Wind: Renewable Energy Policies and Air Pollution Avoided,” *American Economic Journal: Economic Policy*, 2015, *7* (3), 291–326.

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

Olley, G. Steven and Ariel Pakes, “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, nov 1996, *64* (6), 1263.

O’Neill, Richard P., Paul M. Sotkiewicz, Benjamin F. Hobbs, Michael H. Rothkopf, and William R. Stewart, “Efficient market-clearing prices in markets with nonconvexities,” *European Journal of Operational Research*, July 2005, *164* (1), 269–285.

Ordin, Andrey, “Investment and Taxation: the Case of Oil and Gas in the Permian Basin,” 2019. working paper.

Ostrovsky, Michael and Michael Schwarz, “Carpooling and the Economics of Self-Driving Cars,” 2018. working paper.

Pakes, Ariel, Michael Ostrovsky, and Steven Berry, “Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples),” *RAND Journal of Economics*, 2007, *38*, 373 – 399.

Patel, April and Emily Geary, “U.S. crude oil and natural gas production in 2019 hit records with fewer rigs and wells,” *EIA Today In Energy*, June 25 2020.

Pindyck, Robert S., "The Optimal Exploration and Production of Nonrenewable Resources," *Journal of Political Economy*, 1978, *86* (5), 841 – 861.

_ , "The Dynamics of Commodity Spot and Futures Markets: A Primer," *Energy Journal*, 2001, *22* (3), 1 – 29.

Porter, Robert H., "The Role of Information in U.S. Offshore Oil and Gas Lease Auction," *Econometrica*, 1995, *63* (1), 1 – 27.

Preonas, Louis, "Market Power in Coal Shipping and Implications for U.S. Climate Policy," 2019. working paper.

Puller, Steven L., "Pricing and Firm Conduct in California's Deregulated Electricity Market," *Review of Economics and Statistics*, 2007, *89* (1).

Qiu, Chenyu, "The Value of Wind Information in Wholesale Electricity Market: Evidence from U.S. Midwest," 2020. working paper.

Rafey, Will, "Droughts, deluges, and (river) diversions: Valuing market-based water reallocation," *Mimeo*, 2021.

Raimi, Daniel, Alan Krupnick, Jhih-Shyang Shah, and Alexandra Thompson, "Decommissioning Orphaned and Abandoned Oil and Gas Wells: New Estimates and Cost Drivers," 2021. preprint.

Reguant, M., "Complementary bidding mechanisms and startup costs in electricity markets," *Review of Economic Studies*, 2014, *81* (4).

Reguant, Mar and A. D Ellerman, "Grandfathering and the endowment effect An Assessment in the context of the Spanish National Allocation Plan," *Center for Energy and Environmental Policy Research*, 2008, *49*, 69–73.

Reiss, Peter C. and Matthew W. White, "Household electricity demand, revisited," *Review of Economic Studies*, 2005, *72* (3), 853–883.

Reynaert, Mathias, "Abatement Strategies and the Cost of Environmental Regulation: Emission Standards on the European Car Market," *Review of Economic Studies*, forthcoming.

_ **and James M. Sallee**, "Who Benefits When Firms Game Corrective Policies?," *American Economic Journal: Economic Policy*, forthcoming.

Riley, John G., "Ex Post Information in Auctions," *Review of Economic Studies*, 1988, *55* (3), 409 – 429.

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987, *55* (5), 999 – 1033.

Ryan, Nicholas, "The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market," *American Economic Journal: Microeconomics*, 2021.

Ryan, Stephen P., "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 2012, *80* (3), 1019–1061.

Sallee, James M., Sarah E. West, and Wei Fan, "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations," *Journal of Public Economics*, 2016, *135*, 61 – 73.

Salop, Steven C., "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 1979, *10*, 141 – 156.

Schmalensee, Richard and Robert N. Stavins, "The SO2 allowance trading system: The ironic history of a grand policy experiment," *Journal of Economic Perspectives*, dec 2013, *27* (1), 103–122.

_ **, Paul L. Joskow, A. Denny Ellerman, Juan Pablo Montero, and Elizabeth M. Bailey**, "An Interim Evaluation of Sulfur Dioxide Emissions Trading," *Journal of Economic Perspectives*, 1998, *12* (3), 53–68.

Scott, Paul T, "Dynamic Discrete Choice Estimation of Agricultural Land Use," 2013.

Shaffer, Blake, "Misunderstanding nonlinear prices: Evidence from a natural experiment in electricity demand," *American Economic Journal: Economic Policy*, 2020, *12* (3).

Shavell, Steven, "The Judgment Proof Problem," *International Review of Law and Economics*, 1986, *6* (1), 45 – 58.

Skrzypacz, Andrzej, "Auctions with Contingent Payments—an Overview," *International Journal of Industrial Organization*, 2013, *31*, 666 – 675.

Slade, Margaret E. and Henry Thille, "Whither Hotelling: Tests of the Theory of Exhaustible Resources," *Annual Review of Resource Economics*, 2009, *1* (1), 239 – 259.

Small, Kenneth A., "Valuation of Travel Time," *Economics of Transportation*, 2012, *1*, 2 – 14.

Souza-Rodrigues, Eduardo, "Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis," *The Review of Economic Studies*, 11 2019, *86*, 2713–2744.

Springel, Katalin, "Network Externality and Subsidy Structure in Two-Sides Markets: Evidence from Electric Vehicle Incentives," *American Economic Journal: Economic Policy*, forthcoming.

Stavins, Robert, "Transaction Costs and Tradeable Permits," *Journal of Environmental Economics and Management*, 1995, *29* (2), 133–148.

Steck, Andrew L., "Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing," 2018. working paper.

Sweeney, Richard L., "Environmental Regulation, Imperfect Competition and Market Spillovers: The Impact of the 1990 Clean Air Act Amendments on the US Oil Refining Industry," 2015. working paper.

Timmins, Christopher, "Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the Arid West," *Econometrica*, 2002, *70* (2), 603–629.

Torchiana, Adrian L, Ted Rosenbaum, Paul T Scott, and Eduardo Souza-Rodrigues, "Improving Estimates of Transitions from Satellite Data: A Hidden Markov Model Approach *," 2020.

Toyama, Yuta, "Dynamic Incentives and Permit Market Equilibrium in Cap-and-Trade Regulation," *Mimeo*, 2019.

Varian, Hal R., "A Model of Sales," *American Economic Review*, 1980, *70*, 651 – 659.

Vreugdenhil, Nicholas, "Booms, Busts, and Mismatch in Capital Markets: Evidence from the Offshore Oil and Gas Industry," 2020. working paper.

_ , "Incomplete Regulation and Spatial Equilibrium in the Offshore Oil and Gas Industry," 2021. working paper.

Ward, Melissa M. and Timothy W. Kelsey, "Natural Gas Lessors' Experiences in Bradford and Tioga Counties, 2010," 2011. Marcellus Education Fact Sheet, PennState Extension.

Whitefoot, Kate S., Meredith L. Fowlie, and Steven J. Skerlos, "Influence of Acceleration Tradeoffs on CO2 Emissions and Costs of Fuel Economy and Greenhouse Gas Regulations," *Environmental Science and Technology*, 2017, *51* (18), 10307 – 10315.

Wolak, Frank, "Designing Nonlinear Price Schedules for Urban Water Utilities to Balance Revenue and Conservation Goals," *National Bureau of Economic Research*, aug 2016.

Wolak, Frank A., "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market *," *International Economic Journal*, 2000, *14* (2), 1–39.

_ , *Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Competitive Electricity Markets*, Cambridge University Press, 2003.

_ , "Quantifying the supply-side benefits from forward contracting in wholesale electricity markets," *Journal of Applied Econometrics*, 2007, *22* (7), 1179–1209.

_ , "Measuring the competitiveness benefits of a transmission investment policy: The case of the Alberta electricity market," *Energy Policy*, 2015.

Wolak, Frank A, "Long-Term Resource Adequacy in Wholesale Electricity Markets With Significant Intermittent Renewables," 2021.

Wolfram, Catherine D., "Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales," *The RAND Journal of Economics*, jan 1998, *29* (4), 703.

_ , "Measuring duopoly power in the British electricity spot market," *American Economic Review*, 1999, *89* (4), 805–826.

Zou, Eric Yongchen, "Unwatched pollution: The effect of intermittent monitoring on air quality," *American Economic Review*, jul 2021, *111* (7), 2101–2126.