ABSTRACT

Title of Dissertation: DYNAMICS OF CAPITAL INVESTMENT AND POLLUTION EXTERNALITIES IN WHOLESALE ELECTRICITY MARKETS

Christopher Holt
Doctor of Philosophy, 2022

Dissertation Directed by: Professor Joshua Linn
Department of Agricultural and Resource Economics

The field of environmental economics was built on the notion of internalizing into markets the social harm caused by pollution. This dissertation examines the implications of putting that idea into practice in the electric power generation sector, with a particular focus on market structure and short- and long-run industry dynamics.

Environmental policy to mitigate climate change seeks to transform the capital composition of industries for the purpose of reducing carbon dioxide emissions. In deregulated wholesale electricity markets, firms may exercise long-run market power through investment and retirement decisions which affect future wholesale price settlement. In Chapter 1, I develop a dynamic structural model of the Texas electricity market spanning 2003-2019 to analyze how long-run market power exercise and environmental policy for reducing carbon emissions affect the capital composition of the industry over time. I find that market power exercise led to significant early fossil fuel plant retirements over this period, with an attendant decrease in consumer surplus on

the order of $1.6 billion annually. Further counterfactual analysis suggests that federal production tax credits for wind power expanded the deployment of wind by approximately 73 percent, but the associated reductions in emissions were more than twice as costly as would have been achieved under a $20-per-ton carbon tax.

In Chapter 2 I delve further into the issue of market structure and long-run dynamics. Economic theory suggests that setting the wholesale price of electricity at the marginal social cost of unmet demand during periods of scarcity results in optimal capacity investment in the presence of perfect competition. I examine the implications of applying this principle in a setting where competition is imperfect, and where the market was structured prior to the introduction of competition (deregulation) and therefore not established through firms' profit maximizing behavior. I build a stylized model that approximates the effects of a scarcity price mechanism under the hourly demand distribution observed in the Texas wholesale electricity market in 2017. I use this model to demonstrate that the scarcity price mechanism may encourage incumbent firms with large portfolios to retire plants, and that firms with a threshold amount of existing infra-marginal generation capacity will be unwilling to invest in new capacity. I then use a dynamic structural model to demonstrate that the scarcity price mechanism introduced in Texas in 2014 accelerated turnover over the period 2014-2019 by driving greater retirement of capacity in addition to greater investment, relative to a counterfactual scenario in which the scarcity price design was not implemented.

In Chapter 3 I shift my focus from long-run industry dynamics and environmental policy concerning a global pollutant (carbon dioxide) to short-run dynamics and harm from a local pollutant (ground-level ozone). NOx emissions are a precursor to ground-level ozone, a pernicious pollutant that is harmful to human health and ecosystems. Despite decades of regulations includ-

ing NOx emissions pricing, and a corresponding precipitous decline in NOx emissions, episodic high-ozone events prevent many areas from achieving air quality standards. Theoretically, spatially or temporally differentiated emissions prices could be more cost-effective at reducing such events than a uniform price. To test this prediction, using data from the EPA and NOAA spanning 2001-2019, we use novel empirical strategies to estimate (1) the link between hourly emissions and high-ozone events and (2) hourly marginal abatement costs. The estimates form the basis for simulations that compare uniform and differentiated emissions pricing. Consistent with economic theory, differentiated emissions pricing is substantially more cost effective at reducing high-ozone events, but this advantage depends on the accuracy of the estimated NOx-ozone relationship.

DYNAMICS OF CAPITAL INVESTMENT AND POLLUTION EXTERNALITIES IN WHOLESALE ELECTRICITY MARKETS

by

Christopher Holt

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Joshua Linn, Chair/Advisor
Professor James Archsmith
Professor Louis Preonas
Professor Chenyu Yang
Professor Lucy Qiu

Foreword

The third chapter of this dissertation is a jointly authored work with Joshua Linn. The Dissertation Committee acknowledges that Christopher Holt made substantial contributions to the relevant aspects of the chapter.

Dedication

For my mother, who drove me to math class, and my father, who taught me how to hypothesize.

Acknowledgments

This dissertation was made possible by the steady and sure guidance of my advisor Josh Linn. I am incredibly grateful for his unyielding patience through many meetings and phone conversations, and his careful, sincere, and thoughtful delivery of sharp insights.

Numerous conversations with and advising from Jim Archsmith, Louis Preonas, and Rob Williams were also invaluable and enlightening. Discussions with Andrew Sweeting and Chenyu Yang were instrumental in helping me to gain a perspective on the field of empirical industrial organization and understand the methods.

The first two chapters of this dissertation were inspired by meetings and conversations I had during my pre-doctoral fellowship at the Environmental Defense Fund in 2019. I am deeply indebted to Kristina Mohlin and Beia Spiller for their excellent mentorship during my time there.

I would also like to thank Anna Alberini, Erich Battistin, Robert Chambers, Jorge Holzer, Ethan Kaplan, Lori Lynch, Lars Olson, and Lucy Qiu for their support of my endeavors during my doctorate program; and Dany Burns, Katherine Faulkner, and Pin-Wuan Lin for their tireless administrative support.

A special thanks to friends and colleagues Blake Baird, Andrew Card, Aldo Gutiérrez-Mendieta, Anna Koyfman, Robert Kulick, Srikant Narasimhan, Eric Reese, Billy Schwartz, Haoluan Wang, David Wescott, and many others at the University of Maryland and elsewhere, and to my family, who have all inspired and encouraged my work over the past several years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AQS	Air Quality System
CEMS	Continuous Emissions Monitoring System
CREZ	Competitive Renewable Energy Zones
CSAPR	Cross-State Air Pollution Rule
EIA	Energy Information Administration
EPA	Environmental Protection Agency
ERCOT	Electricity Reliability Council of Texas
ISD	Integrated Surface Database
MW	megawatts
MWh	megawatt hours
NAAQS	National Ambient Air Quality Standards
NOAA	National Oceanic and Atmospheric Administration
NOx	nitrogen oxides
ppb	parts per billion
PTC	Production Tax Credit
PUCT	Public Utility Commission of Texas
SPM	scarcity price mechanism
VOCs	volatile organic compounds

Chapter 1: Long-Run Dynamics of Industry Transition: A Structural Model of the Texas Electricity Market 2003-2019

1.1 Introduction

Environmental policy to mitigate climate change seeks to direct the transition of entire industries to comprise of low-carbon technologies. In deregulated markets, firms make capital investment or retirement decisions based on the profitability of those choices over time. Effective environmental policy which seeks to reconfigure the capital composition of an industry must therefore consider the perspective of the firm, and the industry dynamics inherent to firm investment and retirement decisions.

Optimal environmental policy to address a pollution externality can be achieved under perfect competition by pricing pollution at its marginal social cost (Pigou, 1924). It is well established that this result, and the cost effectiveness of externality pricing, is complicated by the presence of market power (Buchanan, 1969). There remains a large gap in the literature when it comes to assessing the interaction of an externality-correcting policy and distortions from market power.

Moreover, a dynamic market structure—inherently of central importance to policy that is designed to induce industry change over time—can affect environmental policy outcomes in sev-

eral ways. Regardless of industry concentration, environmental policy can shift production both at the intensive margin (holding capital fixed) and at the extensive margin over time (allowing for capital adjustments), each effect involving different environmental consequences. Evolving economic costs associated with entry and exit may lead to equilibrium capacity levels and technological compositions that change over time. Distortions from market power, as I will demonstrate, further shape the future industry path and its environmental effects.

My paper examines market power and environmental policy in the context of wholesale electric power generation and over what I will call the "long-run," the time horizon over which firms make capital investment and retirement decisions. The electric power generation industry is the most prominent example of an industry that is undergoing a transition to low-carbon capital. It was responsible for 32 percent of energy-related carbon emissions in the United States in 2020 (EIA, 2020), and due to the increasing availability of renewable energy technologies such as wind and solar, decarbonization of this sector seems achievable. Indeed, policies throughout parts of the United States involve explicit timelines for this goal. Institutional features make market power exercise a uniquely important concern in this industry, and therefore the perspective of the firm is critical in informing effective policy. One such feature is highly inelastic demand in the short-run, which arises because consumers of electricity are not exposed to fluctuations in the wholesale price. Wholesale electricity market design also typically involves price sensitivity during periods of supply scarcity, when levels of demand strain available generation capacity. These features together imply that firms with control over any amount of generation capacity can significantly affect market prices if they are able to adjust the amount of supply they make available. In a spot market that is cleared through a uniform-price auction, withholding supply can induce higher prices which translate to greater rents associated with the portion of the firm's

capacity that is not withheld. The high degree of wholesale price sensitivity implies that a firm need not control a large share of the market for this incentive to be meaningful, therefore typical horizontal concentration measures (such as the Herfindahl-Hirschman index) are uninformative when it comes to understanding the potential for market power exercise (Borenstein, Bushnell, and Knittel, 1999).

It is with these features in mind that I characterize power generator entry and exit as a strategic dynamic game: firms choose how much generation capacity to build or retire, and these decisions (along with demand) determine future market prices for the sale of electricity. This framework is akin to a pre-commitment game; Kreps and Scheinkman, 1983 show, for example, that when producers make quantity pre-commitments prior to engaging in Bertrand-style price competition, a unique Cournot outcome prevails. In my model, quantity pre-commitments are effectively implied by capacity choices, and prices are settled according to the dispatch operations of the system.

Capacity withholding in the short run as an exercise of market power in wholesale electricity markets has been studied extensively. A prominent example is Borenstein, Bushnell, and Wolak, 2002, who decompose wholesale electricity payments in California, attributing a significant portion to market power exercise. Woerman, 2019 examines the issue in the Texas electricity market with a focus on ephemeral episodes of localized transmission congestion which create opportunities for plant operators to extract rents. As I will demonstrate, the same fundamental principles can create strategic incentives over longer time horizons, a phenomenon which is not well studied but carries significant welfare consequences. For example, a firm with a portfolio made up of several power plants may be in a position to increase its profits from retiring a plant, if doing so elevates prices such that it earns greater profits on its remaining active generation

assets. This incentive works in both directions: a large firm with many plants may choose not to invest in new generation capacity if doing so significantly lowers wholesale prices, thereby also reducing profits accruing to its existing portfolio. I choose to focus on this long-run variety of market power, and my modeling does not endogenize strategic behavior in the short-run. Short-run market power is monitored and mitigated in my setting by both the system operator and an independent market monitor, limiting its effect. This modeling choice also allows clearer focus on long-run behavior, a significant market distortion which need not involve the ability to behave strategically in spot markets.

The setting for my analysis is the market operated by the Electricity Reliability Council of Texas ("ERCOT"), a system which comprises the flow of electricity throughout most of the state of Texas, over the period 2003-2019 (I will use "ERCOT" and "Texas electricity market" interchangeably). Texas is uniquely suited for an empirical investigation into the issues at hand for three reasons. First, it is mostly isolated from the rest of the United States' power grid, enabling study of a single market free from complications which may arise from intra-regional differences in regulation, or spillover effects associated with interconnected markets. Second, Texas has adopted an "energy only" market design, in which revenues earned in the sale of wholesale electricity in spot markets are the primary driver of cost recovery for investors; in contrast, many systems within the United States operate a distinct market for capacity. The energy only design draws directly from economic theory of optimal investment. This makes the market naturally suited for economic modeling, and the empirical conclusions of interest in testing the efficacy of the market design. Third, Texas is a leader in wind power generation. Wind power in Texas expanded from less than one percent of the generation mix to over 20 percent from 2002-2019. Much of this increase is attributable to federal production tax credits

(PTCs) which compensate investors for each unit of electricity produced from wind power. This large expansion makes Texas uniquely suited for testing the effects of policies and market forces on the adoption of renewable energy.

I estimate a structural model of firms' decisions to build and retire plants over time. These decisions are based on future expected profit flows from the sale of electricity. I begin by constructing a detailed data set of hourly generation from nearly all power plants in ERCOT spanning 2002-2019 and build a short-run model that simulates hourly dispatch and spot price settlement, and therefore firm profits, under various market conditions. I use this model to solve for future expected profit flows under a long-run equilibrium framework. I employ a non-stationary oblivious equilibrium concept based on Weintraub, Benkard, Jeziorski, and Roy, 2010, which allows me to model the industry transition and the individual behavior of many heterogeneous firms. Variation in future expected profit flows enables estimation of economic costs associated with plant investment and retirement. After recovering the industry cost structure I am able to perform several counterfactual analyses of interest.

A central result of this paper is that long-run market power exercise has economically significant effects on the industry path, which I define as the capital composition of installed system-wide generation capacity over time. I arrive at this result by simulating a counterfactual in which firms behave competitively (absent market power incentives), holding other factors fixed. I find that long-run market power exercise resulted in the retirement of approximately 8,850 MW of fossil fuel generation capacity more than in a competitive counterfactual, an amount equal to nearly seven percent the entire Texas system or enough to power the entire state of Connecticut. This finding implies net welfare losses: high prices from early retirements are only partly offset by new investment, and environmental benefits from lower emissions associated with the retired

plants are relatively small.

I then perform environmental policy counterfactual simulations focused on two policy instruments: a subsidy for wind power generation, and a price on carbon emissions. I find that wind production tax credits expanded the buildout of wind power in Texas by 73 percent from 2003-2019, and that a carbon price of $20 per ton would have achieved an even greater buildout at half the welfare cost per ton of emissions reduced. Among the counterfactuals I perform, a carbon price under the competitive scenario yields the least-cost reduction in emissions. To further examine the effects of environmental policy and market power on the industry transition, I decompose emissions reductions into those attributable to changes at the extensive margin (through the changing capital composition of the generation mix) and those attributable to changes at the intensive margin (through re-arranged dispatch order). This analysis suggests that, despite large differences in plant retirements attributable to market power exercise, a $20 carbon price achieves similar extensive-margin emissions paths whether or not the market is competitive. Under a wind production subsidy, in contrast, fossil fuel capital turnover is slowed in the presence of competition, leading to lower emissions reductions relative to the concentrated scenario.

This paper makes contributions mainly to three branches of literature. First, I contribute to a literature which examines the interaction of market structure and environmental policy with an emphasis on long-run dynamics. Ryan, 2012 assesses the welfare impacts of the 1990 Clean Air Act on the Portland Cement industry using the pioneering framework of Ericson and Pakes, 1995, which applies a Markov Perfect Equilibrium concept as a tool for describing oligopolistic behavior in a dynamic setting. Fowlie, Reguant, and Ryan, 2016 extends this work in the same industry, with a focus on how policies to reduce carbon emissions play out over time. Among other results, these papers highlight that policies may exacerbate distortions from short-run mar-

ket power exercise by inducing firm exit and increasing economic barriers to entry. The ability of firms to exercise short-run market power therefore depends on long-run market structure dynamics. While I do not model short-run market power exercise (looking instead at market power exercised through investment and retirement decisions), a similar result holds in my setting: policies and fixed costs affect the equilibrium composition of the industry and how it evolves, with attendant welfare consequences.

The industry setting analyzed by Ryan, 2012 and Fowlie, Reguant, and Ryan, 2016 involves localized markets each composed of a small number of homogeneous firms, and the industry is modeled as evolving according to a stationary process. In contrast, I am interested in modeling a non-stationary industry transition involving changes in capital composition over time, as determined by the behavior of many heterogeneous agents. Methodologically, these features complicate the application of the Ericson and Pakes, 1995 entry and exit framework. A dimensionality problem arises as the number of agents, and therefore the state space, becomes large, leading to computational intractability. Moreover, the traditional framework imposes stationarity, which implies a state space that is visited infinitely many times into the future. Because I am modeling firms which are unique in their composition (made up of many power plants with unique generation attributes), exit of a plant from the market implies that the state space has changed irreversibly, and the previous state is never to be visited again. To overcome these complications I draw from the literature on oblivious equilibria (Weintraub, Benkard, and Van Roy, 2008) and the extension to a non-stationary setting (Weintraub, Benkard, Jeziorski, and Roy, 2010). Related applications include Sweeting, 2015, which applies a non-stationary oblivious equilibrium concept to a platform model of a secondary market for event ticket sales, accommodating rich player heterogeneity, and Igami, 2017, which applies a non-stationary oligopoly model to analyze

innovation in the hard disk industry over time.

Second, I contribute to a literature which examines investment behavior in restructured electricity markets. Bushnell and Ishii, 2007 develops an oligopoly framework for modeling investment decisions and draws qualitative insights in a stylized electricity market setting. Myatt, 2017 uses a two-stage structural model to demonstrate the significance of strategic incentives driving early plant retirements. Davis, Holladay, and Sims, 2021 uses a real options model to recover retirement costs for coal-fired power plants, then predicts the effects of policy on retirement timelines. Karaduman, 2020 builds a structural model of grid-scale energy storage investment in the South Australian electricity market with attention to how different ownership structures affect investment incentives through price effects. I contribute to these works by introducing an emphasis on environmental policy and changing capital composition over time, modeling an empirical setting involving many firms, and modeling both plant investment and retirement decisions.

Third, I contribute to literature examining the long-run effects of environmental policy in the electricity sector. Linn and McCormack, 2019 disentangles the effects of market forces and environmental regulations of local air pollutants on coal plant profitability, finding that market forces were the dominant factor driving retirements. Palmer and Burtraw, 2005 compare cost effectiveness of policies to encourage renewable energy generation. Cullen and Reynolds, 2017 model the effects of environmental policy on investment in a perfectly competitive dynamic setting. I complement these analyses by employing a model which accommodates imperfect competition.

1.2 Electric Power Industry Background

Several industry features unique to electricity markets motivate an analysis of long-run strategic behavior on behalf of firms. First, consumers are not exposed to fluctuations in the wholesale price of electricity, making demand inelastic in the short-run. Short-run wholesale market prices, determined through a uniform price auction, are therefore highly sensitive to the amount of supply being offered on the system, particularly when demand is high. Second, monitoring of supply offer curves and capacity utilization limits the amount of capacity that can be withheld once it is built. Third, ERCOT's scarcity pricing mechanism, designed to encourage capacity investment, implies additional price sensitivity at the extensive margin.

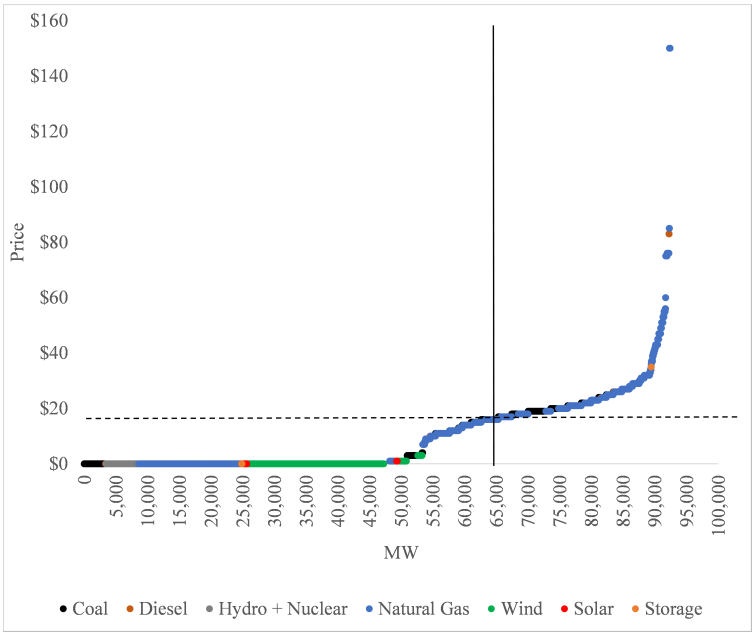
1.2.1 Short-Run Market Operation

Electricity markets are designed to minimize the cost of dispatching electricity such that supply exactly meets demand (or load, in electricity market parlance) in real time.[1] In ERCOT and other deregulated markets, power plant managers submit detailed offer curves which reflect their willingness to sell electricity at a schedule of prices throughout the day subject to the operational constraints of the generation units. Aggregating bids across all generators in the system leads to a market supply curve, illustrated in Figure 1.1. The system operator then implements a constrained optimization routine to determine dispatch according to these bids, load, local transmission constraints, and the need for ancillary services which ensure system reliability. The real-time market in ERCOT is cleared in five-minute intervals and across thousands of nodes which connect sources of generation to consumers of power through the transmission network.

[1] See Cramton, 2017 for an overview of electricity market design.

The clearing price at a point in time and space on the grid reflects the offer curve submitted by the marginal unit—the unit associated with the segment of the aggregate offer curve that intersects with demand. Hypothetical market clearing in the absence of transmission constraints is illustrated in Figure 1.1 where the vertical line represents demand and the horizontal line indicates the market clearing price.

Figure 1.1: Market-Wide Supply Offer Curve and Hypothetical Clearing

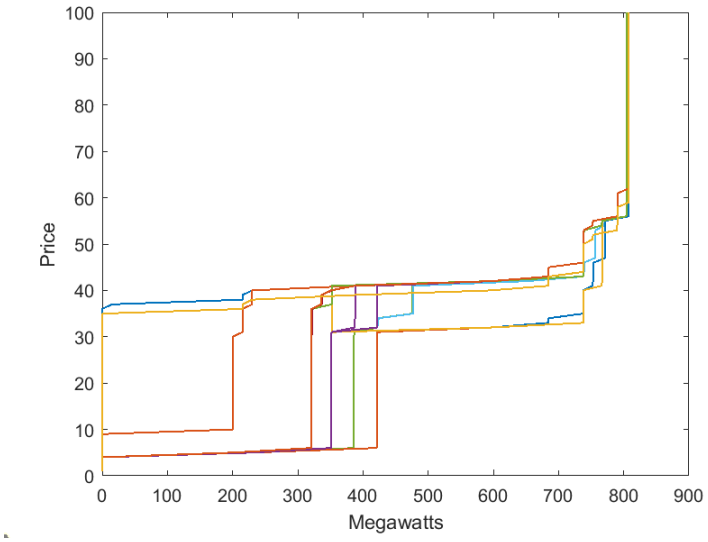


This figure depicts supply offers from July 10, 2019 at 8:30PM DST. Offers are categorized by technology type and represented in 5MW increments. The vertical line represents hypothetical demand and the horizontal line represents the clearing price that would obtain absent transmission and operational constraints. Source: ERCOT

Heterogeneous generation technologies are used to meet load that fluctuates throughout the day. Some plants, like nuclear or coal plants, are built to serve "base" load, or the portion of load that is present throughout all hours of the day. Such plants are not equipped to rapidly increase or decrease output. Wind and solar power generate intermittent energy based on when the wind

is blowing or the sun is shining. Other, typically fossil fuel based plants, are designed to meet the residual "net load," or the difference between total load and the amount of load that is met by base load generators and renewables. Natural gas "peaker" units, for example, are designed to meet demand during "peak" hours and can increase and decrease generation swiftly. Plant operators' offer curves will therefore vary according to the operational attributes of the units including marginal generation costs, ramping constraints (the plant's ability to increase or decrease generation in a given amount of time), startup and shut-down costs, and variation in temporal load patterns. A plant which requires significant ramping time may operate at a loss during periods of low demand in anticipation of earning profits during hours of peak demand when prices become high. Figure 1.2 illustrates variation in submitted offer curves for a plant composed of multiple generation units. Each curve reflects the plant's offer in a different hour of a 24-hour period.

Figure 1.2: 24-Hour Plant Supply Offer Curves, Mountain Creek Natural Gas Steam Turbine Plant



Plants set dynamic offer schedules throughout the day based on anticipated load and operational constraints.This figure depicts supply offers from the Mountain Creek natural gas steam turbine power plant on June 19, 2017. Each color represents an offer curve in a different hour of the day. The x-axis shows the amount of power the plant has offered to generate at the price shown on the y-axis. This plant has three generation units. Source: ERCOT

These dynamics also give rise to incentives to behave strategically in short-run markets. Borenstein, Bushnell, and Wolak, 2002 demonstrate that, unlike other industries, even firms with a small horizontal market presence can exercise market power by withholding supply to effectuate a higher clearing price. Certain safeguards have been put in place in response to the threat of economic harm to consumers arising from short-run market power exercise. An independent market monitor performs routine scrutiny of a measure called the "output gap," or the amount of capacity on the system which would be economically viable for generating electricity but is not dispatched. In addition, ERCOT itself imposes limits on the amount of markup above cost generators can earn as implied by their submitted offer curves, taking generator-specific costs

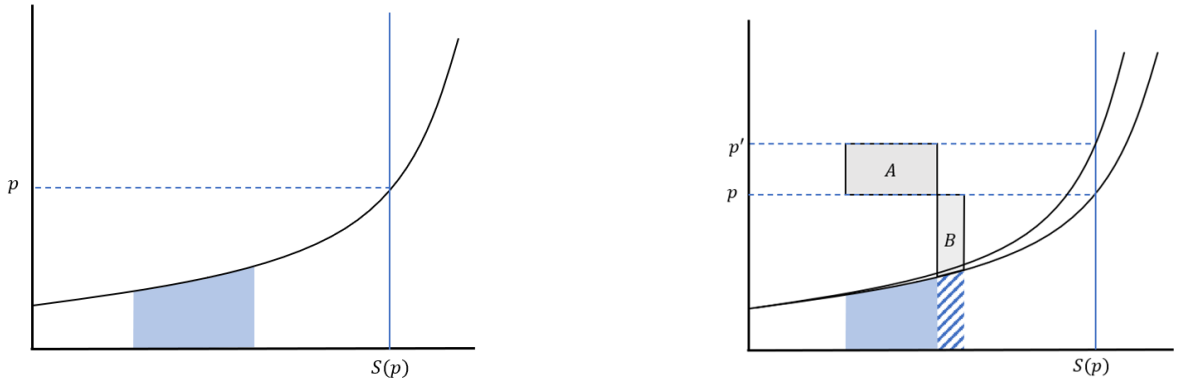
into account. If a plant operator submits an offer curve that is in excess of the pre-determined markup, ERCOT mitigates that offer curve before determining dispatch. Such safeguards only serve to limit, rather than eviscerate, firms' ability to exercise market power (see, e.g., Woerman, 2019, which assesses market power exercise arising from episodes of localized transmission congestion).

1.2.2 Long-Run Efficiency

Deregulated electricity markets are also designed to ensure there is sufficient generation capacity installed to maintain system reliability and prevent shortages. Many systems operate a distinct market for capacity, where investors submit bids to build generation capacity, and projects are approved at least cost subject to a set reliability target. Texas, in contrast, operates an "energy only" market, in which the sale of electricity is the primary means for firms to recover their investment costs.

Figure 1.3a shows an illustrative market-wide supply curve at its intersection with load and the prevailing clearing price. The shaded region represents generation assets owned by a hypothetical firm. At this level of demand, the firm is earning "infra-marginal" rents on its generation units, as the assets are below the point on the curve at which supply intersects with demand. A firm's investments will be profitable if it earns rents in excess of its investment and operational costs.

Figure 1.3: Price Effects of Plant Retirement



(a) (b)

$S(p)$ (x-axis) represents the quantity of electricity a firm offers to generate at price p (y-axis). A firm with generation assets indicated by the blue shaded region (both panels) may increase its profits by retiring generation assets if the amount of profits on remaining assets from the attendant price increase (rectangle A) exceed the lost profits from the retired assets (rectangle B).

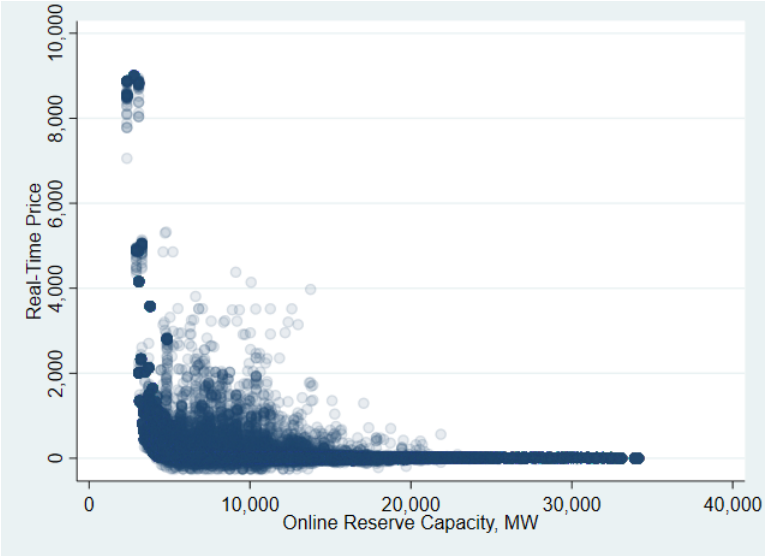
Firms with capacity that is frequently infra-marginal have an incentive to exercise market power by closing an existing plant. Figure 1.3b illustrates the change in infra-marginal rents that the hypothetical firm would earn at the same level of demand if it subtracted one of its plants from the market supply curve. Rectangle A shows the increase in rents earned by the firm's remaining portfolio, and rectangle B the lost rents from its retired plants. If the resulting change in rents (A minus B) over time exceeds the cost of closing the plant, a firm can earn more profits from retiring the plant than from keeping it operable.

Theoretically optimal capacity investment in electricity markets results from internalizing the social cost of a generation shortage. By setting the price of electricity equal to the "value of lost load" during periods when demand exceeds supply, the marginal cost of new capacity is equal to the incremental value it brings to the system and optimal investment is achieved.[2] Since

[2]See Stoft, 2002 for a treatment.

2014, this theory has been put into practice in Texas through a scarcity price mechanism. During periods when generation resources are constrained due to high levels of demand, the scarcity price is triggered. The mechanism is designed such that the price reflects the probabilistic value of a generation shortage and is capped at $9,000 per MWh, the administratively set value of lost load. Figure 1.4 illustrates this relationship.

Figure 1.4: Real-Time Price vs. Online Reserves, 2015-2019



Online Reserve Capacity (x-axis) is the amount of generation capacity availability that can be quickly dispatched to serve demand. ERCOT's scarcity price mechanism is designed to induce a real-time price (y-axis) that reflects the social cost of a potential generation shortage. Source: ERCOT

This mechanism implies high sensitivity at the extensive margin, which has critical implications when it comes to long-run market power exercise. The theoretical results which underpin efficient capacity investment are predicated on the notion of perfect competition. But in a dynamic setting, removal of a power plant may generate significant supra-competitive price effects which are only gradually offset by potential entrants. Moreover, incumbents with significant

infra-marginal assets may be disincentivized from adding capacity due to price effects, while barriers to entry may discourage competition from new entrants.

Economic costs of entry and exit play a critical role in long-run dynamics. Entry costs include acquisition of leases, interconnection, construction, equipment, financing and taxes, and in the case of wind, monitoring and study of wind availability. Plant retirement costs are also often significant and exhibit variation based on how the property is repurposed (Raimi, 2017). Some retired plants can be readily converted to industrial uses while others may require more extensive demolishing, scrapping of materials, and environmental cleanup and restoration.

1.2.3 Environmental Policy Influence on Short- and Long-Run Objectives

Each generation technology involves a trade-off between generation costs and capital costs. Wind plants have near-zero marginal generational costs, but require a large up-front investment. Dispatchable natural gas "peaker" plants designed to ramp up quickly to meet periods of high demand may involve relatively low capital costs but high operating costs which vary with the price of fuel. Environmental policy tends to affect these trade-offs and the relative attractiveness of investing in a particular type of power plant. Production tax credits (PTCs) for generation from renewable sources, made available through federal legislation since 1992, offer a monetary incentive to investors for each megawatt hour (MWh) of electricity produced over the first ten years of a plant's operation. The value of the PTC has varied due to occasional lapses in its renewal and phase-outs; in 2019 it was worth approximately $15 per MWh. (For comparison, the average price of electricity in the wholesale market in Texas was $38 in 2019.) These and

similar investment-side incentives allow investors to overcome the various entry costs associated with renewable generation.[3]

Though subsidies for renewables may encourage their deployment and therefore reduce carbon emissions, a price on carbon emissions is regarded as the first-best policy for internalizing harm from pollution generally (Pigou, 1924). Pricing carbon is also typically more cost effective than other policy instruments because it incentivizes emissions reductions in myriad ways. In electricity markets, a carbon price affects the marginal generation cost of units based on how much they pollute; coal plants which are relatively carbon-intensive will face greater cost increases than cleaner natural gas plants, and therefore be deployed less. These changes in deployment at the intensive margin immediately reduce carbon emissions in the short-run market. In the long-run, a carbon price implies that low-carbon technologies will be more profitable (due to relatively lower costs), thus encouraging investment in them, along with retirement of costly, carbon-intensive units. A subsidy for renewable electricity, in contrast, reduces emissions mostly through the extensive margin by encouraging deployment of low-carbon generation.[4]

1.3 Data

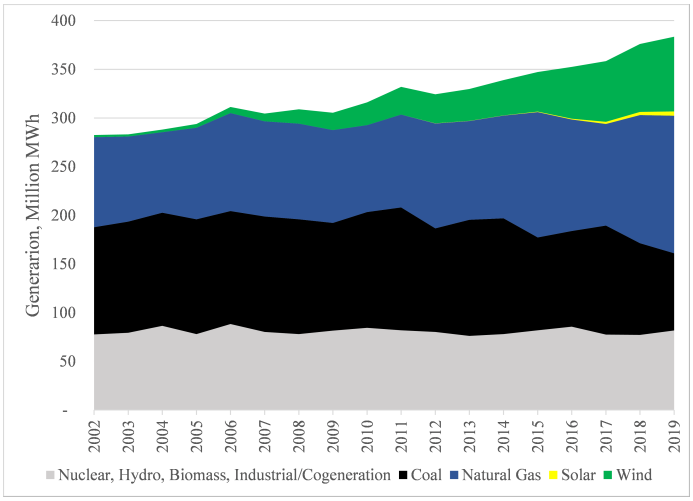
I construct a detailed and comprehensive data set of hourly electricity generation in ERCOT from 2002-2019 using data from the EPA's Continuous Emissions Monitoring System (CEMS),

[3]The Investment Tax Credit (ITC) is a parallel incentive designed to offset 30 percent of capital costs. For onshore wind most recipients have opted for the production tax credit. From 2009-2013 the federal government also made cash grants available to alleviate financial constraints as part of the American Reinvestment and Recovery Act. Aldy, Gerarden, and Sweeney, 2018 analyze the differential effects of production vs. investment subsidies. See Sherlock, 2020 for an overview of these subsidies.

[4]A third channel through which carbon prices lower emissions is through a demand response. Higher electricity prices reduce consumption which further reduces emissions. In contrast, a subsidy for renewable energy decreases electricity prices, thus encouraging consumption and therefore emissions. This channel is prominent in some dynamic analyses which demonstrate that carbon prices are relatively more cost effective than other instruments. See, e.g. Fell and Linn, 2013.

the Energy Information Administration (EIA), and ERCOT itself. The agent in my model is the firm, and firms often own multiple power plants of various technologies. Fossil fuel-based plants are in turn made up of multiple "generation units" such as combustion turbines in the case of natural gas. CEMS tracks hourly fuel input and electricity output at the generation unit level for fossil plants larger than 25 MW, excluding industrial and commercial cogeneration plants. EIA provides monthly net generation and fuel consumption data for all power plants larger than 1 MW. I supplement these data with ERCOT dispatch data which include all resource types and span 2011-2019. I impute hourly wind generation at the plant level prior to 2011 using historical wind profile data provided by ERCOT, which I map geographically to individual plants. Figure 1.5 illustrates the generation technology mix in ERCOT over time. Two major shifts are apparent in this figure: wind power has grown from comprising less than one percent of the generation mix in 2002 to greater than 20 percent in 2019, roughly matching the increase in demand over this period, and fossil generation has shifted away from coal and toward natural gas.

Figure 1.5: Generation by Technology 2002-2019

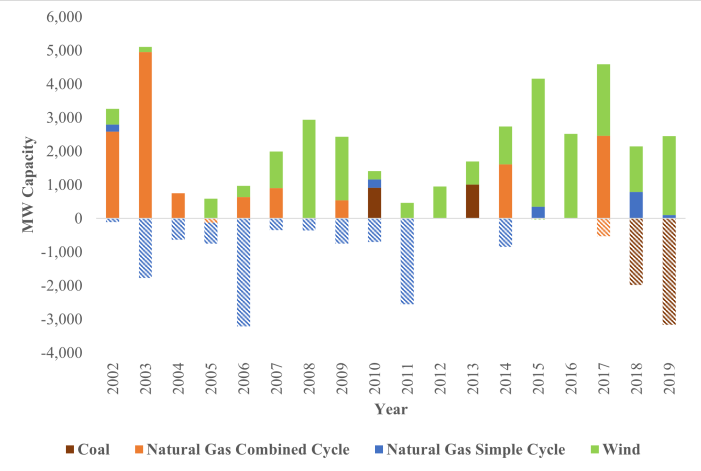


Generation has increased to meet steadily rising demand in Texas from 2002-2019. Generation from wind power (in green) has increased from less than one percent to more than 20 percent of the mix over this time. Source: ERCOT

I compute marginal generation costs using average monthly fuel prices from EIA and observed heat rates from CEMS and EIA. I use ERCOT's historical pricing data to assign real-time hourly settlement prices to power plants. From 2001-2010, ERCOT operated a "zonal" market made up of four geographic zones. Real-time energy market prices were settled in each zone according to security-constrained economic (least-cost) dispatch based on portfolio-level offer curves submitted by generation entities, constrained by transmission infrastructure. In December 2010, ERCOT transitioned to a nodal-level energy market which settled real-time prices in five-minute intervals across thousands of nodes. ERCOT's data records provide a mapping of settlement point prices to generation resources.

I use EIA data to track when a new plant is built or an existing plant is retired, and to determine plant attributes such as technology type. Figure 1.6 illustrates new additions (on the positive y-axis) and retirements (on the negative y-axis). Early investments in combined cycle generation marked the beginning of an ERCOT administrative overhaul in 2002 and 2003. The majority of plant retirements have been older and less efficient simple cycle, steam turbine-driven natural gas plants. Investment in wind generation has proliferated over this time period with larger increases in capacity during years in which federal tax subsidies were available. Investment in wind is also greater in the latter part of the sample following a large investment in transmission infrastructure connecting the windier portion of West Texas with the rest of the grid. This investment, known as the Competitive Renewable Energy Zones investment ("CREZ"), was initiated in 2009 and completed in 2013.

Figure 1.6: Entry and Exit by Technology Type, 2002-2019



Simple cycle natural gas fired plants (mostly steam turbine) made up the majority of retirements with significant coal generators coming offline in 2018 and 2019. A wave of investment in natural gas combined cycle generation (orange) marked the beginning of a fully restructured ERCOT in 2002 and 2003. Wind power investment (green) made up the majority of new capacity. Source: EIA

I use EIA records and filings from the Public Utility Commission of Texas ("PUCT") to track plant ownership over time. EIA tracks entities listed as plant owners, but these entities are often business structures set up for plant operations that are owned by a large firm or investor. The PUCT requires power generation asset owners to register and list parent companies. Because I am interested in the incentives of firms which may own multiple assets across the system, I reviewed PUCT records to trace ownership back to the parent company. This exercise reveals a market with approximately seven large entities accounting for 40 percent of plant ownership in ERCOT, with the remaining share composed of 65 smaller firms as of 2019. The seven largest entities are listed in Figure 1.1. I also find that firms are mostly specialized in either renewable or fossil-fuel based consumption, which informs how I define the action space in my long-run model—as a simplifying assumption I will limit firms' actions according to which technology

type they specialize in. Approximately 84 percent of market capacity is owned by independent power producers, and the remaining 16 percent by electric utilities.

Table 1.1: Capacity Shares 2019

Entity	**Specialization**	**Capacity Share**
Luminant	Fossil (100%)	14.6%
NRG	Fossil (98%)	8.6%
CPS Energy	Fossil (100%)	4.0%
Calpine	Fossil (100%)	3.8%
Exelon	Fossil (96%)	3.4%
Lower Colorado River Authority	Fossil (100%)	3.3%
NextEra	Renewable (100%)	3.0%
Remaining 65 Firms		59%

Seven firms made up 41 percent of generation capacity in 2019 with the remainder composed of 65 smaller firms. Most firms are specialized in either fossil fuel or renewable technologies. Source: EIA, Public Utility Commission of Texas. Shares exclude nuclear, hydro, biomass, and industrial and commercial cogeneration plants.

Appendix A.1 contains further details regarding data set construction.

1.4 Model

1.4.1 Characterizing the Long-Run Equilibrium

I characterize the long-run equilibrium as a dynamic game where firms make investment and retirement decisions under expectations of a non-stationary industry path. The industry path is composed of state variables which include exogenous market variables (demand, fuel costs, environmental policy) and the endogenous generation mix of the industry (installed capacities of various generation technologies). Describing the industry evolution as a non-stationary process is fitting due to the transitory nature of electricity market structure: as market conditions evolve,

firms adjust their expectations about the future and the industry evolves accordingly.

1.4.1.1 Model Setup

I first present a simplified version of the model for ease of exposition. I will then explain several institutional details which I accommodate during implementation. I begin with a fixed number of incumbent firms which are endowed with a set of power plants.[5] Firms operate under a finite horizon in response to an expected industry path which evolves in discrete time from the present period to a terminal period T. With T sufficiently large, firms' strategies converge to that of an infinite-horizon game. I set $T = 5$ and define a period as one year.[6] I define "entry" and "exit" as the firms' decision to introduce a new plant to the system or permanently retire an existing plant, respectively.[7] Exogenous state variables include demand, fuel costs, and environmental policy (in the baseline scenario, this involves a subsidy for wind power generation). I assume that both are deterministic and that demand is perfectly inelastic. The sequence of industry events is as follows:

1. Firms earn current-period profits in the short-run market through sales of electricity, and receive a random draw of logit shock parameters. These parameters are associated with the actions a firm might take: building a new plant or retiring an existing plant.

2. Firms take the decision which maximizes expected future profits and incur the fixed cost associated with the chosen action.

[5] I use the total number of unique firms observed throughout the data set; the number of incumbents is fixed throughout my analysis.

[6] Appendix A.2 discusses the sensitivity of the equilibrium path to the choice of T.

[7] On occasions where plants are mothballed, or put out of commission for several years, I treat the mothball period as exogenous. If the mothballed plant never returns to operation, I treat it as retired, starting the year that it first ceases operation.

3. The market advances to the next period: firms realize investments (thus adjusting their portfolio of plants), incur investment costs, and the next period's exogenous characteristics evolve.

Firm i maximizes its expected value in period $t < T$ according to the following Bellman equation:

$$V_{it}(x_{it}, s_t, \varepsilon_{it}) = \pi_i(x_{it}, s_t) + \max_{a_{it}}\{(-C(a_{it}) + \theta\varepsilon_{ait}) + \beta\mathbb{E}_\varepsilon V_{it+1}(x_{it+1}, s_{t+1}, \varepsilon_{it+1})\} \quad (1.1)$$

with the following primitives:

- $\pi_i(x_{it}, s_t)$ are profits to firm i in period t (one year's worth of hourly profits from electricity sales)
- $a_{it} \in \mathcal{A}$ is a firm's action
- Action space $\mathcal{A} \equiv \{$ build new plant, retire existing plant, keep portfolio the same $\}$
- x_{it} is a vector describing a firm's existing generation portfolio and plant attributes which determine profits from electricity sales. Specifically, x_{it} contains the generation capacity of each plant and the parameters which determine its supply offers in the hourly spot market. A firm's actions in period t will affect its portfolio in period $t + 1$.
- $s_t \equiv (x_{i,t}, x_{-i,t}, z_t)$ describes the industry state variables from the perspective of firm i. $x_{-i,t}$ includes capacity and plant attributes for firm i's rival plants operating in the market. z_t includes the hourly demand profile throughout year t, fuel prices, and environmental policy. $s \equiv s_0, ..., s_T$ describes the entire T-period industry path.
- $C(a)$ is a fixed cost function which I will assign a functional form upon estimation. $C(a)$

represents the economic costs associated with taking action a. If no action is taken in a given period then $C(a = \text{no action}) = 0$.

- $\varepsilon_{a,i,t}$ is a random i.i.d. type I extreme value shock, with dispersion parameter θ. If no action is taken in a given period, no cost shock is incurred.

- β is a discount factor. I use $\beta = 0.9$.

In Equation 1.1, uncertainty enters only through the action-specific logit shock. I estimate cost parameters, and use a short-run profit model to solve for profits in the stage game, described in Section 1.4.2. The value function in the terminal period is:

$$V_{i,T}(x_{iT}, s_T) \equiv \sum_{t'=0}^{\infty} \beta^{t'} \pi_i(x_{iT}, s_{iT}) \tag{1.2}$$

That is, in period $T-1$, final investments are made and the market ceases to evolve in period T.

Define a firm's policy function σ_i as a function that maps its portfolio and the industry path to an action in each period:

$$\sigma_i : x_{it}, s_t \mapsto a_{i,t} \text{ for } t = \{0, 1, ..., T\} \tag{1.3}$$

A non-stationary oblivious equilibrium is achieved when:

- Policy function σ_i^* is profit-maximizing for all firms i given each firm's initial endowment x_{i0} and the expected industry path s

- s is equal to the expected industry path given profit-maximizing policy functions σ^*

1.4.1.2 Simplifying Modeling Assumptions

Institutional features and computational constraints warrant several adjustments to the above simplified version of the model. First, I model three mutually exclusive specialized firm types: wind-specializing firms ("wind firms", indexed as type w), fossil-specializing incumbents with older power plants that are candidates for retirement (the "fossil retirers", indexed as type r), and fossil-specializing firms which build new power plants (the "fossil builders", indexed as type b). I define the action space of these firms as follows:

$$\mathcal{A}^w \equiv \{ \text{ build new wind plant, keep portfolio as-is } \}$$

$$\mathcal{A}^b \equiv \{ \text{ build new combined cycle plant, keep portfolio as-is} \}$$

$$\mathcal{A}^r \equiv \{ \text{retire existing plant, keep portfolio as-is } \}$$

This specialization assumption, while strong, is largely grounded in the empirical experience of the Texas market. Most firms in ERCOT are specialized with respect to investment in renewable vs. fossil fuel-based technology type over the course of the sample period. Solar power, hydropower, and biomass make up a small portion of the generation mix with little change in capacity over the sample period. Investment in and generation from nuclear power also changes rarely throughout. A limitation must also be imposed on the technology type that is undertaken when it comes to new fossil fuel investment given the small amount of observed investment in technologies other than combined cycle natural gas. I treat investment in peaker natural gas units and coal as exogenous; these include two late coal entrants (in 2010 and 2013) and seven natural gas peaker units which made up 1.6 percent of total market capacity in 2019.

Second, to maintain tractability I also constrain firm retirement decisions to adhere to an ordinal ranking that approximates likelihood of retirement. I assign each firm an internal ranking of its plants based first on the observed order of retirement (for those that were retired) and second on plant age. When estimating and solving the model, firms are allowed only to divest plants according to their divestment order. This assumption is strong as it precludes retirement of plants which may be of greater marginal benefit to the firm. Endogenizing the choice of plant to retire is an area of interest for continuation of this research.

Third, I assume that profits to wind firms reflect long-run contracts. Most wind investment over the sample period was undertaken with long-term power purchase agreements, but the parameters of these agreements are largely unobserved. I therefore compute expected profits for new wind plants based on the net present value of all expected future profit flows, including expected subsidies for wind power generation, from the perspective of the period in which the action was taken. This assumption implies that wind firms' investment decisions do not affect profits on the firms' existing portfolios. Fossil firms are assumed "merchant" and earn profits in the spot market in real time; unlike wind, plant profits for newly built plants are therefore allowed to fluctuate in periods beyond the investment decision.

1.4.2 Short-Run Profit Model

1.4.2.1 Overview

ERCOT takes a holistic approach to market settlement in the spot market in that all available generation resources are brought into the short-run market operation by way of their submitted supply schedules, and dispatched to equate generation with load at least cost subject to system

constraints. Behind the scenes of this sophisticated operation, the majority of electricity sales between buyers and sellers in Texas have already been agreed upon through bilateral contracting. Bilateral contract terms are typically confidential and not observed by the researcher. I therefore proceed by estimating firm profits as expected profits that would obtain from generation and sales in the real-time market. This assumption supposes that the expected value to a firm upon entering a contractual agreement is equal to expected future profits from spot market sales of electricity. As noted by the market monitor, "real-time energy prices set the expectations for prices in the day-ahead market and bilateral forward markets and are, therefore, the principal driver of prices in these markets where most transactions occur." (IMM, 2020)

I develop a short-run profit model which is able to reasonably approximate hourly market clearing prices and dispatch, and therefore allows me to compute firm profits in a manner that accounts for a rich set of market dynamics with low computational burden. To accomplish this I first parameterize functions which map hourly markups to unit-level supply based on the historical operations of each unit. Hourly markups are defined as the difference between real-time prices and generation costs. These offer curves are identified from rich hourly variation in load. I estimate offer curves for each unit and for each of five market states, which are categorized using the unit's position in the merit order, relative to demand. Merit position is defined using installed capacity and is therefore exogenous in the short-run. Aggregating the estimated offer curves gives a system-wide supply function. I then solve for the hourly price that equates supply and demand. This clearing price, however, does not account for localized transmission congestion or ERCOT's system-wide scarcity mechanism, both of which generate rents that have important implications for firm cost recovery and therefore investment and retirement decisions. I model expected rents from these factors using a simple reduced form approach which allows expected

rents to vary as a function of system-wide capacity availability.

I take this approach to modeling the short-run market in favor of more structural approaches employed elsewhere in the literature due to its computational expedience and attention to dynamics which produce profit estimates of reasonable precision. For example, incorporating structure to account for startup, shutdown, and ramping dynamics (e.g. Reguant, 2014, Cullen and Reynolds, 2017) or employing a detailed dispatch model (e.g. Johnston, Henriquez-Auba, Maluenda, and Fripp, 2019) is not suitable to my application because I require solving a short-run equilibrium over many hours and simulated random states.

A limitation of my approach is that it does not endogenize short-run market power exercise. Close monitoring of this type of strategic behavior, by both ERCOT and an independent market monitor, suggests it may not be a large concern for the purposes of my analysis. ERCOT mitigates generator supply bids if they imply markups that are in excess of a system-wide limit, and monitors markups closely based on resource-specific attributes. The independent market monitor conducts an annual review of an "output gap"—the difference between a unit's observed operating level and an operating level it deems competitive—ensuring further scrutiny of market power exercise.

These market power mitigation efforts limit the extent of strategic behavior in the short-run and complicate the application of equilibrium models which do not appropriately limit the extent of supply withholding to reflect offer curve mitigation. Moreover, research suggests that short-run market power exercise in ERCOT occurs during brief periods of localized congestion. Woerman, 2019 gives an estimate of a surplus transfer of $540 million from consumers to producers due to such market power exercise annually, representing approximately 4.5 percent producers' total revenue. While this is a non-trivial sum, such behavior will not be captured

without a detailed model accounting for transmission constraints, which would also be computationally prohibitive for my purposes. Though it does not endogenize short-run strategic behavior, my model is nonetheless able to capture average rents from transmission congestion and scarcity pricing.

I do not model firm profits from the sale of ancillary services, which make up a small fraction of revenues. Because I am not modeling reserves (which are offered as ancillary services), my modeling does not address questions related to reliability or blackouts. While the question of ensuring reliability is an interesting one, it is outside of the scope of this paper.

1.4.2.2 Approximating Offer Curves

Two key variables are useful for this exposition. First, define generation unit j's merit index $m_{j,h}$ in hour h as:

$$m_{j,h} \equiv \frac{L_h}{\sum_{n=1}^{N} \bar{q}_n \mathbf{1}(c_n < c_j)}$$

where L_h is net load, n indexes each unit in the merit order (inclusive of all N installed units on the system), $\bar{q}_n$ denotes operational capacity of unit n, and c_n denotes marginal generation cost of unit n. This index creates a measure of net load relative to a fossil unit's position in the merit order. Second, define "system-wide reserves" as:

$$r_h = G - Q_h \tag{1.4}$$

where G represents the total dispatchable (fossil fuel-based) generation capacity of the system (G varies by year), and Q_h is total hourly generation from dispatchable sources. System-wide

reserves are the total available dispatchable supply on the market minus the amount that is being used for power generation in hour h. Low reserve levels r_h are associated with higher prices.

I define short-run profits to firm i from electricity sales, summed across all hours H within period t (suppressing the t subscript for convenience) as:

$$\pi_i(x_i, s) = \sum_{h=1}^{H} p_h S_{ih}(p_h^c) - \sum_{h=1}^{H} \sum_{j=1}^{J_i} c_{ij} q_{ijh} \tag{1.5}$$

where p_h is the real-time price; $S_{ih}(p_h^c)$ is the firm's generation in hour h as determined by its offer curve and what I will refer to as the uncongested market-clearing price p_h^c; firms have constant marginal generation costs c_{ij} where j indexes generation units operated by the firm; and q_{ijh} is a unit's generation in hour h. I approximate individual firm supply curves $S_{ih}(p)$ using piecewise linear functions at the generation-unit level, the most granular level available, combined with a logit model which characterizes unit-level on-off decisions. I define $K = 5$ "states of the world" based on bins (five quantiles) of the merit index m_{jh} for each generation unit. Each bin represents different market conditions in which the unit may or may not be operating. For example, a high merit index value reflects an hour in which demand is high relative to the unit's merit position; in such an hour the unit is likely to be infra-marginal. In other periods, the unit may be close to the margin, and in other periods may be supra-marginal. I approximate supply functions for each generation unit in each state of the world based on observed behavior during periods in which the unit is operating (I refer to these as conditional supply functions). In each state of the world k, I use a k-means algorithm to identify three clusters of each unit's generation status relative to its capacity (and therefore bound between 0 and 1), and add 0 and 1 to the set of possible generation

levels: $f_{jkl} \in \mathcal{F}_{jk} \equiv \{0, \hat{f}_{jkl}, \hat{f}_{jkl}, \hat{f}_{jkl}, 1\}$. The piecewise-linear function can be written as:

$$f_{jk}(p - c_j) \equiv \sum_{l=0}^{4} d_{jkl}(p - c_j) f_{jkl} \tag{1.6}$$

where $f_{jk0} = 0, ..., f_{jk4} = 1$, and $d_{jkl}(p - c_j) = 1$ if markup $p - c_j$ is in segment l, 0 else. Segments l are defined using the median observed markup associated with each k-means cluster of the set $\mathcal{F}_{jk}$. Unit-level quantity generated is then:

$$q_{ijh} = \sum_{k=1}^{5} g_{jkh} f_{jk}(p - c_j) \bar{q}_{ij} \tag{1.7}$$

where $g_{jkh} = 1$ if unit j is in state k (if its merit index is in the k'th quintile in hour h), 0 else; and $\bar{q}_{ij}$ is the operational capacity of unit j belonging to firm i.

1.4.2.3 Unit On-Off Decisions

I nest conditional supply functions in a logit model which captures the probability that a fossil generation unit will be operating in a given hour. This decision (denoted $y_{jh} = 1$ if the unit is switched on in hour h, 0 if not) is implied by the entity's submitted supply offer which determines its dispatch. I model this decision as follows:

$$\hat{y}_{jh} = Pr(y_{jh} = 1 | X_{jh}) = \frac{1}{1 + \exp(-X'_{jh}\beta_j)} \tag{1.8}$$

where X_{jh} is the generation unit's merit order in hour h and a constant term. I estimate individual logit parameters and conditional supply function parameters for each generation unit and for each

year. Firm-level supply functions are the product of the predicted operation probability and unit-level quantity:

$$S_{ih}(p) = \sum_{j=1}^{J_i} \hat{y}_{jh} q_{ijh} \tag{1.9}$$

1.4.2.4 Scarcity and Congestion Rents

The Texas market currently settles prices across thousands of nodes—points on the system from which power is drawn from generators. Clearing prices therefore vary from temporal demand patterns and also due to localized transmission congestion and episodes of system-wide scarcity. Backing out offer curves as described above allows me to form a market-wide offer curve, but the resulting clearing price (the price that equates aggregate supply with demand) will not reflect localized congestion or system scarcity. Such prices are critical to capture in the short-run model because they affect investment decisions. Congestion rents encourage investment in places where supply is constrained, and the scarcity price is designed to ensure sufficient total capacity in meeting market-wide demand.

To account for these features of the market, similar in spirit to Coulon, Powell, and Sircar, 2013 I define the expected hourly market price p_h as the probabilistic combination of the "uncongested" market-clearing price $p_h^c$and the "spike price" p_h^s:

$$p_h = (1 - \delta_h) p_h^c + \delta_h p_h^s \tag{1.10}$$

where δ_h represents the probability of a spike price. The uncongested market-clearing price is the price that equates aggregate supply to net load: $\sum_i S_{ih}(p_h^c) = L_h$. I hold generation from renewable and other non-fossil sources, imports, and industrial and commercial sources as

exogenous and define net load L_h as the difference between total load and generation from these exogenous sources.

System-wide reserves r_h as defined in Equation 1.4 are negatively correlated with the probability and magnitude of the spike price. I estimate the probability of a spike regime δ_h using a logit model with system-wide reserves as the explanatory variable. I define a spike regime as any hour in which the real-time price is in the top percentile. I use a simple OLS regression model to estimate the log of the spike price as a function of system-wide reserves.[8]

1.5 Solution Method and Structural Estimation

1.5.1 Solution Method

To solve the model, I use a solution method that is akin to policy iteration and draws from methods employed by Sweeting, 2013. The algorithm begins with an initial guess at the expected industry path s. Given the industry path, I simultaneously solve a series of single-agent finite horizon problems for each firm using backward induction. This results in a set of conditional choice probabilities for each firm in each period. I use the resulting choice probabilities to compute a new probabilistic industry path s. I repeat these steps until the industry path converges to a specified tolerance.

The industry path s is made up of exogenous and endogenous elements. The exogenous elements include firm expectations about future load, fuel prices, carbon prices (for counterfactual analysis), and wind generation subsidies in periods $t = 1, 2, ...T$. The endogenous elements

[8]Texas introduced a scarcity price mechanism in 2014. I therefore estimate two sets of OLS parameters for the spike price, one pre-2014 and one for 2014 forward, to capture the difference in expected spike prices as a function of r_h.

include the firm's present and future investment path, and the industry path reflecting the present and future composition of the rest of the market, namely, the collective investments and retirements made by the firm's rivals in period $t = 1, 2, ...T$. Although each player is playing a single-agent game, each player's actions will affect the industry path. The model converges when all firms' actions, collectively, represent the expected industry path upon which those actions are based.

I solve an equilibrium for each observed year in the data using contemporaneous forecasts of exogenous state variables associated with that year. For example, when solving the equilibrium for the year 2003, I use load forecasts generated by ERCOT in 2003 for T years into the future. I assume a non-stationary industry path that converges to a stationary equilibrium in $T = 5$ periods: in periods beyond T, I assume the hourly demand pattern remains fixed at its period-T level, and no further investment or retirement actions are taken. Given the Bellman equations expressed in (1) and (2), firm i's *ex ante* value function in $t < T$ is:

$$\begin{aligned} EV_{it}(x, s_t) &= \mathbb{E}_{\varepsilon}[\max_{a}\{\pi_i(x, s_t) + \theta\varepsilon_a - C(a) + \beta V_{t+1}(x(a), s_{t+1}(a)\}] \\ &= \pi_i(x, s_t) + \sum_{a} P(a|x, s_t)[-C(a) + e(P(a|x, s_t)) + \beta EV_{i,t+1}(x(a), s_{t+1}(a))] \end{aligned}$$

$e(P(a))$ is the conditional logit shock $e(P(a)) = \theta(\xi - ln(P(a)))$ where ξ is Euler's constant, $C(a)$ is the fixed cost function (defined with specific functional forms for different firm types below) associated with action a, θ is the logit dispersion parameter (which I estimate separately across firm types), and $P(a|x, s_t)$ are conditional choice probabilities. Because demand is deterministic, the only uncertainty arises from the logit shock ε. Each firm's profit function is individualized according to the attributes of its portfolio, and therefore indexed with i.

1.5.1.1 Plant-Level Profits

Solving for profits in each iteration using the short-run model would be prohibitively burdensome computationally. Instead, I employ a linear sieves approach to approximate profits at the plant level.

$$\pi_{io}(x, s) \approx \phi(x, s)\lambda \tag{1.11}$$

where $\phi(x, s)$ is a vector of functions of state variables and λ is a set of parameters, and o indexes the plants owned by firm i.[9] To ensure external validity of my profit estimates I generate a set of widely varied random industry states and solve for profits using the short run model described in Section 1.4.2. I use a total of 36,000 random states, each representing the market composition—the generation assets owned by each firm and their attributes—and hourly demand in one period. Because this process is only required to be performed once, I can generate a large and varied synthetic data set for estimating λ. I ensure that the state spaces reflected in this data set are feasible in the short-run model (the market clears given parameterized supply functions from the generation mix), and that they include, for some portion of the sample, values of variables that will be investigated in my counterfactual analysis (a non-zero carbon price for some random draws). Taking these measures helps to ensure precision of the linear approximation for state

[9]For wind plants, I use a quadratic function of the following variables interacted with inverse peak load: total fossil capacity, total wind capacity, and fossil capacity binned using four quantiles of marginal generation costs within three technology types (coal, combined cycle, and simple cycle). For fossil plants, I use a quadratic function of the following variables interacted with inverse peak load and total fossil capacity: total wind capacity, fossil capacity binned using four quantiles of marginal generation costs within three technology types (coal, combined cycle, and simple cycle), marginal generation cost, and the system-wide price cap. To control for changes in fuel costs I scale all variables by the natural gas fuel price. I estimate profits per MW capacity. See Chen, 2007 for an overview of the sieves method.

spaces that may be realized in counterfactual modeling.

1.5.1.2 Backwards Induction Step

For a given industry path $s = \{s_0, s_1, ...s_T\}$, value functions for a firm are approximated in the terminal period as:

$$EV_{iT}(x_{i,T}, s_T) = \sum_{t'=0}^{\infty} \beta^{t'} \sum_{o} \phi_o(x_{o,T}(a_{T-1}), s_T(a_{T-1}))\lambda$$

and in periods $t < T$:

$$\begin{aligned} EV_{it}(x_{i,t}, s_t) = &\sum_{o} \phi_o(x_{o,t}, s_t)\lambda \\ &+ \sum_{a} P_i(a|x_{i,t}, s_t)\Big(- C(a) + e\big(P_i(a|x_{i,t}, s_t)\big) + \beta EV_{it}\big(x_{i,t+1}(a), s_{t+1}(a)\big)\Big) \end{aligned}$$

where o indexes plants within a firm's portfolio, and conditional choice probabilities $P_i(a|x_{i,t}, s_t)$ are computed as:

$$P_i(a|x_{i,t}, s_t) = \frac{\exp\Big[\frac{1}{\theta}\Big(- C(a) + e(P_i(a|x_{i,t}, s_t)) + \beta EV_{i,t+1}(x_{i,t+1}(a), s_{t+1}(a))\Big)\Big]}{\sum_{a'} \exp\Big[\frac{1}{\theta}\Big(- C(a') + e(P(a'|x_{i,t}, s_t)) + \beta EV_{i,t+1}(x_{i,t+1}(a'), s_{t+1}(a')\Big)\Big]}$$

1.5.2 Structural Parameter Estimation

I define fixed cost functions $C(a)$ as follows:

$$C(a) = \begin{cases} \gamma_{0,w} + \gamma_{1,w}\kappa_w & \text{for new wind plants} \\ \gamma_{1,b}\kappa_b & \text{for building new combined cycle natural gas plants} \\ \gamma_{0,r} & \text{for divesting a fossil fuel plant} \end{cases} \tag{1.12}$$

where κ_l is capacity of type $l \in \{w, b, r\}$. I use industry estimates to inform per-MW capital costs $\gamma_{1,w}$ and $\gamma_{1,b}$.[10] I estimate lump-sum costs $\gamma_{0,w,}, \gamma_{0,r}$ and, because types are mutually exclusive, three separate dispersion parameters $\theta_w, \theta_b, \theta_r$. To estimate these parameters I nest the oblivious equilibrium solver described above in a policy iteration procedure which follows a similar intuition. The steps to this procedure are as follows:

1. With parameters γ^k, θ^k (starting with an initial guess at $k = 0$), solve for value functions and s that satisfy the oblivious equilibrium by running the equilibrium solver algorithm described above.

2. Solve for parameters that rationalize observed behavior. For wind firms, I achieve this using the following criterion function:

$$\gamma'_w, \theta'_w = \underset{\gamma,\theta}{\operatorname{argmin}} \sum_u \sum_i (y_{iu} - P_i(y_{iu}|x_{iu}, s_u, \gamma^k, \theta^k))^2$$

[10] In particular, I use annual estimates of average wind turbine transaction prices per MW capacity from Lawrence Berkeley National Laboratory, and estimates of the capital cost of building a new combined cycle natural gas plant from EIA's Annual Energy Outlook 2020, which are fixed over time in my model. While heat rates have improved over time for natural gas plants (which I exogenously account for in my model), per-MW capital costs for combined cycle plants have changed little.

where $y_i = 1$ if firm i made the active choice in the action space (in this case, to build a new wind plant) and $P_i(y_{iu}|x_{iu}, s_u, \gamma^k, \theta^k)$ are the corresponding choice probabilities in observed periods u as produced by the solver. Fossil fuel investments are relatively sparse (13 new combined cycle plants were built over the sample period). I use the following criterion function for fossil parameters:

$$\gamma_l', \theta_l' = \underset{\gamma,\theta}{\operatorname{argmin}}((\frac{1}{N}\sum_u \sum_i y_{iu}) - (\frac{1}{N}\sum_u \sum_i P_i(y_{iu}|x_{iu}, s_u, \gamma^k, \theta^k)))^2$$

3. Set $(\gamma^{k+1}0, \theta^{k+1}) = \psi(\gamma', \theta') + (1 - \psi)(\gamma^k, \theta^k)$ and repeat steps 1-2 until convergence is achieved at a specified tolerance: $|(\gamma_{l0}^k, \theta_l^k) - (\gamma^{k+1}, \theta^{k+1})| < \epsilon$ for all firm types l.

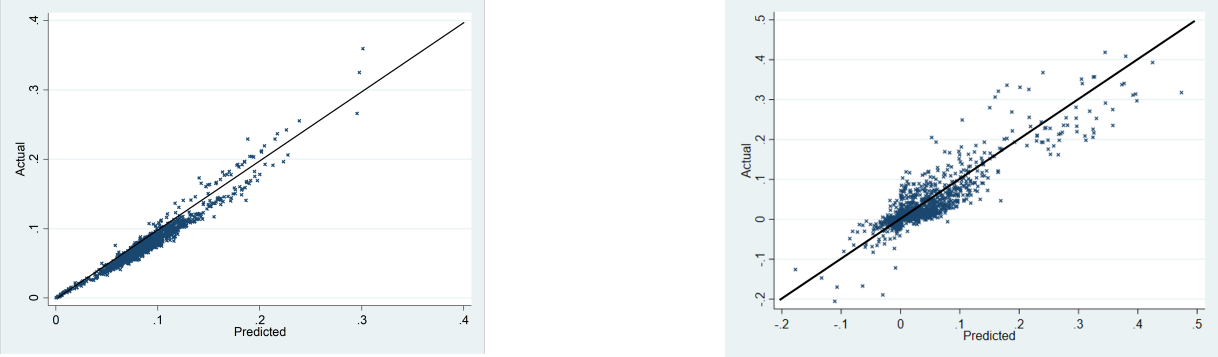
I set $\psi = 0.3$ and $\epsilon = \$10,000$. Identification is achieved from variation in firm expectations about future profit flows. For fossil firms, cross-sectional variation arises from differences in plant attributes (manifested through supply function parameters in the short-run model), and firm size—large firms will be relatively more sensitive to the price effects of adding or removing a plant. Temporal variation arises from changes in fuel prices and the evolving state space (generation mix among competitors, changes in expectations about future load). With respect to wind, cross-sectional variation arises from differences in wind favorability, and differences in the efficiency of turbines used by firms. Temporal variation comes from technical efficiency improvements and changes in the availability of the production tax credit.

1.6 Estimation Results

1.6.1 Short-Run Model

The primary objective of the short-run model is to achieve a sufficiently precise approximation of expected profits from the sale of electricity in the spot market under a wide range of feasible state spaces that could obtain in counterfactual simulations of the long-run industry transition. Annual plant-level profits are the most relevant because these are the building block of firm profits and the basis for firm decisions, which are taken on a yearly basis. Precision of profit estimates are summarized in terms of the prediction error in Figure 1.7 and Table 1.2. I define actual profits for a given firm as the observed real-time spot market price minus fuel costs associated with that firm's observed generation, summed over all hours of the year. I define predicted profits analogously, where prices and fossil dispatch are predicted using the short-run model, and wind quantities are taken as exogenous. Because the model does not account for geographic variation in transmission constraints, which lead to localized price spikes, some dispersion from actual profits is expected. This dispersion is more pronounced in fossil plants, where dispatchability enables plant operators to capture more localized price variation.

Figure 1.7: Actual vs. Predicted Annual Plant Profits per MW Capacity



(a) Wind (b) Fossil

Actual profits per nameplate capacity are shown on the y-axis and defined as observed revenues (using the real-time price) net of generation costs (which are zero in the case of wind). Predicted profits per nameplate capacity are shown on the x-axis and are produced using the short-run market model. The black lines indicate a perfectly predicted profit.

Table 1.2: Actual vs. Predicted Annual Plant Profit per MW Capacity, $ MM

Variable	Obs	Mean	Std. dev.	Min	Max
Fossil					
Prediction Error	1,423	-0.003	0.031	-0.159	0.158
Absolute Error	1,423	0.021	0.023	0.000	0.159
Wind					
Prediction Error	1,092	-0.011	0.009	-0.046	0.059
Absolute Error	1,092	0.012	0.007	0.000	0.059

Prediction errors are defined as the difference between actual and predicted profits per MW capacity at the plant level. Absolute errors are equal to the absolute value of the predicted error.

1.6.2 Entry and Exit Cost Parameter Estimates

Estimates of the parameterized fixed cost functions from Equation 1.12 are shown in Table 1.3. For firms choosing an action, the estimated parameter γ_0 is accompanied by a logit shock that is distributed according to the estimated scale parameter θ. Firms taking an action (building

or retiring a plant) can be thought of as drawing a random logit shock from the given distribution, comparing it with the value of future discounted profits that would obtain from building a new plant net of capital costs, and deciding whether to follow through with the action. In Table 1.3, the rightmost column gives the total action costs associated with the average-sized plant and the unconditional mean of the logit shock. This amount is comparable to the average total economic costs associated with each action.

For wind, I estimate separate costs associated with the periods before, during, and after the CREZ transmission infrastructure project which was undertaken from 2009-2013. Profits in my model take into account improvements in wind turbine capacity factors and changes in capital costs (which I hold exogenous); cost parameter estimates presented here control for these variables. My results suggest declining lump sum economic costs γ_0 and scale θ, which translate to declining unconditional average costs over time. One possible explanation for this result may be the limited availability of resources for developing wind plants in earlier years compared with later years. The CREZ project, for example, made vast areas of the Texas available for development and interconnection, thus lowering economic costs associated with building a new wind plant. My model produces results for fossil retirements costs which are larger than other estimates in the literature. Davis, Holladay, and Sims, 2021 estimate average retirement costs for coal plants at $71.6 million for active plants, and $37.4 million for retired plants, significantly less than the $262.5 million unconditional average lump sum retirement cost estimate presented here. Further examination of these costs is a natural area for further attention.

Table 1.3: Estimated Entry and Exit Parameter and Cost Estimates, $MM

	Cost γ_0 (estimated)	**Ave. Capital Costs (given)**	**Scale θ (estimated)**	**Unconditional Average Costs (Plant)**
Wind Pre-CREZ (2002-2013), $MM	475.8	188.2	179.7	560.2
Wind CREZ (2002-2013), $MM	192.3	235.5	92.6	374.3
Wind Post-CREZ (2014-2019), $MM	146.1	194.9	59.2	306.9
Fossil Entry (Combined Cycle) $MM	-	859.6	85.2	810.4
Fossil Retirement, $MM	314.9	-	90.9	262.5

Lump sum cost parameters γ_0 and scale parameters θ are estimated. Capital costs reflect per-MW capacity costs multiplied by the average plant size $\bar{\kappa}$. Lump sum costs for combined cycle entry and per-MW costs of fossil retirement are assumed zero. Unconditional average costs = $\gamma_0 + \gamma_1\bar{\kappa} - \theta\xi$ where ξ is Euler's constant. Capital cost data source: EIA, Lawrence Berkeley National Laboratory

1.6.3 Long-Run Model Performance

Following estimation of structural parameters, I perform simulations of the model, holding exogenous market conditions at their observed values to assess model performance compared with the actual, observed industry path. An individual simulation is performed as follows:

1. From an initial period, solve for the market equilibrium based on firm expectations about the future industry path

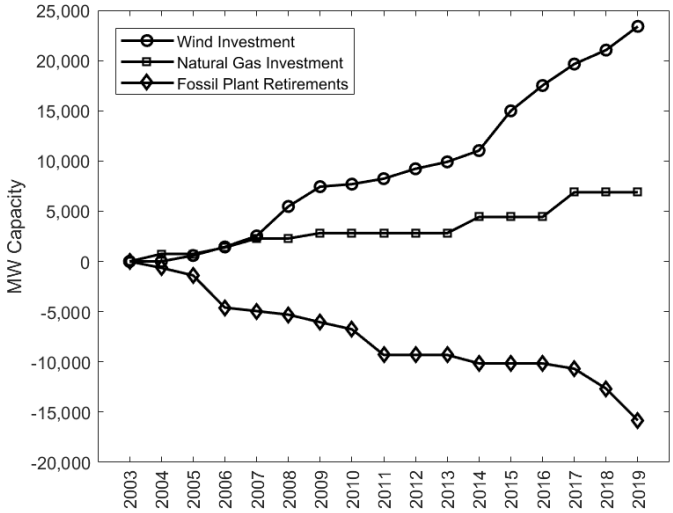
2. Select a set of random draws of logit shocks and advance the model by choosing discrete firm actions which maximize future expected profits given the randomly drawn shocks

3. Repeat steps 1 and 2 until the simulation is complete for all periods

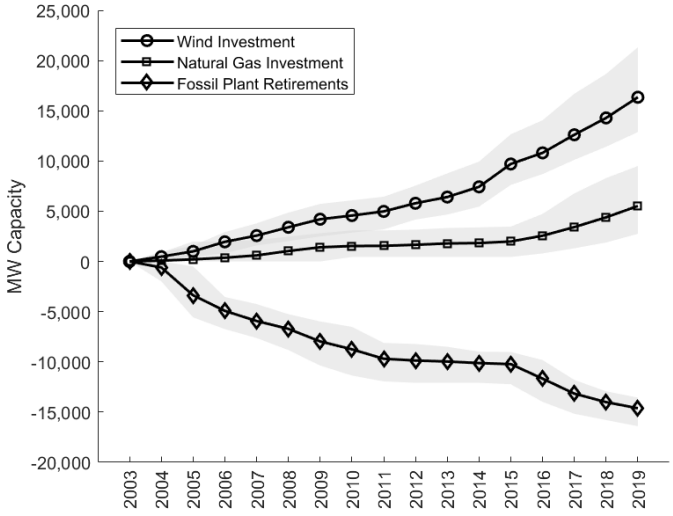
I perform $N = 50$ simulations and begin at 2003, advancing the model through 2019. Figure 1.8 illustrates results of this exercise for wind investment, fossil investment, and fossil retirements. The model follows the observed paths of investment and retirements fairly well. I will refer the mean industry paths shown in Figure 1.8b as the baseline scenario, making comparisons against

this scenario in the counterfactual analysis.[11]

Figure 1.8: Observed vs. Predicted Industry Evolution



(a) Observed Industry Paths



(b) Predicted Industry Paths

Panel (a) shows observed investment (cumulative added capacity) and cumulative retirement capacities over time. Panel (b) shows investment paths as predicted by the structural model. Grey shading indicates 95 percent confidence interval for $N = 50$ random draws.

[11] Although my data set begins in 2002 (the year in which an overhaul of market rules was introduced in ERCOT), I use 2003 as a starting point due to an outsize amount of new natural gas capacity that was introduced that year and which totaled 5,000 MW, or 34 percent of the total invested over the entire sample period.

1.7 Counterfactual Simulations

I present results from eight sets of simulations. I first compare counterfactual scenarios with a "baseline" case in which firm concentration and the environmental policies of interest are taken as observed in the actual world (wind production subsidies in the form of the federal PTC, and no carbon price). Holding market concentration fixed I then simulate an environmental policy counterfactual scenario which removes the wind subsides, another which replaces it with a carbon emissions tax of $20, and one which includes both wind production subsidies and a $20 carbon tax.

I repeat the same four scenarios (wind production subsidies only, a carbon tax only, both wind production subsidies and a carbon tax, and no policy) while adjusting firm concentration, simulating a "competitive" market in which price effects are mostly removed from the firm's objective function. I achieve this set of scenarios through a simple adjustment of firms' conditional choice probability computation. For firms that build plants, I remove profits obtained from the firm's existing portfolio from the objective function. Doing so results in conditional choice probabilities which are based solely on profits from the new plant in question, and not on effects of that investment on future profits accruing to its other plants. For the retirement decision, firms choose between incurring an exit cost plus zero future profits (retiring the plant), or future expected profits from continuing to operate the plant. Table 1.4 lists these simulations and their features. I will refer to the scenarios in which market concentration is taken as observed as the "concentrated" scenarios, and those in which price effects are absent as the "competitive" scenarios.

Table 1.4: Counterfactual Simulations

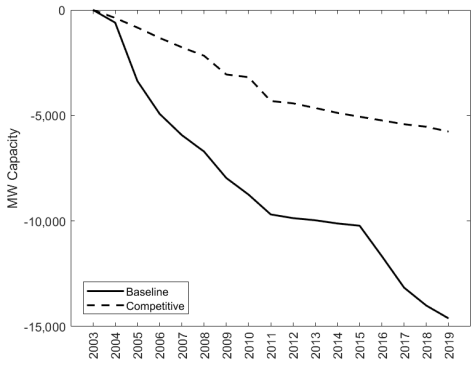
Counterfactual	Competitive/ Concentrated	Wind Production Subsidies	Carbon Price
1 (No Policy)	Concentrated	No	$0
2 (Baseline)	Concentrated	Yes	$0
3	Concentrated	No	$20
4	Concentrated	Yes	$20
5	Competitive	No	$0
6	Competitive	Yes	$0
7	Competitive	No	$20
8	Competitive	Yes	$20

1.7.1 Market Power Effects

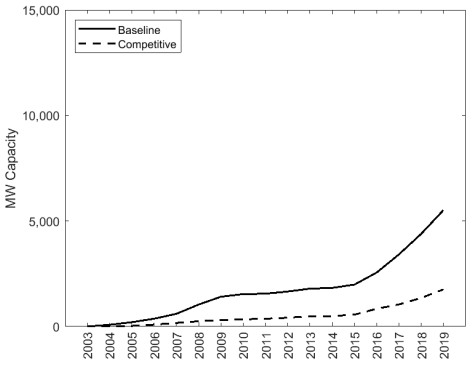
Figure 1.9 compares the industry path that obtains under the baseline (concentrated) scenario with that of the competitive scenario (6) which holds environmental policy fixed. A central result of this analysis is that fossil fuel plant retirements are significantly greater in the presence of strategic investment incentives relative to this competitive scenario, as shown in panel 1.9a. This result suggests that long-run market power exercise incentives lead to an additional 8,850 MW in retirements—amounting to nearly seven percent of the entire Texas system in 2019, or more than enough to power the entire state of Connecticut.

Retirements have an economically significant effect on electricity prices, which in turn lead to lower profits for new plants. This effect gives rise to second-order effects: fewer retirements in the competitive scenario translate to lower investment in new natural gas capacity (Figure 1.9b) and slightly lower investment in wind (Figure 1.9b).

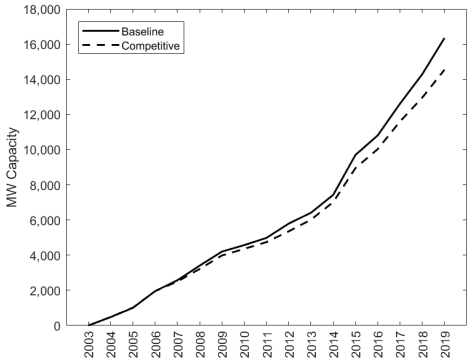
Figure 1.9: “Baseline” vs. “Competitive” Industry Path



(a) Fossil Retirements



(b) Fossil (Natural Gas) Investment



(c) Wind Investment

1.7.2 Environmental Policy Counterfactuals

Figure 1.10 illustrates industry paths involving both competitive and environmental policy counterfactuals. The solid lines indicate the baseline investment scenario which involves the federal production tax credits as they were made available over the sample period. The industry path absent those subsidies is indicated by the dashed line. Finally, the effect of a $20 carbon price in lieu of subsidies is represented by the dotted line. This price is incorporated into the model by adding to plant generation costs according to plant-specific emissions rates. This price level is modest relative to the current social cost of carbon set in the US ($51) as well as recent legislative proposals. The panels on the left-hand side reflect scenarios involving market concentration

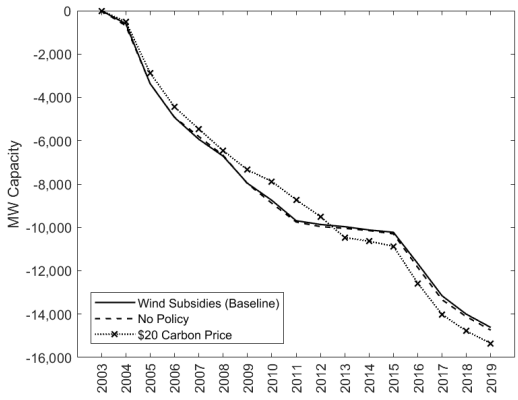
(holding plant ownership fixed), and the panels on the right-hand side reflect the competitiveness assumption.

The top two panels illustrate the effects of environmental policies on the fossil plant retirement path in the concentrated scenario (Panel 1.10a) and competitive scenario (Panel 1.10b). Both policies have little effect on the pace of retirements. This may be attributable to the price on carbon being relatively low; 2021 find that a higher carbon price of $51 advanced retirement dates two years into the future. My results also suggest that having significantly less wind on the system (in the absence of subsidies), associated with higher average prices, would not significantly affect the pace of fossil plant retirements.

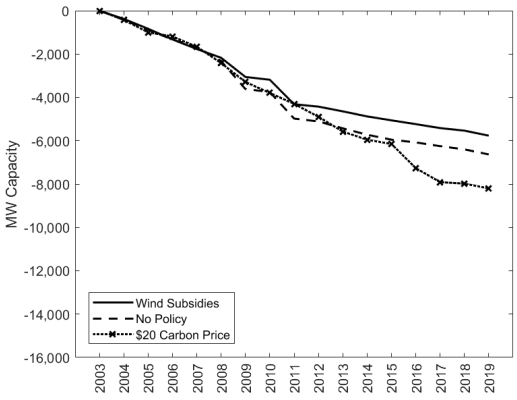
Replacing subsidies for wind with a carbon tax has a larger effect on new natural gas investment, as shown in Panels 1.10c and 1.10d, relative to both the baseline scenario (which includes wind subsidies) and the scenario that does not include wind subsidies. In these figures the second-order effects of fewer retirements are again apparent. In Panel 1.10c, high levels of retirements combined with a carbon price result in over 7,000 MW of combined cycle natural gas capacity by 2019. In Panel 1.10c this effect is diminished to 5,400 MW due to the large amount of existing fossil fuel based capacity staying online.

Panels 1.10e and 1.10f illustrate wind investment paths. The large difference in investment in Panel 1.10e suggests that, from 2003-2019, approximately 6,900 MW of wind capacity additions are attributable to the federal production tax credits for renewable energy that were made available over that time, an increase of 73 percent. Under a $20 carbon price wind investment to a path in excess of that effectuated by wind production subsidies, an increase of 104 percent or 9,900 MW. Wind investment is slightly lower in the competitive scenario, likely due to the lower market prices associated with fewer fossil fuel plant retirements.

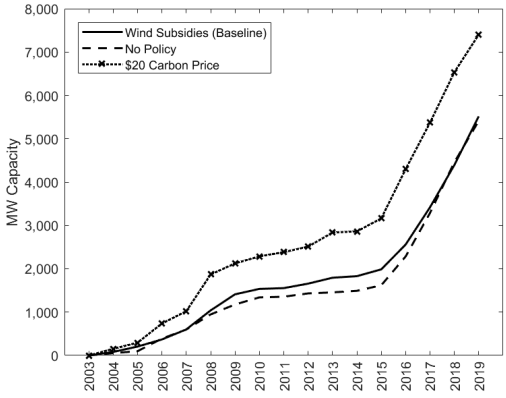
Figure 1.10: Environmental Policy, Market Power Effects, and Industry Path



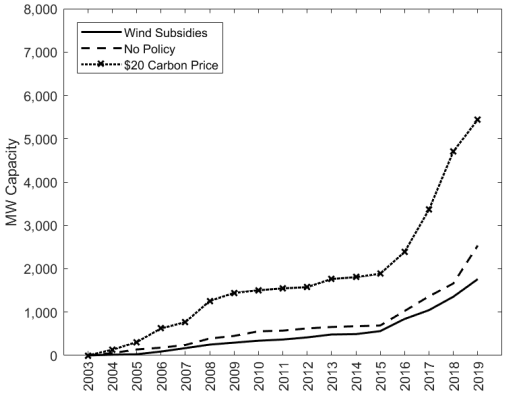
(a) Fossil Retirements, Concentrated



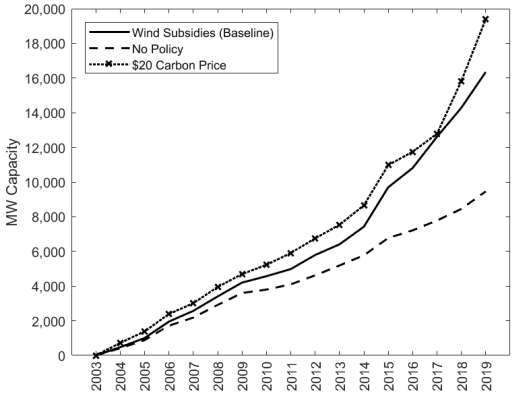
(b) Fossil Retirements, Competitive



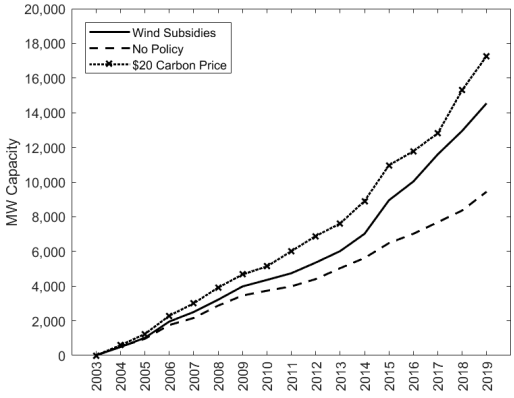
(c) Fossil (Natural Gas) Investment, Concentrated



(d) Fossil (Natural Gas) Investment, Competitive



(e) Wind Investment, Concentrated



(f) Wind Investment, Competitive

1.7.3 Emission Reductions and Welfare Analysis

The confluence of economic factors reflected above requires a holistic approach to analyzing the environmental and welfare effects of the industry over time. For example, market concentration accelerates the retirement of fossil units, which has the effect of elevating market prices. These elevated market prices in turn encourage investment in other technologies that would not have arisen otherwise—as shown in Figure 1.9, wind and natural gas investment are significantly higher in the presence of firm concentration. Moreover, significant changes in emissions arise from a carbon price due to a re-arranged dispatch order of generators, regardless of retirements and new investment. To assess the net effects of these various factors I compare key welfare metrics across the counterfactual simulations. Results are shown in Table 1.5. In this analysis I assume the carbon price is implemented as a tax and revenues are transferred to consumers in lump sum. Results show changes relative to the "no policy" counterfactual (1), which holds concentration as given but involves neither environmental policy; this comparison allows assessment of the cost effectiveness of the environmental policies both in the presence and absence of long-run market power exercise. Cost effectiveness is computed as the change in total (discounted) surplus associated with each counterfactual relative to the no-policy counterfactual, divided by emissions reductions. These results suggest that a carbon price of $20 in the concentrated scenario (3) would effectuate more than six times the emissions reductions as the federal production tax credits, and at less than half the cost in total welfare. The competitive scenario with a carbon tax yields the greatest cost effectiveness at $22.46 per ton of emissions reduced.

Emission reductions from a carbon price involve reductions at the intensive margin, from a re-arranged dispatch order, as well as reductions at the extensive margin, which develop over

Table 1.5: Counterfactual Welfare Analysis Summary

	Competitive/ Concentrated	Carbon Tax ($)	PTC	Δ Consumer Surplus ($MM)	Δ Producer Surplus ($MM)	Δ Total Surplus ($MM)	Δ Emissions (MM tons)	Cost Effectiveness ($ / ton CO_2)
(1)	Concentrated	$0	No	$0	$0	$0	0	-
(2)	Concentrated	$0	Yes	$1,578	-$5,858	-$4,281	-63	$67.99
(3)	Concentrated	$20	No	$7,479	-$18,045	-$10,566	-396	$26.66
(4)	Concentrated	$20	Yes	$7,047	-$23,782	-$16,735	-493	$33.94
(5)	Competitive	$0	No	$10,376	-$7,887	$2,489	32	-
(6)	Competitive	$0	Yes	$10,924	-$11,201	-$277	-10	$27.41
(7)	Competitive	$20	No	$14,683	-$22,979	-$8,295	-369	$22.46
(8)	Competitive	$20	Yes	$14,637	-$27,095	-$12,458	-436	$28.60

Changes in surplus and emissions are shown relative to scenario (1). Cost effectiveness is equal to total decrease in total surplus divided by the reduction in emissions. Consumer welfare changes reflect carbon tax revenue transferred as a lump sum to consumers. The PTC is treated as a transfer from consumers to producers. Results do not reflect the social value of reduced emissions.

time as capital is replaced. All of the emissions reductions achieved by a subsidy for renewable energy are attributable to changes at the extensive margin. The welfare analysis above reflects the total combined extensive and intensive emissions reductions. It is useful to compare how extensive margin emissions reductions arise under either scenario to better understand how the composition of the generation mix is changing over time. I make this comparison by decomposing emissions reductions associated with a carbon price into their intensive and extensive margin subparts. Define $E(x(\cdot), \tau)$ as the emissions that obtain under industry composition x (defined as the installed generation mix of all firms) if carbon price τ were to be applied, and holding demand fixed. $x(\text{policy})$ indicates the market composition that obtains if an environmental policy is put in place, and $x(\text{no policy})$ indicates the composition that obtains if no policy is put in place. The

decomposition is as follows:

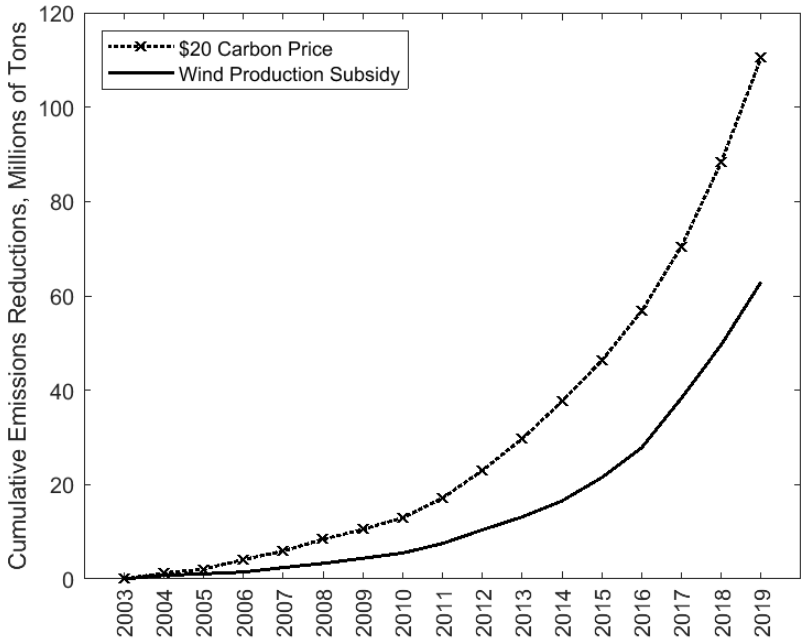
$$\begin{aligned}\Delta E &= E(x(\text{no policy}), \tau = 0) - E(x(\text{policy}), \tau = 20) \\ &= \underbrace{E(x(\text{no policy}), \tau = 0) - E(x(\text{policy}), \tau = 0)}_{\text{Extensive Emissions Reductions}} + \underbrace{E(x(\text{policy}), \tau = 0) - E(x(\text{policy}), \tau = 20)}_{\text{Intensive Emissions Reductions}} \\ &\equiv \Delta E_{\text{ext}} + \Delta E_{\text{int}}\end{aligned}$$

ΔE_{ext} can be thought of as the amount of emissions reductions that obtain only from changes in the long-run capital composition that arises from the environmental policy, if the market were to operate in the short-run with no carbon price. ΔE_{int} can be thought of as the additional reductions in emissions obtaining from introducing a carbon price to short-run market clearing, holding changes in the capital composition fixed. This decomposition allows insight into how the environmental policy affects the carbon intensity of the installed generation mix over time.

The decomposition analysis shows that 28 percent of the emissions reductions achieved from a carbon price are achieved at the extensive margin, if a carbon price were to supplant the wind production subsidy in the concentrated scenario (comparing emissions from counterfactual (3) with those from counterfactual (1)). Figure 1.11 illustrates that these extensive emissions reductions, totaling 110 million tons over the entire sample period, are well in excess of the 63 million tons that I estimate resulted from a wind production subsidy (comparing emissions from counterfactual (2) with counterfactual (1)). The difference is driven largely by the buildout of new natural gas capacity, which replaces older, more carbon intensive fossil generation in the dispatch order. The new natural gas capacity buildout is encouraged by the carbon price signal

and does not arise to the same degree under the wind subsidy, as shown in Figure 1.10c.

Figure 1.11: Cumulative Extensive Emissions Reductions,
Carbon Price vs. Wind Production Subsidy

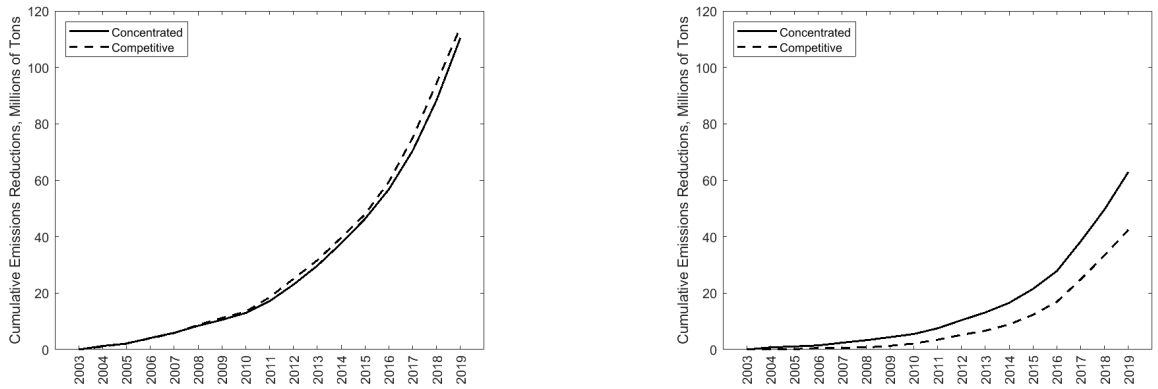


Emissions reductions attributable to changes in capital composition under a $20 carbon price (counterfactual (3)) are 110 million tons from 2003-2019 vs. 63 million arising from a wind production subsidy (counterfactual (2)). Reductions are calculated relative to counterfactual (1).

Next I examine the effects of market concentration on extensive emissions reductions. Figure 1.12 compares cumulative emissions reductions at the extensive margin under each of the two environmental policies (a carbon price and a wind production subsidy) with and without the competitiveness assumption. Figure 1.12a shows a nearly identical path under a carbon price whether the market is competitive vs. concentrated, despite the large difference in the amount of capacity retired. This result reflects infrequent dispatch of plants which are retired in the concentrated scenario, and that wind and natural gas capacity buildout are comparable under both the competitive and concentrated scenarios which involve a carbon price. In contrast, under a wind

production subsidy, imperfect competition leads to greater emissions reductions at the extensive margin relative to the competitive counterfactual. This result follows from the relatively lower amount of fossil fuel capital turnover in the competitive scenario, leading to greater use of the existing, carbon intensive generators.

Figure 1.12: Cumulative Extensive Margin Emissions Reductions: Competitive vs. Concentrated Counterfactuals



(a) $20 Carbon Price

(b) Wind Production Subsidy

Emissions reductions attributable to changes in capital composition under environmental policies, with and without market concentration.

1.8 Conclusion

I have demonstrated the economic significance of long-run market power incentives in determining the evolution of electricity market structure over time. Long-run market power exercise may upset the fundamental economics undergirding market design for efficient investment, thereby affecting outcomes associated with policies for ensuring reliability and decarbonizing electricity systems. While electricity market designs vary across systems in the United States, these fundamentals apply broadly in deregulated markets. The effects are most striking when it comes to early retirement of existing capacity, but implications for future investment are increas-

ingly important as "deep decarbonization" becomes a more urgent policy objective that is best achieved through robust competition among firms specializing in advanced renewable technologies and electricity storage.

My counterfactual analysis adds to a large literature demonstrating the efficiency of pricing pollution externalities as compared with the use of a subsidy. While federal production tax credits for wind generators encouraged a large buildout of wind plants in Texas, the attendant reduction in emissions associated with these plants is small relative to that resulting from a carbon tax which could have achieved buildout of wind at a comparable scale, and which reduces emissions at a significantly lower cost per ton. Changes in emissions which arise at the intensive margin are also critical to appreciate with respect to environmental policy. Despite finding significant early plant retirements associated with market power exercise, changes in emissions volume associated with those retirements are relatively small absent a direct externality price.

This paper invites several avenues of further research. In the context of electricity markets, one area of interest is an assessment of scarcity pricing mechanisms such as the one employed by ERCOT, addressing the question of how a design based on theoretical principles of optimal investment may be affected by long-run market power incentives (see Holt, 2022b). With respect to environmental policy research, I have examined only two environmental policy instruments. Future work may take a broader look at second-best policies for reducing carbon emissions that are of practical policy interest such as clean electricity standards, and how these policies interact with long-run market power incentives. While my analysis covers 2003-2019, modeling beyond this time period is also of interest in understanding the progression of a decarbonized electricity system. Such an analysis may incorporate the increased presence of solar power and grid-scale storage technologies, which are being rapidly deployed presently, beyond the sample period.

Chapter 2: Scarcity Pricing and Imperfect Competition in Wholesale Electricity Markets

2.1 Introduction

Optimal pricing in wholesale electricity markets ensures efficient investment in power generation capacity. A fundamental theoretical result suggests that optimal pricing is achieved by setting the wholesale price equal to the "value of lost load" during periods of supply scarcity. This result is derived under an assumption of perfect competition. While it has been acknowledged that the presence of market power could affect this result, surprisingly little work has been undertaken to further understand how. Practical applications of this fundamental theoretical principle include "energy only" market designs currently in operation in various markets around the world, including Texas, Australia, New Zealand, and Singapore. In these markets, the notion of imperfect competition upsetting the efficiency of capacity investment does not seem to be a significant concern among system operators, though other types of market power exercise are monitored carefully. Nevertheless, whether the energy-only market design ensures sufficient resource adequacy and reliability remains a point of contentious debate.

Few (I am not aware of any) empirical analyses have been undertaken to assess the *ex post* effectiveness of scarcity pricing designs, or to understand how such designs interact with imper-

fect competition. In this paper I use a stylized model based on the scarcity price mechanism (SPM) design and market conditions of the Electricity Reliability Council of Texas (ERCOT) to examine the implications of imperfect competition on capacity investment outcomes. I parameterize the model using values from ERCOT planning documents and ERCOT's observed load (demand) distribution. This simple modeling serves to demonstrate that price effects associated with capacity investment are significant in the presence of an SPM to the extent that firms above a modest threshold size will find further investment unprofitable. Moreover, because many firms in ERCOT were structured prior to deregulation, their firm size upon deregulation (and thereafter) was not a result of market competition. My analytical modeling highlights how an SPM can encourage large firms to retire capacity to capture greater rents.

I then use a dynamic structural model of firms' power plant investment and retirement decisions to empirically assess the effectiveness of the SPM. Consistent with the analytical results, I find that on net, large firms retired more capacity in the presence of the SPM than in its absence over the period 2014-2019. These retirements, however, are somewhat offset by the addition of new capacity from smaller competitors in response to the SPM. In particular, I find that 2,400 MW of additional fossil fuel capacity was retired as a result of the SPM, while 1,400 MW of natural gas capacity and 1,600 MW of wind capacity was added from 2014-2019 as a result of the mechanism.

This paper complements the theoretical literature on electricity market pricing for efficient investment. Seminal work in Boiteux, 1960, Turvey and Anderson, 1977, Chao and Wilson, 1987 establishes the fundamental efficiency result. The efficiency result is analogous to externality pricing in that the optimal spot price internalizes the social cost caused by a failure to meet demand. In a concise summary, Stoft, 2002 points to an important shortcoming of the theory,

noting that "[a]lthough basic theory ignores risk and market power, it provides valuable insights and a basis for discussing more subtle theories of setting energy prices."

A recurring theme in these works is the characteristic lack of demand response in electricity markets, which prevents settlement prices from reflecting consumers' preferences for reliability. As such, administratively implemented SPMs were recognized as a means for achieving reliability. Hogan, 2005 and 2013 present the case for an SPM which was ultimately implemented in ERCOT. These works address the market power issue explicitly, but put most emphasis on short-run market power. Joskow and Tirole, 2007 consolidates much of the theoretical work in examining the relationship between non-market-based features of electricity markets (such as reliability requirements, price caps, and contracting obligations) and optimal pricing and investment. The article includes a discussion of oligopolistic competition in the market for peak-demand power generation, but imperfect competition is not a central focus. My paper builds on these works by examining how the presence of imperfect competition might upset the efficiency of scarcity pricing designs in the "long-run" sense that firms can effectuate different pricing outcomes through their power plant investment and retirement decisions.

This long-run market power incentive was recognized in Myatt, 2017, which uses a structural oligopoly model to demonstrate a large incentive for the early retirement of coal plants. My paper is the first that I am aware of that closely examines these incentives as they relate to scarcity pricing.

This paper also complements the literature that directly examines the effects of an SPM on investment outcomes. Carden and Dombrowsky, 2021 is one example of regular analysis undertaken on behalf of ERCOT to determine proper calibration of its SPM and reserve margin targets. These assessments typically estimate equilibrium investment based on annualized capital costs,

and do not contemplate economic barriers to entry or market concentration. Other empirical assessments of ERCOT's SPM include Bajo-Buenestado, 2021, which notes that increased wind penetration diminishes rents from the ERCOT SPM, thereby negatively affecting investment incentives. Levin and Botterud, 2015 examine the effectiveness of ERCOT's SPM relative to other designs in the presence of increasing generation from intermittent renewable sources. I add to these works by taking an industrial organization perspective, directly addressing how imperfect competition may upset the intuition which forms the basis of the ERCOT market design.

2.2 A Simple Model of Capacity Investment

2.2.1 Model Overview

There are $h = 0, 1, ..., H$ hours in the year and planning occurs over a one-year horizon without discounting. Load is conditional on the previous period: next period load L_{h+1} is a random variable distributed over the interval $[0, 1]$ according to the distribution $f(\cdot|L_h)$ (load in period 0 can be taken as given). Denote installed generation capacity of a homogeneous technology $G \in [0, 1]$. The probability that load is greater than capacity (a "scarcity event") is equal to $1 - F(G|L_h)$. When load exceeds generation capacity, load of the amount $L_h - G$ is unmet, or "lost." Denote the value of lost load (VOLL) as v, the marginal social cost incurred for each unit of demand that is unmet. I assume zero variable cost and linear marginal investment costs κ. This model is similar to the Boiteux-Turvey style model used in Schmalensee, 2019 and, as the next section will highlight, is a general form of the actual SPM as it is employed in ERCOT.

2.2.1.1 Social Optimum

A social planner may wish to find the social cost-minimizing amount of generation capacity G according to the following objective function:

$$\min_{G} \sum_{h} \int_{G}^{\infty} (X - G) f(X|L_h) dX + \kappa G \tag{2.1}$$

Taking the first order condition, the optimum is realized when:

$$v \sum_{h} [1 - F(G|L_h)] = \kappa \tag{2.2}$$

The second order condition is $v \sum_h f(G|L_h) > 0$ which implies that the solution is cost-minimizing. Perfect competition ensures that the optimum is realized when the price of electricity p is set equal to v. This outcome reflects the classical result of optimal investment in electricity markets. Define the competitive equilibrium as G^C which satisfies:

$$p \sum_{h} [1 - F(G^C|L_h)] = \kappa \tag{2.3}$$

2.2.1.2 Oligopolistic Competition

In an imperfectly competitive setting with firms $i = 1, ..., N$, firm i's objective function under VOLL pricing ($p = v$) can be expressed as:

$$\max_{G_i} p \sum_{h} [1 - F(G_i + G_{-i}|L_h)] G_i - \kappa G_i \tag{2.4}$$

where $G_{-i} = \sum_{j \neq i} G_j$ and G_i is firm i's capacity choice. The resulting first order condition is:

$$p \sum_h [1 - F(G^{IC}|L_h) - f(G^{IC}|L_h)G_i] = \kappa \tag{2.5}$$

for all firms i, where $G^{IC} = \sum_{i=1}^{N} G_i$ is the equilibrium market capacity under imperfect competition. Note that $G^C > G^{IC}$ since $F(G|L)$ is increasing in G and $f(G|L)G_i > 0$. This simple result demonstrates that market concentration results in sub-optimal capacity investment (consistent with typical oligopoly models).

2.2.1.3 Oligopolistic Competition with Endowments

Suppose that firms do not start at zero capacity and make a choice of how much to install, but rather start with endowments $\omega_i \geq 0$ and can either add capacity $x_i \geq 0$ or subtract capacity $y_i \geq 0$ resulting in final capacities $G_i = \omega_i + x_i - y_i$. Firms incur marginal investment cost κ when they add capacity, and retirement cost η when they subtract capacity. This setup is analogous to the real-world scenario in which regulated utilities built up portfolios of generation assets and subsequently became engaged in a deregulated competitive wholesale market (wholesale deregulation occurred in Texas in 1995). The endowments represent the firms' pre-existing capacities upon entering a competitive market. The firm objective function can be written as:

$$\max_{x_i, y_i} p \sum_h [1 - F(\omega_i + x_i - y_i|L_h)](\omega_i + x_i - y_i) - \kappa x_i - \eta y_i \tag{2.6}$$

subject to:

$$x_i \geq 0, y_i \geq 0, \omega_i + x_i - y_i \geq 0$$

The Karush-Kuhn-Tucker formulation of the problem:

$$L = p\sum_{h}[1 - F(\omega_i + x_i - y_i|L_h)](\omega_i + x_i - y_i) - \kappa x_i - \eta y_i + \lambda(\omega_i + x_i - y_i) \tag{2.7}$$

The interior solution first order conditions are:

$$p\sum_{h}[1 - F(G|L_h) - f(G|L_h)(\omega_i + x_i)] = \kappa \tag{2.8}$$

if $x_i > 0$, and

$$-p\sum_{h}[1 - F(G|L_h) - f(G|L_h)(\omega_i - y_i)] = \eta \tag{2.9}$$

if $y_i > 0$.

2.2.2 An Operationalized Model

ERCOT planning documents define a function analogous to $F(G|L)$ used in the simplified model above. The probability of lost load in hour $h + 1$ is given according to a normal distribution conditional on the amount of spinning and non-spinning reserves available in hour h. The function also has a discontinuity built in which increases the scarcity price to the value of lost load if reserves go below an administratively determined threshold X. Specifically, the scarcity price function is $p_h^S = v[1 - H(R_h)]$ where:

$$H(R_h) \equiv [\frac{1}{2}H_S(R_{S,h}) + \frac{1}{2}H_{NS}(R_{NS,h})] \tag{2.10}$$

and

$$H_S(R_{S,h}) = \begin{cases} \Phi(\frac{R_{S,h}-X-0.5\mu_S}{0.707\sigma}) & \text{if } R_{S,h} > X \\ 1 & \text{if } R_{S,h} \leq X \end{cases} \tag{2.11}$$

$$H_{NS}(R_{NS,h}) = \begin{cases} \Phi(\frac{R_{S,h}+R_{NS,h}-X-\mu_S}{\sigma}) & \text{if } R_{NS,h} > X \\ 1 & \text{if } R_{NS,h} \leq X \end{cases} \tag{2.12}$$

$R_{S,h}$ is the level of spinning reserves in hour h; $R_{NS,h}$ is the level of non-spinning reserves in hour h; X is an administratively set reserve threshold; and σ, μ_S are determined based on historical fluctuations of hourly load.

I make two adjustments to the function employed by ERCOT to make it tractable for the purposes of my analysis. First, I abstract away from the distinction between spinning and non-spinning reserves ("system reserves"). Reserves in concept are equal to the amount of available dispatchable generation capacity minus the amount of net load (load net of generation from renewable sources). In practice, this relationship is complicated by whether reserves are "spinning" and therefore immediately deployable, or "non-spinning" and therefore require 30 minutes to become available. ERCOT uses a weighted average function of spinning and non-spinning reserves, as shown in Equations 2.11 and 2.12 above. Distinct from these system reserves, I define "economic reserves" for use in this model as the difference between installed generation capacity and net load, or $G - L_h$. This approach is consistent with models used in the literature (e.g. Stoft, 2002). One can think of generation capacity in this sense as being standardized based on how much it contributes to the availability of system reserves.[1] Second, I remove the administratively-

[1]For example, one MW of wind generation provides a different contribution to system reserves than one MW of natural gas peaker capacity. ERCOT planning documents use conversion factors to approximate the contribution of reserve availability across different technologies.

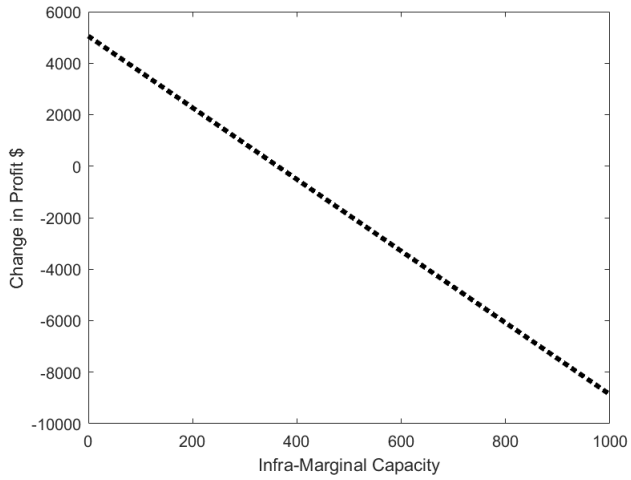
set discontinuity, thereby allowing the function to be differentiable. This adjustment will understate the amount of revenue firms will earn from scarcity pricing, and therefore understate price effects associated with a change in system capacity.

The resulting scarcity price function I employ in the context of the formulae presented in Section 2.2.1 is as follows:

$$F(G|L_h) \equiv \Phi\left(\frac{G - L_h - X - \mu_S}{\sigma}\right) \tag{2.13}$$

where $\Phi(\cdot)$ is the standard normal CDF.[2] Figure 2.1 compares the scarcity price used in my model vs. that used by ERCOT (while still abstracting away from spinning vs. non-spinning reserves).

Figure 2.1: Comparison of Price Mechanism used in Modeling vs. used by ERCOT



The simplified model series reflects the price obtained using Equation 2.13: $p = v[1 - F(G|L)]$. The ERCOT series reflects Equation 2.10: $p = v[1 - H(R)]$ where R holds non-spinning reserves fixed at zero and sets spinning reserves equal to $G - L$.

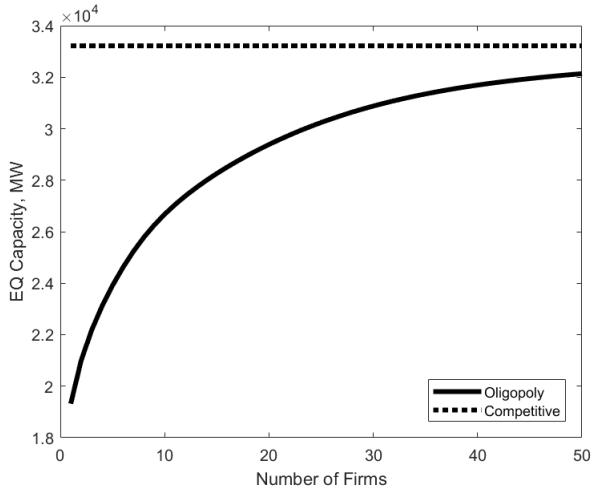
[2] This implies $f(G|L_h) = \frac{1}{\sigma}\phi\left(\frac{G - L_h - X - \mu_S}{\sigma}\right)$.

2.2.2.1 Equilibrium Comparison

I use observed hourly net load L_h from 2017 and suppose that firms compete to build capacity for 35,000 MW (roughly 30 percent of the ERCOT system).[3] I use annualized investment costs from ERCOT planning documents to inform κ (the annualized cost of new entry). Parameters X, μ_S, and σ are also taken from ERCOT planning documents.[4]

Figure 2.2 illustrates the difference in equilibrium investment under perfect competition, and under imperfect competition for n=1,...,50 symmetric firms. Equilibrium capacity under oligopolistic competition is significantly lower than the perfectly competitive outcome, even with a large number of firms.

Figure 2.2: Equilibrium Capacity Investment, Perfect Competition vs. n-firm Oligopoly



[3]This choice of level is without loss of generality because any economic reserves level in excess of 35,000 MW implies a scarcity price of zero.

[4]In particular, $\kappa = \$93,500, X = 2,000, \mu_S = 246.01, \sigma = 1,275.11$.

2.2.2.2 Duopoly Mutual Best Responses

Figure 2.3 illustrates mutual best response functions under a duopoly, when both firms start with zero endowments $\omega_1 = \omega_2 = 0$. Equilibrium capacity investment occurs at the intersection of the best responses.

Figure 2.3: Mutual Best Responses, $\omega_1 = \omega_2 = 0$

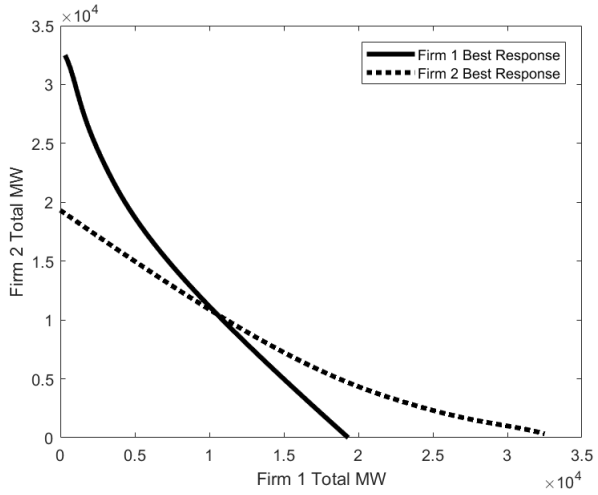
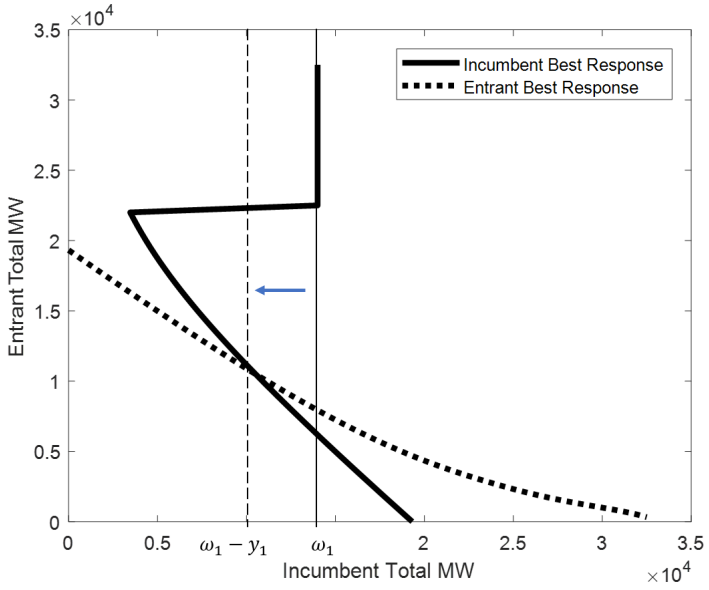


Figure 2.4 presents a simple but important result that is central to this paper: a firm with a large endowment may wish to reduce capacity. For ease of illustration I assume $\eta = \kappa$, or that removing 1 MW capacity costs the same as installing 1 MW capacity. Figure 2.4 illustrates mutual best response functions under a duopoly when Firm 1 has an initial endowment of $\omega_1 = 12,000$ MW. Firm 1's best response function is shown by the thick solid black line and its initial endowment is shown by the thin vertical solid line. If Firm 2's capacity were sufficiently low, it would be optimal for Firm 1 to add capacity (as shown when its best response quantity is greater than its initial endowment). However, Firm 2's best response function (shown by the dotted line) intersects Firm 1's best response function to the left of Firm 1's endowment. It is therefore optimal for Firm 1 to reduce its capacity to $\omega_1 - y_1$ as indicated by the vertical dashed line.

Figure 2.4: Mutual Best Responses, $\omega_1 = 12,000, \omega_2 = 0$



2.3 ERCOT Application

The analytical model presented in the previous section examines the SPM in a vacuum. Many other factors, such as spot prices (prior to the addition of the scarcity price), economic barriers to entry, and technological heterogeneity will also determine how firms respond to the scarcity price signal. Testing whether the more general results highlighted above will hold in a real-world setting requires modeling that takes into consideration a complex market structure and dynamics. In this section I employ a detailed dynamic structural model capable of capturing important industry features to conduct an *ex post* analysis of the ERCOT SPM.

2.3.1 ERCOT Market Overview

I begin with a high-level overview of firm behavior in ERCOT which appears consistent with the economic intuition described above. Using plant ownership records collected by the

Public Utility Commission of Texas, I summarize the size of firms (in terms of MW capacity owned) making investment and retirement decisions. ERCOT implemented its SPM in the summer of 2014. Since then, new gas-fired power plants have been added almost exclusively by small firms. Table 2.1 illustrates that of the seven new fossil fuel fired power plants added from 2014-2019, only two firms had non-zero infra-marginal capacities prior to investing in the new plant. Moreover, fossil fuel based retirements in ERCOT have come mostly from larger firms, as shown in Table 2.2.

Table 2.1: New Fossil Fuel Plant Additions in ERCOT, 2014-2019

Year Added	**Firm**	**New Plant Type**	**Firm Size Before Build (MW)**
2014	Panda	Combined Cycle	0
2015	Invenergy	Peaker	0
2017	Exelon	Combined Cycle	3771.5
2018	Agilion Energy	Peaker	0
2018	BTEC	Peaker	0
2018	Quantum Utility	Peaker	0
2019	Agilion Energy	Peaker	242

Table 2.2: Size of Firms Retiring Plants 2014-2019

Year Retired	**Firm**	**Firm Size Before Retirement MW)**
2014	NRG	9,326
2017	Blackstone	705
2018	Luminant/Vistra	16,577
2019	City of San Antonio - (TX)	3,712
2019	Luminant/Vistra	15,002

An analysis of firm incentives at the margin also serves to illustrate the nature of the price effects introduced by ERCOT's SPM. The price adder is system-wide, therefore any units that are operating at the time the scarcity price is induced will earn revenue equal to the product of the

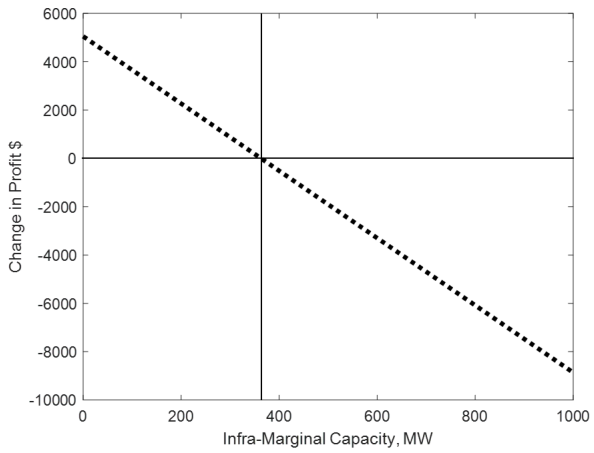
scarcity price and the amount of power being generated. Defining the profit function with respect to system reserves as:

$$\pi_i(R_S, R_{NS}) = \sum_h p[1 - H(R_{S,h}, R_{NS,h})]G_i \tag{2.14}$$

using $H(\cdot)$ as defined in Equation (2.10) and observed system reserve levels from 2017, Figure 2.5 plots the relationship between $\partial\pi_i/\partial R_S$ (the amount of revenue a firm would earn if it were to add one megawatt of additional spinning reserve capacity to the system for the entire year, all else fixed) and the amount of infra-marginal generation G_i operated by the firm. The figure suggests a threshold amount of infra-marginal capacity at which one additional unit of reserve capacity on the system implies a loss in revenue. This threshold occurs at approximately 400 MW, suggesting that for any firm with existing assets of a modest size (the average peaker plant, for example, is approximately 200 MW) the SPM makes investment in new capacity unprofitable.

This relationship is illustrated at observed levels of system reserves; in equilibrium, one would expect rents to be competed away and therefore an unprofitable outcome from adding system reserves at the margin is perhaps not surprising. However, even at lower reserve levels, a fairly low zero-profit threshold is realized. For example, repeating the same exercise when system reserves are uniformly 2,000 MW lower for the duration of the year, a firm with approximately 600 MW of infra-marginal capacity would fail to earn revenue from adding one unit of reserves to the system.

Figure 2.5: Marginal Revenue from Additional Unit of Reserves
vs. Infra-marginal Capacity



2.3.2 Structural Modeling Results

I employ the model used in Holt, 2022a to perform counterfactual simulations of the Texas electricity market. In this model, firms make yearly investment and retirement decisions based on expectations about the future industry path and profit flows. The model is suitable for this analysis because it takes into account a number of market features which determine plant capacity investment and retirement decisions. The structural model reflects estimated entry and exit costs which are consistent with realized firm behavior from 2003-2019. A nested short-run profit model captures hourly dispatch dynamics across hundreds of heterogeneous generation units. Perhaps most importantly, the equilibrium concept used in the model allows that each firm's decisions reflect the collective anticipated decisions of its rivals. These features taken together allow the model to capture how the market will respond to each decision that is taken, including the extent to which plant capacity that is retired by one firm will be replaced by new capacity from competing firms.

A limitation is that "system reserves"—spinning and non-spinning reserves sold on the ancillary services market in ERCOT—are not explicitly modeled (this definition of reserves is distinct from that of "economic reserves" which I use throughout this paper). Plant operators face a decision to use their dispatchable generation capacity to sell electricity into the energy market, or dedicate it for use as system reserves, sold in the day-ahead market as an ancillary service. For example, an operator of a peaker unit can make bids to generate power during peak demand through the real-time market electricity auctions, or can instead bid its peaker capacity into the ancillary services market to offer system reserves. System reserves refer to generation capacity that is available for immediate dispatch but is not contributing power to the system; system reserves can be called upon to provide power if needed. Economic reserves, in contrast, are the difference between installed dispatchable generation capacity and net load, or $G - L$ as used in the literature and in my stylized modeling above. By not modeling this decision, the firm's objective function inherent in the model does not account for incentives with respect to the value of offering reserves. While the modeling allows insight into how installed capacity changes over time, this approach precludes an analysis of reliability in any specific sense—for example, this modeling does not predict the occurrence of blackouts arising from insufficient system reserves.

This modeling approach does not examine short-run market power exercise, focusing instead on long-run responses to the SPM. The extreme price sensitivity introduced by the SPM raises the question of whether firms can induce greater scarcity prices through capacity withholding in short-run market operations, a question which represents an interesting avenue for future research. This and other strategic short-run behavior only affects the analysis presented in this paper to the extent that it influences long-run decisions. Policing of short-run market power, performed both by an independent market monitor and ERCOT itself, suggests firms' opportunities

to extract rents in this manner would be limited.[5] However, evidence suggests that generators may nevertheless behave strategically when the opportunity presents itself, particularly during periods of localized transmission congestion (Woerman, 2019).

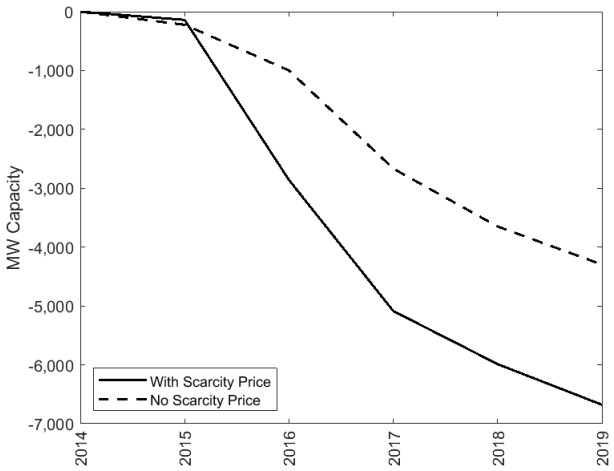
Counterfactual analysis begins with a baseline scenario in which electricity settlement prices reflect the SPM as it was implemented in the actual world. I simulate forward from 2014 (the year the SPM was introduced) through 2019, recording firms' decisions to retire existing power plants or build new wind or combined cycle natural gas capacity. I then repeat this exercise, omitting the SPM from the short-run profit model, to simulate a counterfactual industry path in which there is no SPM. An overview of the implementation of the counterfactual modeling is presented in Appendix B.

Figure 2.6 illustrates the results of this modeling with respect to power plant retirements. The solid line represents the cumulative amount of fossil fuel-based generation capacity, in MW, retired from 2014-2019 under the SPM; the dashed line represents cumulative retirements without the SPM. Results suggest that approximately 6,700 MW of fossil fuel-fired capacity was retired under the SPM, where only 4,300 was retired in the counterfactual with no SPM. That is, 2,400 MW more capacity was retired as a result of the SPM. This result is contrary to the intended purpose of the mechanism (to improve resource adequacy) but consistent with the economic intuition described above; firms with a significant amount of infra-marginal generation capacity may profit from the early retirement of power plants if the attendant reduction in system-wide capacity profitably increases activation of the SPM. The ability of a firm to extract rents in this manner depends on entry and exit costs. If entry costs are sufficiently low, the retirement of

[5] An independent market monitor regularly reviews whether capacity on the system is underutilized as a means for detecting strategic capacity withholding. ERCOT mitigates supply offers if they are deemed excessive with respect to pre-determined generator-specific markups.

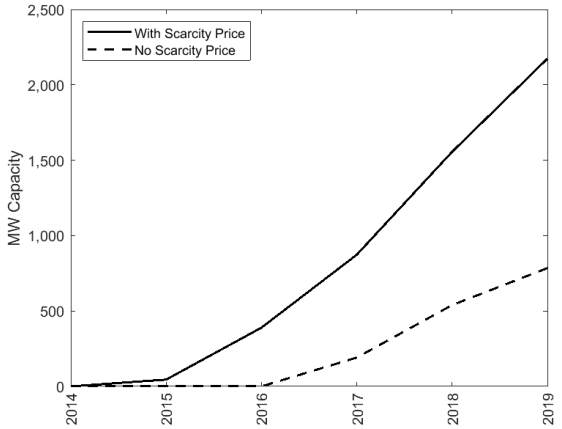
a plant will be readily offset by competitors. In the presence of economic barriers to entry, a window of opportunity may exist which makes it profitable to engage in this form of long-run market power exercise. Estimated entry and exit costs inherent to the model suggest that it is profitable for some firms to undertake such strategic behavior.

Figure 2.6: Simulated Retirements

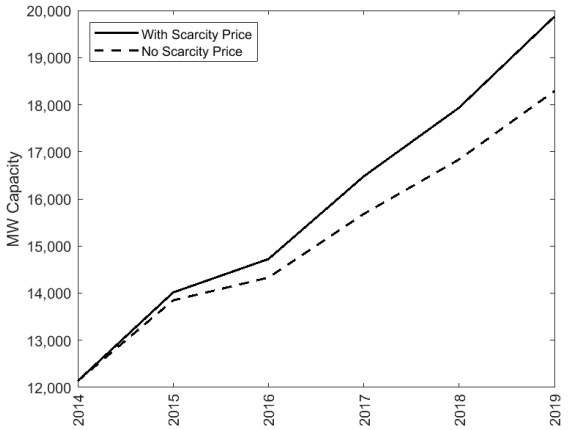


Figures 2.7a and 2.7b suggest that the SPM was effective at increasing investment in new natural gas and wind capacity. Natural gas and wind capacity additions in the presence of the SPM were approximately 1,400 MW and 1,600 MW greater, respectively, relative to the counterfactual scenario in which there was no SPM. This result is again consistent with the intuition described above; smaller firms (and in particular new entrants) will profit from adding new capacity to the system as a result of a scarcity price.

Figure 2.7: Simulated Natural Gas and Wind Investment Paths



(a) New Natural Gas Additions

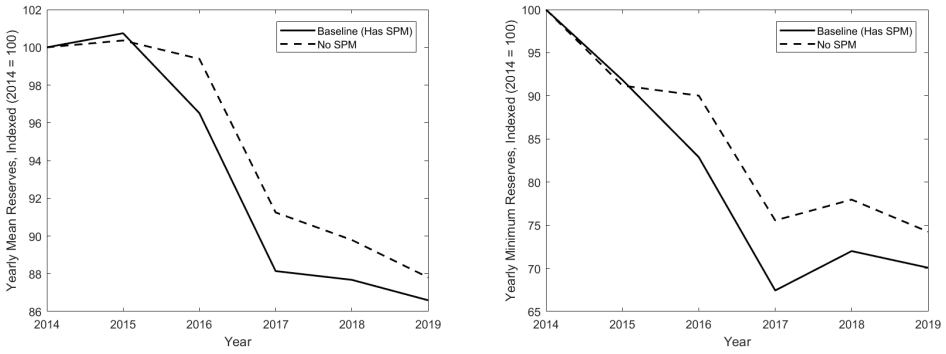


(b) Wind Additions

In light of these countervailing effects of the SPM—more capacity is retired, but more is also built—it is useful to examine net effects in terms of reliability. The structural model I employ captures economic reserves (availability generation capacity net of load), while system reliability depends on a slew of factors such as spinning vs. non-spinning reserves, unforeseen outages, and demand shocks which I do not model here. Nonetheless, an examination of economic reserves is informative as it provides a sense of how changes in the investment and retirement of heterogeneous technologies affect the overall availability of generation capacity relative to demand.

Figure 2.8 compares annual mean and minimum available reserves for each of the simulated scenarios. On net, reserves are slightly *lower* in the presence of the SPM compared to the counterfactual with no SPM. This disparity, though small in magnitude, may cut against the goal of improving reliability. However, the overall effect on reliability depends on the precise technological composition of the capacity being added and subtracted and its contribution to system reserves, which I do not analyze here. In both scenarios, economic reserves decrease over the sample period; this decrease corresponds with heavy fossil retirements and increasing demand.

Figure 2.8: Mean and Minimum Annual Economic Reserves



(a) Mean Economic Reserves (b) Minimum Economic Reserves

The capacity additions and subtractions depicted in Figures 2.6 and 2.7 are enumerated in columns (1) through (4) of Table 2.3. In both scenarios, significantly more fossil fuel-fired generation capacity is retired than is built. Upon simulating the baseline and counterfactual scenarios, I use the short-run model to compute profits, entry and exit costs, and consumer expenditures over the six-year period. Table 2.3 also summarizes results which provide insight into the welfare consequences of the SPM. These figures are expressed as totals over the six-year period, discounted to 2014 using a discount factor of 0.9. Short-run profits to producers (shown in Column (5)) are significantly greater ($9.9 billion) in the presence of the scarcity price mechanism; consequently,

consumer expenditures (column (10)) are greater by a comparable magnitude ($9.7 billion). Consistent with market power exercise, only $3.3 billion in additional investment is realized under the SPM, or approximately one third of the $9.9 billion welfare transfer from consumers to producers. Investment in wind increases 24 percent in terms of dollars spent under the SPM, and investment in natural gas more than triples.

Table 2.3: Counterfactual Simulation Results Summary

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Scenario	MW Fossil Added	MW Fossil Retired	Net MW Fossil Added	MW Wind Added	Producer Short-Run Profits	Wind Investment Costs	Nat. Gas Investment Costs	Retirement Costs	Producer Net Profits	Consumer Expend-itures
[A]	Baseline (with scarcity price)	2,174	6,678	-4,504	7,733	$35,470	$10,052	$1,973	$1,379	$22,067	$66,406
[B]	No Scarcity Price	785	4,296	-3,512	6,153	$25,587	$8,087	$666	$1,217	$25,500	$56,705
[A]-[B]	Difference	1,389	2,382	-993	1,580	$9,884	$1,964	$1,307	$162	-$3,433	$9,701

Profits, costs, and expenditures are expressed in millions of dollars, discounted to 2014 using a discount factor of 0.9.

2.4 Concluding Remarks

The analytical modeling and *ex post* analysis of ERCOT presented above suggests that an SPM accelerates capital turnover in a manner which may undermine its intended purpose of improving resource adequacy, and at a significant expense to consumers in the form of higher electricity prices. Indeed, the SPM may perhaps even degrade resource adequacy by inducing firms to retire power plants early as a form of market power exercise. The specific welfare implications associated with the greater turnover induced by an SPM depend on the particular features of the setting. In a simple setting with a homogeneous generation technology, it would be inefficient for some firms to incur costs to reduce their size while other firms incur cost to build capacity. In a more complex setting with heterogeneous technologies, these inefficiencies

may be counteracted by the environmental benefits of retiring old technology, or lower prices from installing capacity with lower variable costs. In the case of ERCOT, my analysis suggests that the SPM resulted in less overall dispatchable capacity and a significant transfer of welfare from consumers to producers. Several implications for policy and market design and avenues for further research follow.

From a policy perspective, the modeling I present above emphasizes that the effectiveness of an SPM requires low barriers to entry and a pool of competitive potential entrants. Any constraint on competitive entry may result in sub-optimal investment in capacity and supra-competitive settlement prices. This finding is of particular interest as regulators seek to overhaul interconnection processes which have created a bottleneck in the entry of new power plants.

This paper draws into question both the extent to which firms are incentivized to retire plants early, and are reluctant to invest in new generation capacity once they reach a certain size. Further research is needed to understand how these incentives will play out as wind, solar, and energy storage account for a larger portion of electricity supply. These technologies differentially affect scarcity pricing. Wind and solar, through their inherent intermittency, are most likely to affect prices during certain hours of the day. Utility-scale storage has the potential to directly offset scarcity events, thereby dampening the incentives introduced by the SPM.

Adjusting the SPM in a manner that corrects the market power incentive seems prohibitively difficult as it would require a firm-by-firm assessment of costs, markups, and counterfactual scenarios. Alternative approaches used by other system operators—such as a separate capacity market, or setting reliability obligations for load-serving entities—could be used (and have been considered) in place of the SPM. Under these approaches, capacity and reliability targets are predetermined, which may serve to undermine the market power incentive by removing the ability of

large firms to influence the amount of overall capacity on the market, and ensuring that investors will be compensated for new capacity at a rate that is determined through a competitive process such as an auction.

Chapter 3: Targeted Regulation for Reducing High-Ozone Events

3.1 Introduction

Ground-level ozone harms human health and the environment. Emissions of nitrogen oxides (NOx) and volatile organic compounds (VOCs) cause ground-level ozone, and decades of regulating these emissions from electricity generators, industrial facilities, vehicles, and other sources have dramatically reduced ozone levels in much of the country. The Environmental Protection Agency (EPA) assesses attainment with ozone standards based not on the average ozone level but on the frequency of ozone exceedance events, where currently an exceedance is defined as an 8-hour period over which the average ozone concentration exceeds 70 parts per billion (ppb). Despite reductions in average ozone levels and the frequency of exceedances, many regions, including areas inhabited by 37 million residents of the northeastern US, fail to reduce exceedances sufficiently to attain current ozone standards.[1]

To reduce these exceedance events and help states achieve the air quality standards, EPA enforces annual and seasonal NOx emissions caps for electricity and large industrial sources in the eastern US (the seasonal caps cover emissions from May through September). Essentially, the annual and seasonal emissions caps introduce a price on emitting NOx emissions from regulated

[1] Source: EPA Greenbook. The number in the text reflects populations of non-attainment counties in Connecticut, Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania, and Virginia.

sources, and all emissions face a uniform price regardless of when or where the emissions occur. Consequently, EPA could continue to tighten the annual and seasonal NOx emissions caps to reduce further the frequency of exceedances.

Alternatively, EPA could introduce spatially or temporally differentiated emissions prices (a temporally differentiated price may vary by hour of day or day of week, for example). Standard economic theory suggests that an emissions price is the efficient policy for a global or uniformly mixed pollutant such as carbon dioxide. The effect of NOx emissions on high-ozone events (which we define throughout this paper as hours in which ozone concentration exceeds 70 ppb) varies over space and time depending on the presence of VOCs, temperature, and sunlight. For example, high ambient air temperature exacerbates the marginal effect of NOx emissions on the probability of a high-ozone event. Consequently, a temporally or spatially varying emissions price that accounts for variation of the marginal effects of NOx emissions could be more efficient than a uniform price. However, Fowlie and Muller, 2019 point out that the marginal effect of emissions must be determined in advance. If the actual effect differs from the expected effect, the uniform emissions price may be more cost effective than the differentiated price.

Motivated by this theory, we consider an alternative to tightening existing emissions caps: a differentiated NOx emissions price that varies according to the hourly and region-specific propensity to cause high-ozone events. We ask whether a spatially and temporally differentiated emissions price would be more cost effective than tightening the emissions caps.

Comparing these two types of policies requires predicting the relationship between NOx emissions and high-ozone events as well as estimating marginal abatement costs. Moreover, we need to allow this relationship and marginal abatement costs to vary spatially and temporally. We are not aware of spatially and temporally varying estimates of the effect of NOx emissions on

high-ozone events or marginal abatement costs.

We implement novel empirical strategies to estimate these marginal effects and marginal abatement costs. We begin by assembling data on emissions, ozone concentrations, and allowance prices from the annual and seasonal cap-and-trade programs. The data cover the Northeast and Mid-Atlantic, which have experienced some of the highest ozone levels in the eastern US. The primary data sources are the EPA Continuous Emissions Monitoring System (CEMS) for hourly unit-level emissions, generation, and fuel consumption at fossil fuel-fired electricity generators; the EPA Air Quality System (AQS) for hourly ozone concentrations; the National Oceanic and Atmospheric Administration (NOAA) Integrated Surface Database (ISD) for hourly meteorological data; and S&P Global and Evolution Markets for NOx allowance prices. The combined data set includes hourly observations from 2001-2019 (except for the allowance prices, which are daily and cover 2015-2017).

We use the emissions, ozone concentrations, and meteorological data to estimate the effect of NOx emissions on high-ozone events, allowing this effect to vary across four sub-regions within the Northeast, hour of the day, and five-year sub-periods between 2001 and 2019. The key explanatory variable is the hourly NOx emissions across fossil fuel-fired generators located within 100 miles of the monitor. The dependent variable is an indicator equal to one for a high-ozone event, meaning that the hourly concentration exceeds 70 ppb. We use the indicator variable rather than the level of ozone because an area's compliance with the air quality standard is evaluated based on the number of days on which ozone levels exceed the threshold (Section 3.4 provides further explanation for the rationale for these measures of emissions and high-ozone events).

A challenge to estimating the effect of emissions on high-ozone events is that ozone levels

depend on other factors besides emissions from electricity generation, such as vehicle and industrial facility emissions. To account for the endogeneity of electricity generator emissions, we follow Johnsen, Lariviere, and Wolff, 2019 and instrument for emissions using a least-cost dispatch model. The instrument is constructed using the model from Linn and McCormack, 2019, and it varies because of time-varying coal, natural gas prices, and electricity demand interacting with variation across generators in their fuel type and efficiency. The emissions instrument is a strong predictor of observed emissions because the dispatch model accurately predicts hourly generation levels. As in Johnsen, Lariviere, and Wolff, 2019, the validity of the instrument rests on the assumption that spatial variation in generator fuel types and marginal costs is uncorrelated with omitted variables (that is, somewhat analogous to a Bartik-style instrument).

Across the four sub-regions, we find fairly consistent patterns. Emissions during late-afternoon hours have the largest effect on ozone levels, and early morning emissions affect ozone levels by about one-third as much as during late afternoon. For all hours of the day, the estimated effects are statistically significant at the one percent level. The effects vary relatively little across the sub-regions within the Northeast, but they vary by factors of three to four across five-year sub-periods. The temporal variation is more stable across time periods than the spatial variation, which demonstrates the importance of considering temporal as well as spatial variation.

Next, we estimate the marginal abatement cost curves. Three of the four sub-regions faced seasonal emissions prices under the Cross State Air Pollution Rule. A region's marginal abatement cost curve is the marginal cost of reducing emissions across all generators located in the sub-region, where abatement is measured relative to the level of emissions if there is no emissions price.

Existing models of NOx emissions abatement, such as the EPA Integrated Planning Model,

characterize investments in abatement equipment. For example, these models estimate investment in technologies such as low-NOx burners, which reduce temperature and NOx formation in the power plant's boiler. However, a spatially or temporally differentiated emissions price could affect emissions through mechanisms other than installing abatement equipment. For example, a power plant operator facing a time-varying emissions price may adjust boiler operation to reduce its emissions during periods of the day when the emissions price is high. The time-varying price may also affect generator utilization, causing a shift from high- to low-emitting sources. Consequently, relying on conventional models to characterize emissions abatement could omit some of the emissions response to a time-varying emissions price.

We take a reduced-form approach to estimate marginal abatement costs by regressing the hourly average NOx emissions rate in a sub-region on the daily NOx allowance price, allowing the price coefficient to vary by sub-region and hour-of-day. Importantly, the price coefficients include responses of generator operation and dispatch. We estimate each region's marginal abatement costs under two assumptions. First, marginal abatement costs are linear in the level of abatement, and second, permit markets are competitive and all firms choose levels of abatement such that marginal abatement costs equal the permit price. Under these assumptions, we estimate the slope of marginal abatement cost curves for each hour of the day and sub-region. We find that abatement costs vary somewhat across hours of the day and more so across sub-regions.

Using the estimated NOx-ozone relationship and marginal abatement costs, we compare a uniform emissions price and an emissions price that varies across hours and sub-regions in accordance with the estimated effect of emissions on high-ozone events. The regulator chooses the emissions price at the beginning of the year, and we estimate the effect of the price on emissions and high-ozone events as well as abatement costs. To facilitate comparison, we calibrate

the uniform and differentiated emissions prices to achieve the same predicted reduction in the frequency of high-ozone events. Because the regulator must set the emissions prices in advance, the differentiated price may be less cost effective than the uniform price if the NOx-ozone relationship is different from expectations at the time the prices are chosen. To examine the effect of imprecisely set emissions prices, we define two differentiated prices: an *ex ante* price, which is based on information prior to the implementation of the policy about the statistical relationship between ozone and NOx, and an *ex post* price, which is based on updated, and therefore more accurate, information about the NOx-ozone relationship. We find that both the *ex ante* and *ex post* prices are 4 to 18 percent less costly than the fixed price in reducing the occurrence of high-ozone events.

This paper contributes to literature examining differentiated environmental taxes (Muller, 2012, Fowlie and Muller, 2019) that stems from the foundational literature on market-based regulation (Baumol et al., 1988, Montgomery, 1972, Tietenberg, 1990). Fowlie and Muller, 2019 show that a spatially differentiated tax for addressing heterogeneous damages welfare dominates a uniform tax in the presence of perfect information; under uncertainty the gains are ambiguous. This and related research often focus on spatially heterogeneous damages. An important distinction of our analysis is that we consider temporal as well as spatial variation in the marginal effect of emissions on high-ozone events. Our results also differ from 2019 in that we find evidence that the spatially and temporally differentiated emissions prices outperform the uniform prices. Our results differ from theirs because we estimate larger temporal variation in the effects of emissions on high-ozone events than spatial variation. Moreover, the temporal patterns are relatively stable across years, reducing prediction error. These results cause the temporally differentiated emissions price to outperform the uniform price; 2019 consider only a spatially but not a temporally

differentiated price.[2]

We also contribute to literature on the statistical relationship between NOx and tropospheric ozone levels (e.g. Xiao, Cohan, Byun, and Ngan, 2010, Grewe, Dahlmann, Matthes, and Steinbrecht, 2012). We are not aware of other papers in this literature that address the endogeneity of observed NOx emissions using an instrumental variables strategy. Our results confirm findings in the literature of a complex and nonlinear relationship between NOx emissions and high-ozone events, and this relationship poses a challenge to policy makers, in that differentiated emissions prices are only cost effective if the estimated relationship is sufficiently accurate. A distinction between our analysis and most of the literature is that we consider how the NOx-ozone relationship has varied over the time period in which NOx emissions decreased dramatically, complementing research that examines the value of improvements in air quality during the same period (e.g. Holland, Mansur, Muller, and Yates, 2020).

In performing our policy analysis we estimate marginal costs of abating NOx emissions using econometric methods. These estimates serve to further highlight the distinction between economic abatement costs and estimates of abatement costs which are formed using engineering data (e.g., Linn and McCormack, 2019 and Fowlie and Muller, 2019). The advantage of the econometric approach is that abatement costs reflect observed rather than predicted behavior and include a large range of choices plant operators have in reducing emissions. A limitation of our approach is that our marginal abatement cost estimates reflect short-run costs. This approach is appropriate for our analysis because we are interested in understanding short-run responses to a dynamic emissions price that varies by the hour and that is announced a short period of time

[2]Our paper also differs from 2019 in that we evaluate cost-effectiveness of policies that achieve a common environmental outcome, whereas they characterize optimal Pigouvian taxes. Our approach is motivated by the structure of the air quality standards and that by statute, the standards are set without considering costs; consequently, the standards need not be set at the economically efficient level.

ahead. However, this approach cannot give insight into long-run responses to a differentiated emissions price, such as the installation of abatement equipment or plant closures. Nonetheless, the econometric approach may be useful in other policy contexts in which short-run abatement costs vary temporally or spatially. For example, NOx and sulfur dioxide emissions contribute to fine particulates, which are linked to a wide range of health problems. Estimating short-run marginal abatement costs for NOx and SO_2 could be useful for evaluating the near-term effects of policies aiming to reduce ambient levels of particulates.

3.2 Background

3.2.1 Overview of Ozone Formation, Health Effects, and Emissions Trading Programs

Tropospheric, or ground-level, ozone has been linked to mortality and morbidity (see, e.g., M. L. Bell et al., 2007, Ito, De Leon, and Lippmann, 2005) and harms plant life (see, e.g., J. N. B. Bell and Treshow, 2002, Davison and Barnes, 1998). NOx and VOCs are the primary ingredients of ozone. Although the process of ozone formation is highly complex, in broad terms, ozone formation occurs in two "regimes": a NOx-sensitive regime, in which an additional amount of NOx in the atmosphere is likely contribute to ozone formation; and a NOx-limited regime, in which ozone formation depends less on the additional NOx emissions, and more on the additional VOC emissions. Which regime a particular mass of air belongs to depends on the ratio of NOx to VOCs, as well as other factors, including the levels of each ingredient, the composition of the particular species of VOCs present, and meteorological conditions. Sunshine and high temperatures are the most important meteorological factors conducive to ozone formation, while

low windspeeds and low relative humidity are also associated with higher ozone concentrations (Sillman, 1999, Blaszczak, 1999).

Ozone pollution is increasingly harmful at higher levels of concentration. The EPA sets the National Ambient Air Quality Standard (NAAQS) for ozone, and requires that the three-year average of the year's fourth-highest 8-hour moving average concentration level be no greater than 70 ppb (Brown, Bowman, et al., 2013). In general, periods of high ozone concentrations tend to be ephemeral, lasting approximately 3.5 hours on average.[3] The structure of the ozone NAAQS motivates our focus on high-ozone events, because the greater the number of events, the more likely an area fails to attain the NAAQS.

Long-distance transport of pollution poses a substantial challenge in many areas. Power plants emit NOx from smokestacks, and prevailing winds may cause pollution from power plants to contribute to ozone levels hundreds of miles away. The Clean Air Act requires states to develop plans to achieve the ozone NAAQS. These plans include, among other tools, strict emissions permitting requirements for businesses located in the states, but many states have had trouble achieving the ozone NAAQS because they are downwind from power plant NOx emissions from other states.

To address the cross-state pollution transport problem, the EPA administers the Cross State Air Pollution Rule (CSAPR), which caps NOx emissions from much of the eastern half of the United States.[4] (The EPA also administers sulfur dioxide trading programs to address cross-state pollution, which are not the focus of this paper.) States are included in these programs if the EPA determines that emissions from those states contribute to ozone non-attainment in other states.

[3] In our 19-year sample, the median duration of ozone concentrations greater than 70 ppb is 3 hours, and the mean duration is 3.5 hours.

[4] The program includes Texas, Oklahoma, Kansas, Nebraska, and all states to the east except Florida and the New England states.

EPA finalized the CSAPR NOx trading program and set state-level emissions budgets (i.e. caps) in 2011, and has updated the budgets several times since. The most recent update set annual NOx emissions budgets on 22 states, and tightened the emissions caps for 12 of the 22 states, including Maryland, Pennsylvania, New Jersey, and New York, which are areas of focus in this paper.[5]

An important feature of the CSAPR annual and seasonal programs is that firms must submit allowances to cover their total annual or seasonal emissions. Because allowances can be traded, firms face a common allowance price for all of their emissions, regardless of how much those emissions contribute to ozone formation.

3.2.2 Policy Framework

We are interested in whether a temporally or spatially differentiated emissions price would be more cost effective at reducing high-ozone events than a uniform emissions price. This focus is motivated by the choice EPA faces: to help states achieve the ozone NAAQS, it could continue tightening the annual and seasonal caps, which would increase the allowance price (all else equal); or it could add a temporally differentiated emissions price. The differentiated price could be implemented by introducing trading rules such that emissions at certain times and from certain sources would be traded at a higher ratio. For example, one ton of NOx emitted between 4pm and 5pm could be counted as two tons for compliance purposes, meaning that the firm would need to submit two allowances to cover the single ton of emissions. Alternatively, EPA could replace the emissions caps with an emissions price that varies in accordance with the marginal effect of emissions on high-ozone events. Presently, we abstract from further implementation details and

[5]The 12 states for which the caps were tightened include Illinois, Indiana, Kentucky, Louisiana, Maryland, Michigan, New Jersey, New York, Ohio, Pennsylvania, virgnia, and West Virgnia.

include a brief discussion of implementation in the conclusion.

According to economic theory, a spatially or temporally differentiated emissions price would be more cost effective at achieving attainment under two conditions: a) the contributions of NOx emissions to non-attainment vary over space or time; and b) the regulator can accurately predict the variation. If the first condition does not hold, emissions make the same contribution to non-attainment regardless of where or when they occur, and a spatially or temporally differentiated price would not have any advantage.

Fowlie and Muller, 2019 discuss the second condition in the context of a spatially differentiated price, and the same argument applies here. If the variation can be predicted, the regulator can create clear rules for credit trading ratios that regulated firms can follow. However, if the variation is predicted with error, the advantage of the differentiated price would be undermined.

For example, suppose that based on modeling analysis performed at a particular time, EPA believes that emissions that occur at 2pm contribute more to non-attainment than emissions that occur at 3pm, and sets trading ratios accordingly. In that case, the temporally differentiated emissions price would cause more emissions abatement at 2pm than would the uniform emissions price (holding marginal abatement costs fixed across hours). However, EPA had to make assumptions about atmospheric conditions and other factors to perform its modeling analysis, and suppose for the sake of argument that those assumptions are inaccurate and emissions that occur at 2pm have the same effect on non-attainment as emissions that occur at 3pm. In this situation, the differentiated emissions price causes more abatement at 2pm than is cost effective.

A final consideration about comparing a differentiated and uniform emissions price is that it is necessary to estimate marginal abatement costs by location and time of day. Since the differentiated emissions price yields a different spatial and temporal pattern of abatement than

the uniform price, we need to estimate how marginal abatement costs vary over space and time to evaluate the cost effectiveness of the two policies.

This discussion motivates the three empirical questions we address in this paper. First, does the effect of emissions on ozone non-attainment vary over time of day. Second, can that variation be predicted accurately. Third, how do marginal abatement costs vary over time of day. After answering these questions, we can assess whether a temporally differentiated emissions price is more cost effective than a uniform price at reducing high-ozone events.

3.3 Data

We combine data from EPA, NOAA, and other sources to construct the data set used in our analysis. EPA's CEMS database records hourly generation, NOx emissions, and fuel consumption at the generation-unit level for every power plant larger than 25 MW (excluding industrial and commercial cogeneration plants). The AQS database records hourly ozone concentrations at hundreds of monitoring stations across the US. We use data from the 57 AQS stations located in non-attainment counties the Northeast and Mid-Atlantic. We use NOAA's ISD for hourly temperature data, mapping each AQS station to its most proximate NOAA station.[6] We use daily NOx allowance trading prices from S&P Global covering the period 2015-2017.

Recent literature demonstrates that local regulators strategically place monitors to facilitate compliance with air qulaity standards (Grainger, Schreiber, and Chang, 2018, Grainger and Schreiber, 2019). Nonetheless, readings from these monitors determine counties' attainment status. We therefore rely on data from these monitors as we are interested in assessing costs asso-

[6]The average distance between an AQS station and a NOAA station is 15 miles; we exclude AQS stations that do not have a corresponding NOAA station within 25 miles. Our regression analysis is limited to 48 AQS stations because nine stations did not have a proximate NOAA temperature monitor within 25 miles.

ciated with EPA-defined attainment, noting that monitor placement may cause biased estimates of a county's air quality.

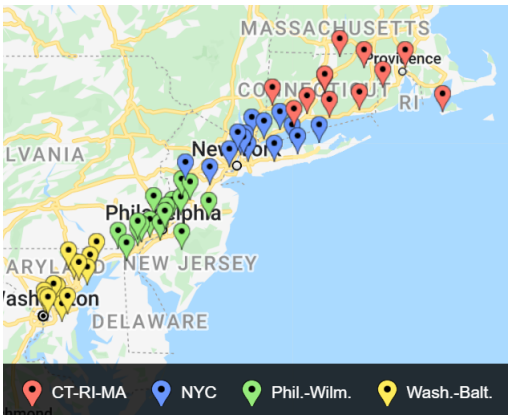
We limit our analysis to the "ozone season"—May through September—when 96 percent of all high-ozone events in our sample occurred. We focus on counties that failed to achieve attainment—those that did not reduce ozone concentrations in compliance with the NAAQS—in the coastal mid-Atlantic and northeastern United States. Ozone pollution has been a persistent problem in this highly populated region due in part to heavy industrial activity and the eastward transport of NOx and VOCs from other regions.[7]

Our region of focus represents a subset of the areas that have failed to achieve ozone standards. Our analysis of other regions, which is not reported in this paper, reveals a highly varied relationship between ozone concentration and local NOx emissions. We maintain a focus on one particular region as illustrative, noting that application of a differentiated emissions price requires estimating region-specific relationships between NOx emissions and ozone.

Figure 3.1 shows the geographic sub-groupings used in our analysis and the location of each AQS station within a sub-region. Stations are grouped geographically based approximately on their proximity to major metropolitan locations: Washington, DC and Baltimore, Maryland; Philadelphia, Pennsylvania and Wilmington, Delaware; New York City, New York; and stations north of the New York City area in Connecticut, Rhode Island, and Massachusetts.

[7]Cross-state pollution transport has been the basis for extensive litigation. For example, in February 2020 Attorneys General from Connecticut, Delaware, Massachusetts, New Jersey and New York sued the U.S. EPA for failure to address "harmful pollution blowing into our states," namely NOx and VOCs. See New Jersey vs. Wheeler, Complaint for Declaratory and Injunctive Relief, Civil Action No. 20-cv-1425, US District Court for the Southern District of New York, Filed Feb. 19, 2020.

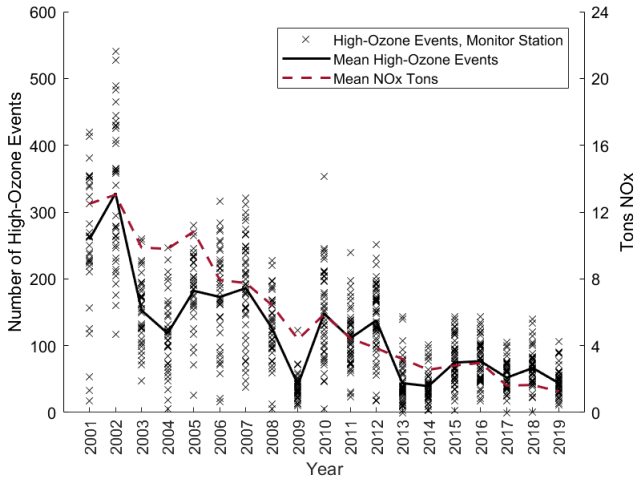
Figure 3.1: Geographic Groupings



Each marker represents an air monitoring station used in our analysis. We perform separate analyses for four sub-regions, which we define based on proximity to major metropolitan areas.

Figure 3.2 shows the mean number of high-ozone events (left axis) and emissions (right axis) between 2001 and 2019 across the monitors and power plants in the data. The individual x's indicate the number of events at individual stations, revealing quite a few stations that experience hundreds of high-ozone events annually during the first decade. The figure also illustrates a precipitous decline in NOx emissions and high-ozone events over time. However, since 2013 the number of high-ozone events has remained relatively flat. Studying this region therefore allows insight into how additional reductions in high-ozone events can be achieved in the presence of an electricity system that is already characterized by relatively low NOx emissions intensity. In other words, much of the initial gains from installing abatement technologies and replacing highly polluting coal plants with relatively cleaner-burning natural gas generators has already been achieved; further improvements in air quality may require alternative solutions to the uniform NOx caps currently in place.

Figure 3.2: NOx Emissions and Occurrence of High-Ozone Events, 2001-2019

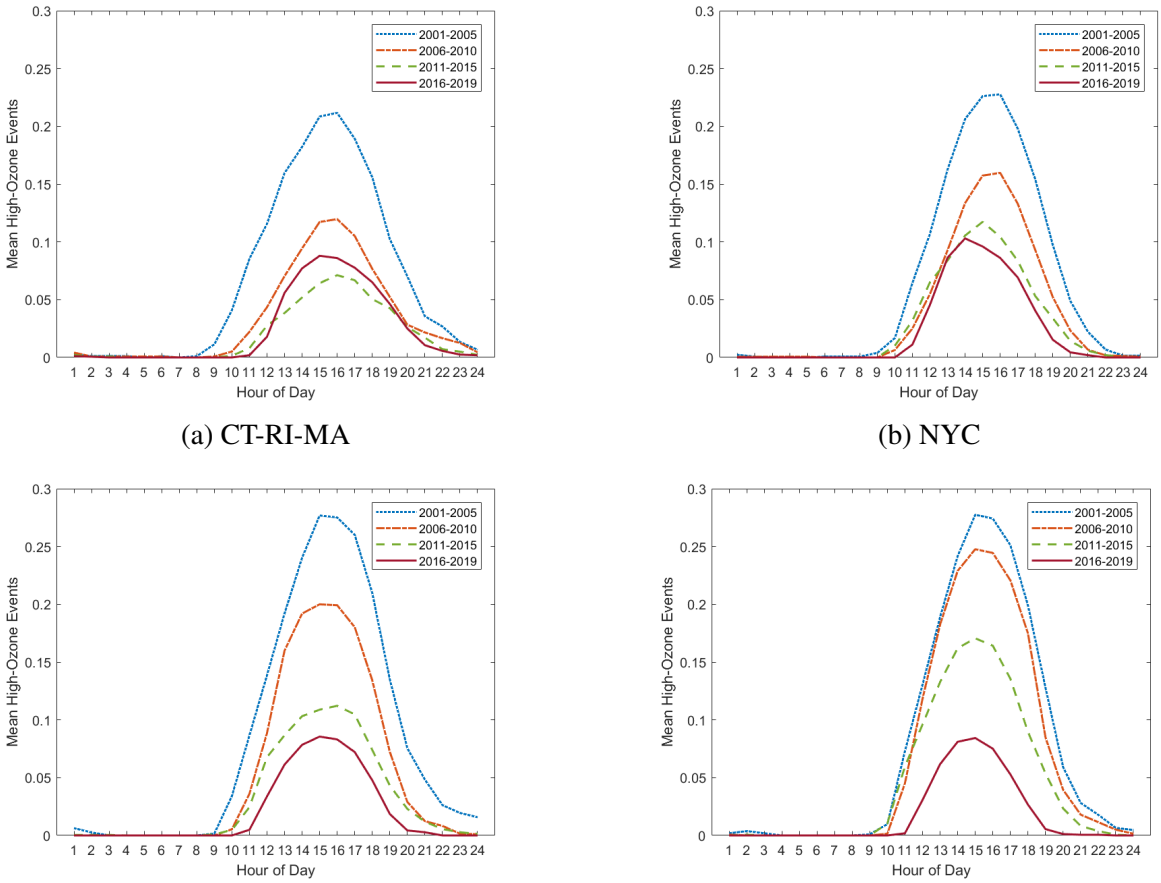


The left vertical axis shows the mean number of high-ozone events across monitors (solid line) and the number of high-ozone events at each monitor (x's). The right axis shows the average NOx emissions (tons) of plants located within 100 miles of a monitor.

Figure 3.3 shows changes in the average frequency of high-ozone events by hour-of-day for each sub-region over time. The y-axis measures the mean number of high-ozone events in each hour and sub-region. The probability of a high-ozone event exhibits a strong diurnal pattern that corresponds with sunlight and temperature; events occur most frequently in the mid-afternoon and dwindle in the evening hours, with virtually no high-ozone events occurring in the late night and early morning. Each line in each of the charts pertains to a different sub-period within the 19-year sample. Across all four sub-regions, the occurrence of high-ozone events declined significantly (roughly halved) between the earlier sub-periods (2001-2005, 2006-2010) and the later sub-periods (2011-2015, 2016-2019). This decline continued in the southern sub-regions (Philadelphia-Wilmington and Washington-Baltimore). In the northern sub-regions, however, the decline was arrested, and even slightly reversed in Connecticut-Rhode Island-Massachusetts.

Notwithstanding the overall downward trends, even in the most recent period, high-ozone events are frequent enough in all sub-regions to cause these areas to be in non-attainment for the ozone NAAQS.

Figure 3.3: Mean High-Ozone Events Over Time



(a) CT-RI-MA

(b) NYC

(c) Phil.-Wilm.

(d) Wash.-Balt.

Each panel shows the mean number of high-ozone events by hour of day for the sub-region indicated in the panel heading. Each line shows the mean across days in the indicated sub-period.

3.4 Estimating the Effects of NOx Emissions on Ozone Formation

The empirical analysis consists of (1) estimating the effect of NOx emissions on high-ozone events and (2) estimating marginal abatement costs. This section discusses the estimation

strategy and results pertaining to (1).

3.4.1 Estimation Strategy

As discussed in the introduction, hypothetically, if a policy reduces NOx emissions or even the average ozone level without affecting the number of high-ozone events, the policy would not help an area achieve attainment with respect to the NAAQS. For this reason, we aim to quantify the costs of policies that reduce the number of high-ozone events. This motivates our use of a high-ozone event as the dependent variable in the econometric analysis.[8]

Because we are interested in policies that affect power plant emissions, we estimate the effect of power plant emissions on high-ozone events. The estimating equation within each region is:

$$y_{i,t} = \beta_0 + \sum_{h=1}^{24} \beta_h \mathrm{H}_h \mathrm{N}_{i,t} + \sum_{j=2}^{4} \alpha_j \mathrm{T}_{i,j} + \gamma_i + \theta_t + \varepsilon_{i,t} \tag{3.1}$$

where $y_{i,t} \in \{0, 1\}$ is an indicator for whether ozone concentration exceeds 70 ppb in hour t at monitor i. The key independent variables are hour-of-day indicators (H_h) interacted with total NOx emissions from power plants within 100 miles of monitor i ($\mathrm{N}_{i,t}$). The equation includes fixed effects for quartiles of the temperature distribution ($\mathrm{T}_{i,j}$); γ_i is a station fixed effect; θ_t is a vector of fixed effects for year, month, day-of-week, and hour-of-day; and $\varepsilon_{i,t}$ is an error term.

We definie $\mathrm{N}_{i,t}$ as the total emissions across generators within 100 miles of the monitor

[8] We obtain similar results when defining the dependent variable as equal to one if the 8-hour moving average exceeds 70 ppb. As Appendix C.3 shows, defining the dependent variable this way yields an hourly pattern that lags that of our main results, but is otherwise similar. We choose to use an indicator variable based on contemporaneous ozone rather than the 8-hour average for our main analysis for ease of interpretation, and because reducing hourly high-ozone events remains of interest from a policy perspective regardless of whether there is an 8-hour high-ozone event.

for simplicity. The radius is consistent with typical atmospheric conditions that cause emissions greater than 100 miles from the monitor to be unlikely to affect ozone levels at the monitor within the same hour. However, this definition includes two implicit assumptions. First, emissions from all plants within this radius affect ozone levels equally and within one hour. Second, emissions from power plants outside this radius do not affect ozone levels at the monitor. We have tried alternative specifications that relax these assumptions, such as by including separate variables for emissions within 25 or 50 miles, or by including one or two-hour lags of emissions from generators located more than 100 miles away. We have also used results from the HYSPLIT model, which characterizes dispersion of air parcels in the atmosphere, to determine which generators' emissions may affect the hourly ozone reading in a particular hour.

In Equation (3.1), the coefficients on NOx emissions are identified by deviations of emissions from the corresponding hour relative to other days, and after controlling for mean emissions at the monitor and during the same hour-of-day, day-of-week, month, and year. For example, if a generator operates more on one day than it did at the same hour on the same day during the previous week, the higher emissions help identify the coefficient for that hour.

Equation (3.1) allows the coefficient on NOx emissions to vary by hour of day, and the estimated effect of emissions on high-ozone events can differ freely over the day. For example, the data determine whether we estimate a larger marginal effect for daytime versus nighttime emissions. The temperature variables control for the fact that high-ozone events are more common when the temperature is high, conditional on emissions. The monitor fixed effects control for the average high-ozone event probability, emissions, and temperature at the monitor. The temporal fixed effects in θ_t control for annual, monthly, day-of-week, and diurnal patterns of ozone levels, emissions, and temperature.

We make two remarks about Equation (3.1). First, NOx emissions have a linear effect on the probability of a high-ozone event. As robustness checks, we can relax the linearity assumption by including higher-order polynomial terms in NOx emissions and adding interactions between emissions and meteorological variables such as temperature.

Second, the equation does not include explicit controls for other sources of NOx emissions or VOCs or for emissions from beyond the 100-mile radius. The monitor fixed effects and θ_t control for the average levels of these emissions at the monitor as well as the average emissions for each year, month, day-of-week, and hour-of-day. For example, θ_t controls for diurnal patterns in vehicle emissions of NOx and VOCs.

However, these variables likely do not control perfectly for other emissions sources. For example, suppose an industrial facility burns coal and consumes electricity. If the facility operates at an unusually high level on one day, its direct emissions of NOx will be higher than usual due to additional coal consumption. Moreover, its electricity consumption causes power plants to burn more fuel and emit more NOx than they would normally. If those power plants and the industrial facility are located near the monitor, $\mathrm{N}_{i,t}$ in Equation (3.1) would be correlated with the omitted industrial NOx and VOC emissions, causing inconsistent estimates of the NOx coefficients.

An additional concern is emissions from upwind, non-local sources. That several states have filed lawsuits against the EPA because out-of-state upwind emissions contribute to their non-compliance with ozone standards attests to the importance of NOx transport from outside of the 100-mile radius and one-hour window that we use to compute $\mathrm{N}_{i,t}$. Because of diurnal patterns in electricity demand and other factors, out-of-state emissions are also likely to be correlated with contemporaneous emissions from proximate sources.

To address the potential endogeneity of the NOx emissions, we instrument for $\mathrm{N}_{i,t}$. Fol-

lowing Johnsen, Lariviere, and Wolff, 2019, the instrument is the emissions predicted from a computational model of generator operation. In a standard economic dispatch model, generators in the system are dispatched to minimize the costs of meeting electricity demand. We use the model from Linn and McCormack, 2019, which extends the standard dispatch model by accounting for startup and adjustment costs; they show that the model outperforms a standard dispatch model by a substantial margin.[9]

The exogenous variables in the generation model are fuel prices and the total hourly fossil fuel-fired generation for the interconnection. Fuel prices are the monthly delivered prices computed from EIA Form 923. The total generation is the average for the corresponding hour across the years 2005 through 2019. For example, for 3pm on March 31, we compute the mean total fossil fuel-fired generation across the 15 observations of total fossil fuel-fired generation at 3pm on March 31 between 2005 and 2019. Consequently, total generation in a particular hour reflects typical behavior rather than events specific to that year. This reduces the potential correlation between the instrument and omitted economic activity, such as the reduction in industrial activity during the 2008-2009 economic recession.

The instrument isolates variation in unit-level emissions driven by time varying fuel prices and demand and cross-sectional variation in generator fuel type and efficiency. For example, natural gas prices were higher in 2008 than in 2009. In 2009, some coal-fired generators were more costly to operate than some natural gas-fired generators, which was not true in 2008. Therefore, the operational model predicts higher generation levels and emissions from some coal-fired

[9]In a standard dispatch model, each generator operates at full capacity if its marginal costs do not exceed the electricity price and does not operate if costs exceed the price. However, it is costly for thermal generators to start up after not operating for several hours or to adjust their output quickly. Because of these costs, many generators operate at low levels of generation even if their (static) marginal costs exceed the price. 2019 use a reduced-form strategy to approximate these costs and reproduce this behavior.

power plants in 2009 than in 2008. Because coal-fired generators typically have higher NOx emissions rates than gas-fired generators, the model predicts lower NOx emissions in 2009 than 2008 because of the decrease in natural gas prices.

Moreover, the model predicts a larger emissions decline between 2008 and 2009 during hours when aggregate generation is low. This is because when aggregate generation is high, the coal-fired generators operate even if they are more costly because those generators are needed to meet aggregate generation. Thus, the model yields variation in emissions across years and across hours within a year.

First stage results (see Appendix C.2) show that the emissions instrument is a strong predictor of the observed emissions. The sources of variation of the inputs to the operational model help support the exclusion restriction that the emissions instrument is uncorrelated with the error term in Equation (3.1). Returning to the example of the industrial facility above, because we use typical hourly generation as an input, variation of the facility's electricity consumption across years is uncorrelated with the aggregate generation input to the model. The underlying assumption is that variation in fuel prices across years yields emissions levels that are uncorrelated with within-year shocks to emissions from other sources. This assumption is the same as in Johnsen, Lariviere, and Wolff, 2019, where the main difference is that we are using a different operational model to predict each generator's emissions.

3.4.2 Estimation Sample Summary Statistics

Table 3.1 provides summary statistics for the full sample of data that includes each ozone season (May through September) from 2001-2019 for 57 AQS stations. Each panel shows statis-

tics for the sub-region indicated in the panel heading.

Table 3.1: Ozone Regression Summary Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
CT-RI-MA					
Ozone Parts Per Billion	596,734	33.15	17.22	0.00	158.00
Ozone > 70 PPB	596,734	0.03	0.16	0.00	1.00
NOx Tons	596,734	3.01	3.38	0.02	42.63
Simulated NOx Tons	596,734	2.88	3.98	0.00	40.63
Temperature (degrees Fahrenheit)	596,734	68.36	9.72	3.02	102.02
NYC					
Ozone Parts Per Billion	361,650	31.56	18.41	0.00	167.00
Ozone > 70 PPB	361,650	0.03	0.17	0.00	1.00
NOx Tons	361,650	5.07	5.26	0.25	55.23
Simulated NOx Tons	361,650	5.15	5.80	0.00	52.68
Temperature (degrees Fahrenheit)	361,650	70.60	9.41	30.92	107.96
Phil.-Wilm.					
Ozone Parts Per Billion	914,284	33.05	18.69	0.00	150.00
Ozone > 70 PPB	914,284	0.03	0.18	0.00	1.00
NOx Tons	914,284	6.59	5.42	0.44	57.06
Simulated NOx Tons	914,284	7.33	5.13	0.21	51.55
Temperature (degrees Fahrenheit)	914,284	72.76	9.65	35.60	105.08
Wash.-Balt.					
Ozone Parts Per Billion	554,407	32.79	19.43	0.00	156.00
Ozone > 70 PPB	554,407	0.03	0.18	0.00	1.00
NOx Tons	554,407	5.82	4.75	0.18	45.43
Simulated NOx Tons	554,407	7.25	4.18	0.33	39.04
Temperature (degrees Fahrenheit)	554,407	73.18	9.75	33.98	107.06

The table reports summary statistics of hourly observations for the sub-region indicated in the panel heading. Ozone $>$ 70 ppb is an indicator equal to one if the hourly ozone concentration exceeds 70 ppb. NOx emissions are computed for power plants located within 100 miles of the monitor. Simulated NOx tons is the emissions predicted by the simulation model (see text for details).

The average ozone concentration was about 32-33 ppb across sub-regions. The ozone level exceeded 70 ppb in approximately three percent of all hours in the sample. The bottom two rows of each panel show the variation of the simulated emissions and temperature, where the simulated emissions are the instrument for $N_{i,t}$ in Equation (3.1). Importantly for identification of the first

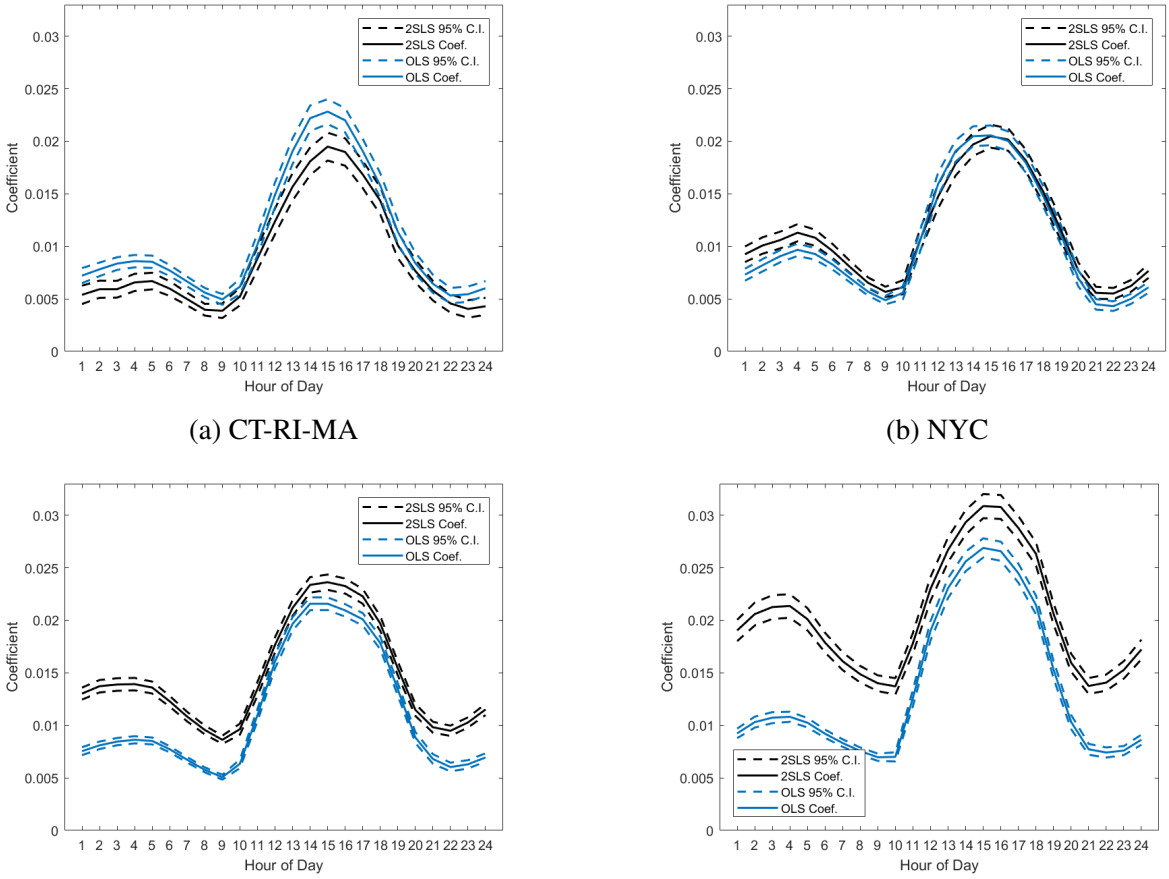
stage, the simulated emissions have similar variance and range to the observed emissions.[10]

3.4.3 Estimation Results

Figure 3.4 plots the coefficients on the interactions of NOx emissions and indicator variables for each hour of the day. The coefficients represent the incremental change in the probability of a high-ozone event caused by a one ton emissions increase across generators located within 100 miles of the AQS monitoring site. The blue lines in Figure 3.4 show the results from OLS regressions, and the black lines show results from 2SLS regressions. For reference, Appendix C.2 presents the point estimates and standard errors of these interactions. Dashed lines indicate 95 percent confidence intervals.

[10]The mean simulated emissions differ from the observed emissions by a few tons per hour. This difference is explained by our use of mean total fossil fuel-fired generation across 2005-2019 rather than the observed hourly generation each year. If we use the observed total generation rather than the mean, the operational model accurately predicts mean emissions, as reported in Linn and McCormack, 2019.

Figure 3.4: OLS and 2SLS Regression Results, 2001-2019



(a) CT-RI-MA

(b) NYC

(c) Phil.-Wilm.

(d) Wash.-Balt.

For the sub-region indicated in the panel heading, the figure plots the estimated coefficients on the interactions between hourly NOx emissions and a set of hour-of-day fixed effects (see Equation (3.1)). Dashed lines show the 95 percent confidence intervals, where standard errors are robust to heteroskedasticity. The blue line reports results from estimating the equation by ordinary least squares (OLS), and the black line reports the results from instrumenting emissions with predicted emissions (see text).

The figure illustrates a distinct diurnal pattern that matches the diurnal ozone levels in Figure 3.3. An additional ton of NOx emissions is statistically significantly more likely to contribute to a high-ozone event in the mid-afternoon than at other times of day. The pattern is intuitive because temperatures are typically highest and sunlight exposure greatest in mid-afternoon, both of which are conditions conducive to ozone formation. Overall, the temporal variation of the coefficients is greater than the spatial variation across sub-regions.

Disparity between the OLS and 2SLS coefficients would indicate that the OLS estimates could be biased by omitted variables. In principle, the bias could be negative or positive, depending on whether power plant NOx emissions are negatively or positively correlated with omitted variables. In the top two panels, the OLS estimates are larger than the 2SLS estimates, which indicates a positive correlation between emissions and omitted variables. In contrast, the bottom two panels indicate that the OLS estimates are downward biased.

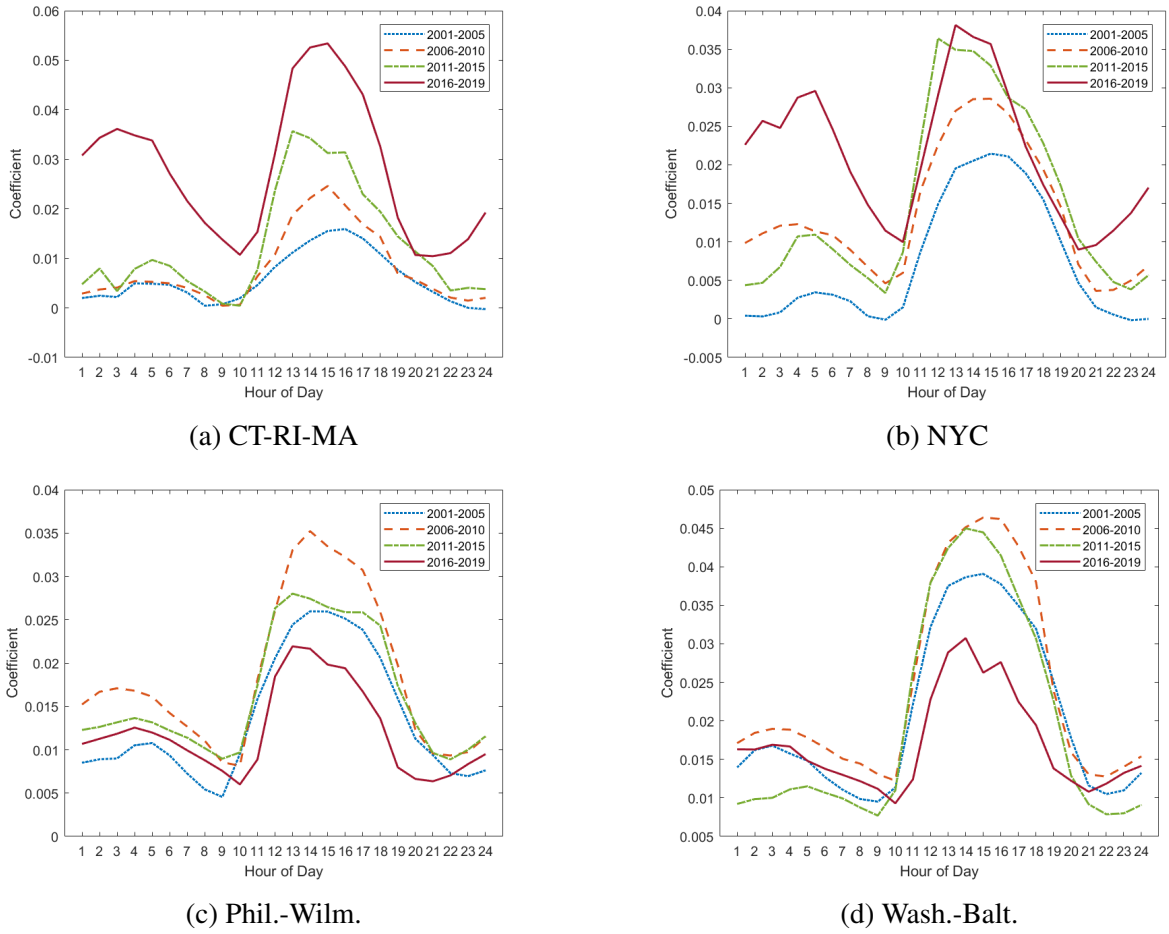
The variation of the marginal effects of emissions on high-ozone events implies that a differentiated emissions price may be more cost effective at reducing high-ozone events than a uniform emissions price. The cost-effective emissions price should be higher during mid- and late afternoon than at other times of day.

As we discussed in Section 3.2, the advantage of a differentiated emissions price depends on the accuracy with which one can predict temporal variation in the effect of emissions on high-ozone events. One way to assess this accuracy is to suppose that the EPA predicts the temporal variation in year t using data prior to year t.

Figure 3.5 demonstrates how the emissions coefficients change across time periods in each sub-region. The precipitous decline in NOx emissions from 2001-2019 in all sub-regions could cause the coefficients to become larger in magnitude if the lower NOx emissions overall increase the likelihood that ozone formation is NOx-limited. However, we do not observe a monotonic change in the emissions coefficients. In the northern sub-regions, NOx sensitivity has increased over time; in the southern regions, it has declined in recent years. These results underscore the importance of recognizing the changing atmospheric conditions which can also vary widely across regions. That the coefficients vary across time periods implies that if EPA uses data from one time period to predict the effect of emissions in the following period, the prediction is likely

to be inaccurate. Consequently, a differentiated emissions price may be less cost effective than a uniform emissions price.

Figure 3.5: 2SLS Regression Results, Year-Groupings



(a) CT-RI-MA

(b) NYC

(c) Phil.-Wilm.

(d) Wash.-Balt.

The figure shows hourly coefficients for each sub-region. Each line plots coefficients estimated using data from the years shown in the legend. All coefficients are estimated by instrumenting for NOx emissions analogously to Figure 3.4.

3.5 Estimating Marginal Abatement Costs

The policies considered in the counterfactuals impose emissions prices on all generators. To compare the costs across counterfactuals, we need estimates of marginal abatement costs, allowing for the possibility that these costs vary by hour-of-day or sub-region. This section

discusses the estimation strategy and results pertaining to our estimation of short-run marginal abatement costs.

3.5.1 Estimation Strategy

For a generation unit facing an emissions price, we define the unit's marginal abatement costs as the cost of reducing emissions by one ton of NOx relative to a baseline in which the emissions price equals zero. We expect marginal abatement costs to increase with the level of abatement, measured in NOx tons reduced.

In the short run, firms can reduce emissions by decreasing output from generators or reducing the emissions rates of their generators. For units with post-combustion controls such as selective catalytic reduction, firms can control emissions by deciding whether and how to operate these controls. For other units, firms can adjust the operation to reduce combustion temperatures, which helps reduce emissions rates. Linn, 2008 shows that firms adjust emissions rates by up to 10 percent in response to monthly variation in emissions allowance prices. In the long run, firms can reduce emissions by adopting technologies that decrease emissions rates or by investing in generators that use lower-emissions fuels, such as replacing coal-fired units with gas-fired units or wind or solar-powered generators.

Given these short- and long-run responses to the permit price, each firm has a short- and long-run marginal abatement cost curve, which represents the marginal costs to the firm of reducing emissions below the baseline level of emissions. These costs include reductions in profits from producing less electricity, increases in fuel costs from operating less efficiently, or increases in operating costs of the emissions controls.

Assuming the allowance market is perfectly competitive, each unit reduces emissions until its marginal abatement costs equal the allowance price. Consequently, the aggregate marginal abatement cost curve for a region is the horizontal sum of the marginal abatement cost curve for each firm in that region.

We assume the aggregate marginal abatement cost curve is linear in abatement. Appendix C.1 shows that under this assumption, the slope of the marginal abatement cost curve is inversely proportional to the derivative of the average emissions rate with respect to the allowance price. This relationship motivates the estimating equation in which we regress the emissions rate on permit price:

$$\mathrm{E}_{i,t} = \delta_0 + \sum_{h=1}^{24} \delta_h \mathrm{H}_h p_t + \eta_i + \Xi_t + \nu_{i,t} \tag{3.2}$$

where $\mathrm{E}_{i,t}$ is the average hourly NOx emissions per MWh generated across generators located within 100 miles of station i in period t; H_h is an indicator for hour-of-day; p_t is the daily NOx allowance permit price; η_i is a station-specific fixed effect; Ξ_t is a vector of fixed effects for year, month, day-of-week, and hour-of-day; and $\nu_{i,t}$ is an error term. We perform a separate regression for each sub-region. The Connecticut-Rhode Island-Massachusetts region is excluded from this analysis because these states are not subject to CSAPR.

The key coefficients of interest are the interactions of the hour fixed effects with the allowance price. We expect the coefficients to be negative because a higher permit price incentivizes abatement that reduces the average emissions rate.

By design, three features of Equation (3.2) are parallel to Equation (3.1). First, the two equations have the same unit of observation, which is an hour by monitor, and the sample includes

only observations from May through September. Second, the dependent variable, $E_{i,t}$, is the average emissions rate across the same set of units that are used to compute NOx emissions $N_{i,t}$ in Equation (3.1). Consequently, emissions from a common set of generators are used to identify the effect of emissions on ozone levels and marginal abatement costs. Third, the two equations include the same set of time fixed effects to control for seasonal and diurnal patterns of generator utilization and emissions.[11]

The interactions of the hour fixed effects with the daily allowance price are the key coefficients of interest, and the year fixed effects play an important role in their identification. Because of the fixed effects, the coefficients are identified by within-year permit price variation. As discussed above, changes in generator utilization, boiler operation, and emissions controls can affect the dependent variable; all of these behaviors help identify the coefficients. Because of the year fixed effects, we interpret the coefficients as representing short-run marginal abatement cost, that is, abatement that occurs without the adoption of pollution abatement equipment. This interpretation of the coefficients is consistent with the construction of the counterfactuals, as we explain in the next section.

3.5.2 Estimation Sample Summary Statistics

Table 3.2 shows summary statistics for the dependent variable in Equation (3.2), which is the hourly average emissions rate across generator units located within 100 miles of the monitor. Emissions rates are greater in the southern sub-regions of the sample, which reflects a generation mix that is composed of more NOx-intensive (e.g. coal-burning) technologies.

[11] According to Equation (3.2), the emissions rate is endogenous to the allowance price. Because Equation (3.1) does not include the allowance price, this endogeneity is an additional argument for instrumenting for emissions. By construction, the instrument we use is uncorrelated with the allowance price.

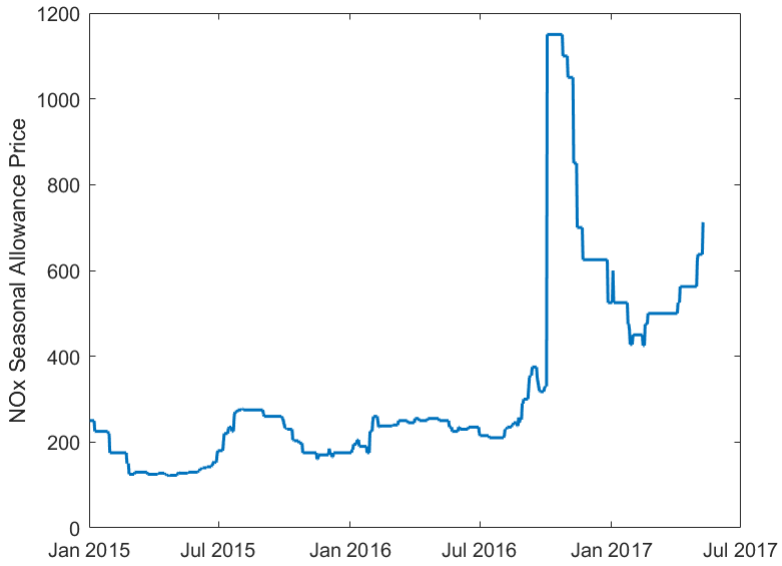
Table 3.2: NOx Emissions Rates (lbs. per MWh) by Sub-Region

Sub-Region	Obs	Mean	Std. dev.	Min	Max
NYC	98,970	0.33	0.11	0.11	1.15
Phil.-Wilm.	123,268	0.49	0.15	0.17	1.28
Wash.-Balt.	82,519	0.72	0.22	0.22	1.93

Observations are by hour and monitor.

Figure 3.6 shows the daily price of an allowance of one ton of NOx emissions from Jan. 2015 to May 2017. These permits were traded in compliance with EPA's CSAPR program, and EPA allows a limited amount of allowance banking. EPA allocates permits each year, and each permit is assigned a vintage. We use the contemporaneous vintage in the figure, such that prices for each calendar year reflect that year's vintage. Although not shown in the figure, we observe that when multiple vintages are traded at the same time, the prices for the allowances are highly correlated with one another, although earlier vintages often sell at a discount because of banking constraints.

Figure 3.6: Seasonal NOx Allowance Price, Jan. 2015 - May 2017

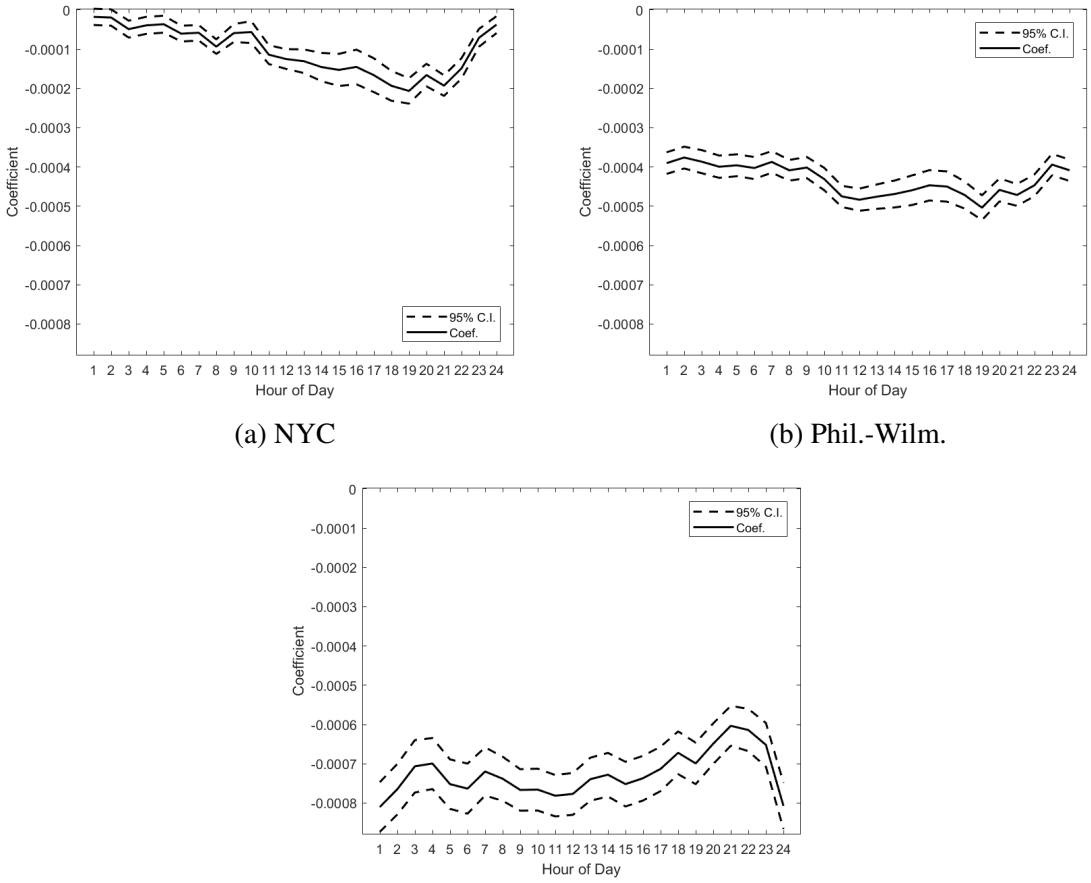


The price of a seasonal NOx allowance under CSAPR varies depending on expected future demand, generation mix, and NOx allowance budgets set by the EPA. Budgets associated with CSAPR were tightened in 2017 leading to higher allowance prices.

3.5.3 Estimation Results

Figure 3.7 illustrates coefficient estimates by hour-of-day (δ_h from Equation (3.2)) for the three sub-regions subject to CSAPR. The coefficients are identified by within-year deviations from hour-of-day and monthly averages of permit prices and emissions rates. We interpret the coefficients as the negative reciprocal of the slope of the short-run marginal abatement cost curves, and we estimate separate coefficients for each sub-region.

Figure 3.7: MAC Regression Results



(a) NYC

(b) Phil.-Wilm.

(c) Wash.-Balt.

Each panel reports the coefficient estimates (solid lines) and 95 percent confidence intervals (dashed lines) from estimating Equation (3.2) for the indicated sub-region. Standard errors are robust to heteroskedasticity.

For all three sub-regions, the interaction coefficients are statistically significant at the 1 percent level in nearly all cases, and they are negative as expected. Because the permit price is measured in dollars per ton of emissions and the dependent variable is measured in pounds of NOx per MWh of generation, a coefficient of -0.0004 (which is a typical estimate for the Philadelphia-Wilmington sub-region) implies that a decrease of the allowance price of $100 per ton reduces the emissions rate by 0.04 pounds of NOx per MWh, which is about eight percent of the mean for this sub-region.

Recall that we interpret the negative of the reciprocal of the price-hour coefficients as the slope of the marginal abatement cost curves. The magnitudes of the coefficients decrease moving from north to south across the sub-regions. In other words, we find that marginal abatement cost curves are less steep in the southern sub-regions. This result may reflect greater flexibility in short-term operational changes or re-dispatch away from NOx-intensive generation in the southern sub-regions. In contrast, generators in the northern sub-region are already relatively low-emitting, and reducing NOx emissions further is more costly.

Although we do not observe as strong a diurnal pattern as in Figure 3.4, for the New York City and Philadelphia-Wilmington sub-regions, the marginal abatement cost curve appears to be less steep in the early evening than during other hours.

3.6 Policy Simulations

3.6.1 Description of Counterfactual Policy Scenarios

We simulate hypothetical policy scenarios to evaluate whether temporally or spatially differentiated emissions prices would be more cost effective than uniform prices at reducing high-ozone events between 2016 and 2019. This time period corresponds to the most recent period used for estimating parameters in Equation (3.1), and it corresponds roughly to the period used to estimate Equation (3.2).

Across the scenarios, we hold constant the average change in the probability of high-ozone events, which effectively fixes attainment status with respect to the ozone NAAQS.[12] Doing so

[12]This is not the same as holding fixed environmental damages, because two policies could cause the same probability of a high-ozone event and cause different ozone levels. We construct the scenarios to achieve the same probability of a high-ozone event because the aim of CSAPR is to facilitate attainment rather than to maximize social welfare.

allows us to compare the costs of policies that achieve the same reduction in high-ozone events. This approach is useful for analyzing costs of achieving attainment regardless of whether attainment goals are socially optimal.

The differentiated emissions price varies by hour-of-day and sub-region in proportion to the marginal effect of emissions on the probability of a high-ozone event, which is estimated in Equation (3.1). Because the regulator must commit to the differentiated price at the beginning of the year, we construct two types of differentiated prices. The first we refer to as the *ex ante* price; this price is proportional to the marginal effects estimated from the preceding (*ex ante*) time period (2011-2015).

For comparison with the *ex ante* price, we define an *ex post* differentiated price that assumes perfect foresight. This emissions price varies according to NOx emissions coefficients in Equation (3.1) that are estimated using *ex post* data (spanning the 2016-2019 ozone seasons). Because the NOx-ozone relationship changes from one period to the next (see Figure 3.5), the *ex post* parameters produce emissions prices that are more accurately targeted to reduce high-ozone events.[13]

Formally, the *ex ante* emissions price is defined as:

$$\tau_{i,t}^{ea} \equiv \lambda_i^{ea} \beta_{i,t}^{ea} \tag{3.3}$$

[13]Pooling observations across several years masks year-to-year variation in NOx sensitivity. To obtain more accurate estimates for the purpose of informing a policy, the policy maker may wish to allow coefficients to vary by year or an even shorter time horizon, or use detailed weather forecasts and air transport modeling.

and the *ex post* emissions price is defined as:

$$\tau_{i,t}^{ep} \equiv \lambda_i^{ep} \beta_{i,t}^{ep} \tag{3.4}$$

where *ex ante* parameters β_i^{ea} are estimated using Equation (3.1) and 2011-2015 data, and *ex post* parameters β_i^{ep} are estimated using Equation (3.1) and 2016-2019 data; and λ_i^{ea} and λ_i^{ep} are scale parameters.[14]

The third policy is a uniform emissions price, $\bar{\tau}$. To calibrate the differentiated prices, we solve for scale parameters λ_i^{ea} and λ_i^{ep} that equate reductions in expected high-ozone events in each region i:

$$\sum_t (\lambda_i^{ea} \beta_{i,t}^{ea})[\delta_{i,t} q_{i,t} \beta_{i,t}^{ep}] = \sum_t (\lambda_i^{ep} \beta_{i,t}^{ep})[\delta_{i,t} q_{i,t} \beta_{i,t}^{ep}] = \sum_t \bar{\tau}[\delta_{i,t} q_{i,t} \beta_{i,t}^{ep}] \tag{3.5}$$

where $q_{i,t}$ is MWh generation in region i and period t, and $\delta_{i,t} q_{i,t}$ is the change in NOx emissions resulting from a one dollar increase in the NOx price in period t. The terms inside the square brackets are the reduction in the probability of a high-ozone event caused by increasing the NOx emissions price by one dollar. Because we assume linear marginal abatement cost curves and a linear relationship between NOx emissions and high-ozone events, the amount by which the probability of a high-ozone event is reduced in period t scales linearly with the emissions price. We set $\bar{\tau} = \$100$. Regression output from the estimation of β_i^{ea} and β_i^{ep} is presented in Appendix C.2.

Next we explain how we compute total abatement costs of each policy. Appendix C.1

[14] t indexes each hour in our sample. For ease of exposition we use the t subscript noting that coefficients vary only by hour-of-day h.

shows that the slope of the marginal abatement cost curve is inversely proportional to the allowance price coefficient from Equation (3.2). Total abatement cost is the integral from 0 to a of marginal abatement costs:

$$\begin{aligned} TC &= \int_{x=0}^{a} -\frac{x}{\delta_1} dx \\ &= -\frac{a^2}{2\delta_1 q} \end{aligned} \tag{3.6}$$

Total costs are proportional to the square of abatement and inversely proportional to generation and the estimated allowance price coefficient in Equation (3.2).

We can express total costs as a function of the allowance price rather than abatement by exploiting the fact that firms abate emissions until marginal abatement costs equal the allowance price. The amount of abatement achieved from price τ is equal to:

$$a(\tau) = -\tau \delta_1 q \tag{3.7}$$

Plugging Equation (3.7) into Equation (3.6) gives the total abatement costs associated with price τ:

$$TC = -\frac{1}{2}\tau^2 \delta q \tag{3.8}$$

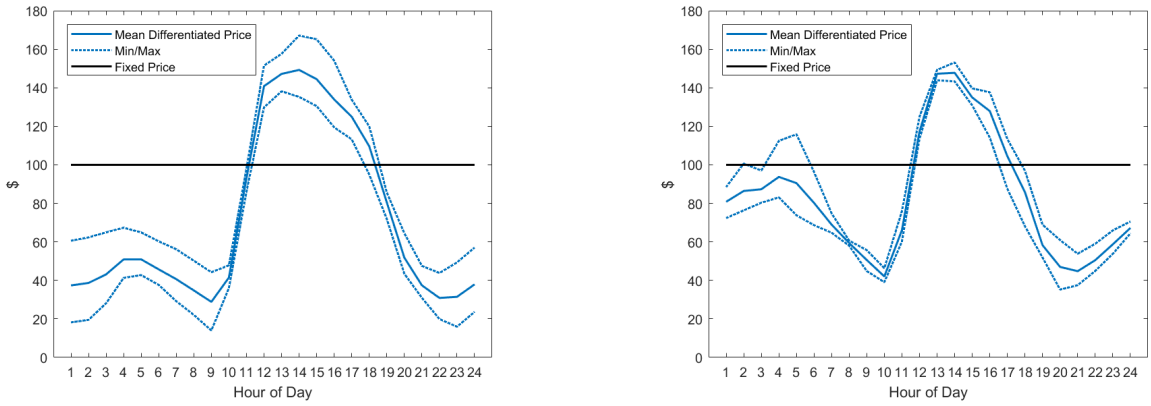
This equation allows us to compute total costs for each hour and site given the assumed allowance price, estimated δ_1, and generation level. Expected reductions in high-ozone events for each hour and site associated with policy τ equal $\tau\delta q$. Cost effectiveness is defined as total costs (summed across hours and sites) divided by the expected reductions in high-ozone events (summed across hours and sites).

3.6.2 Results

For each of the three sub-regions subject to CSAPR seasonal prices, we consider fixed or differentiated prices for the monitor in the sub-region with the highest frequency of high-ozone events; all other monitors in the sub-regions are excluded from these simulations. Note that a generation unit may lie within 100 miles of multiple air quality monitors. Therefore, focusing on a single monitor in each sub-region avoids double-counting abatement costs of the generators that lie within 100 miles of multiple monitors.

Figure 3.8 compares the hour-of-day differentiated emissions prices with the fixed price of $100. The solid line shows the mean differentiated prices by hour of day and sub-region and thee dashed lines show the minimum and maximum prices across days in the 4-year time period of analysis. Panel (a) shows the ex ante prices and panel (b) shows the ex post prices. In both cases, the differentiated prices are higher in the mid-day hours than during other hours, which reflects the fact that NOx emissions during the mid-day hours have larger effects on high-ozone events. On average, the differentiated prices are lower than the fixed price.

Figure 3.8: Mean Differentiated Price by Hour-of-Day



(a) *Ex Ante* Emissions Price

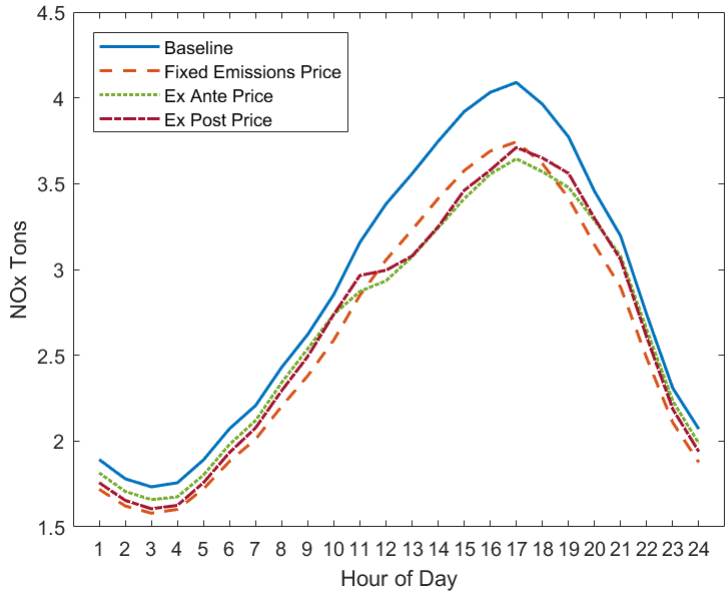
(b) *Ex Post* Emissions Price

The solid black line is the fixed emissions price of $100 per ton and the solid blue lines are the differentiated prices. Panel (a) reports differentiated prices using parameter estimates from 2011-2015 (*ex ante*), and panel (b) reports prices using estimates from 2016-2019 (*ex post*). Min/Max are the minimum and maximum prices across days.

The *ex ante* price shown in panel (a) reflects priors about ozone formation that are outdated; as shown in panel (b), updating those priors leads to a more accurate *ex post* price that is substantially different from the *ex ante* price. The differences between the differentiated prices in the two panels indicates that, in principle, the *ex ante* could be less cost effective than the uniform price.

Figure 3.9 compares average hourly NOx emissions in the baseline (zero emissions price) with those under the fixed and differentiated emissions prices. The fixed emissions price causes a relatively uniform reduction in NOx emissions throughout all hours of the day. This is consistent with the fact that the slopes of the marginal abatement cost curves vary relatively little over the day and that the uniform price does not vary across hours (by assumption).

Figure 3.9: Average Hourly NOx Emissions Under Fixed vs. Differentiated NOx Price



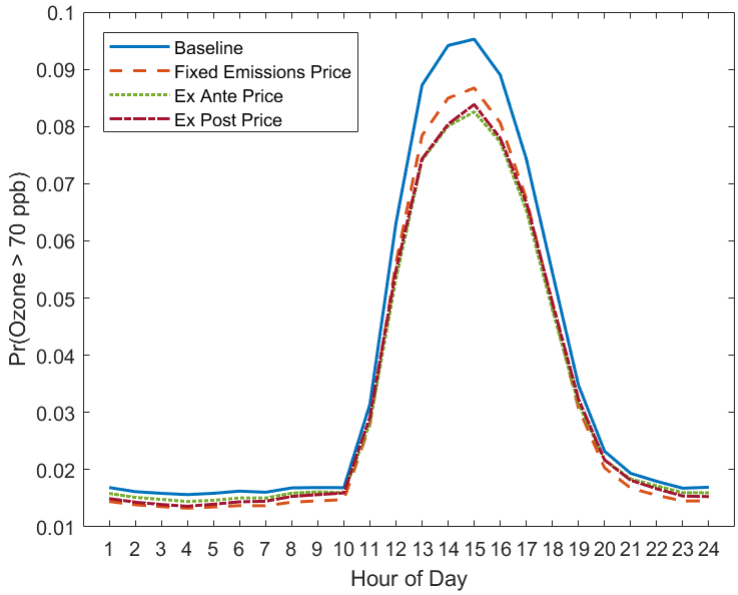
The blue line is the average hourly emissions in the baseline scenario, which reflects an allowance price of zero. The dashed orange curve is the average hourly emissions under the fixed emissions price scenario, and the two dotted curves are the average hourly emissions under the *ex ante* and *ex post* price scenarios.

By design, the differentiated emissions prices reduce emissions more during the middle of the day, when NOx is most likely to contribute to high-ozone events. The changes in emissions under the *ex ante* and *ex post* prices are almost identical to one another, despite the fact that the prices differ from one another (see Figure 3.8).

Figure 3.10 compares the resulting average decline in the probability of a high-ozone event across scenarios. The uniform emissions price reduces the probability during all hours, and by more during the middle of the day than during other hours. The differentiated emissions prices cause greater reductions in high-ozone events during the middle of the day, when these events are most common, than the uniform price. Analogously to Figure 3.9, the *ex ante* and *ex post* cause

nearly identical changes in the probability of a high-ozone event.

Figure 3.10: Average Probability of High-Ozone Event, Fixed vs. Varied NOx Price



The figure shows the average probability of high-ozone events for the baseline and three policy scenarios. Probalities are computed using the allowance prices in Figure 3.8 and the estimated marginal effects of emissions on ozone levels from Equation (3.1).

During mid-day hours, the differentiated prices cause larger reductions in high-ozone event probabilities than the uniform price. The situation is reversed in other hours, when the uniform price causes larger reductions in probabilities. These results imply that overall, the differentiated prices cause less aggregate abatement than the fixed price because the differentiated prices cause greater emissions reductions when those emissions reductions are most effective at reducing high-ozone events.

Table 3.3 compares abatement costs across the uniform and differentiated emissions prices. Policies are calibrated to achieve the same reduction in the expected number of high-ozone events per season. Panel (a) shows that the differentiated prices cause less NOx abatement even though

they achieve the same probability of high-ozone events. By construction, costs per ton of NOx emissions reduced are equivalent under the fixed versus *ex post* price, as shown in panel (b). The *ex ante* price exhibits slightly greater costs per ton of NOx abatement due to inaccuracy of the *ex ante* parameter estimates. Nevertheless, both the *ex ante* and *ex post* prices are more cost effective than the fixed price, with the differentiated prices reducing costs by 4-18 percent.

Table 3.3: Cost Effectiveness Summary

(a)

	Δ NOx Tons			Δ High-Ozone Events		
Region	Fixed	Varied - *Ex Ante*	Varied - *Ex Post*	Fixed	Varied - *Ex Ante*	Varied - *Ex Post*
NYC	280.5	229.1	230.0	5.9 [5.8, 5.9]	5.9 [5.8, 5.9]	5.9 [5.8, 5.9]
Phil.-Wilm.	1,238.5	1,066.8	1,045.5	15.4 [15.4, 15.5]	15.4 [15.4, 15.5]	15.4 [15.4, 15.5]
Wash.-Balt.	1,344.4	1,105.3	1,182.5	23.7 [23.7, 23.8]	23.7 [23.7, 23.8]	23.7 [23.7, 23.8]

(b)

	Cost per NOx Ton Reduced			Cost per Expected High-Ozone Event Avoided			Pct. Savings from Varied Tax	
Region	Fixed	Varied - *Ex Ante*	Varied - *Ex Post*	Fixed	Varied - *Ex Ante*	Varied - *Ex Post*	*Ex Ante*	*Ex Post*
NYC	$50	$55	$50	$2,387 [2375.4, 2399.2]	$2,126 [2115.3, 2137]	$1,958 [1948.9, 1966.9]	11% [0.107, 0.111]	18% [0.178, 0.182]
Phil.-Wilm.	$50	$50	$50	$4,015 [4007, 4023.1]	$3,484 [3475.5, 3491.5]	$3,389 [3381.8, 3397.1]	13% [0.132, 0.133]	16% [0.155, 0.156]
Wash.-Balt.	$50	$58	$50	$2,832 [2826.1, 2838.4]	$2,716 [2708.5, 2723.9]	$2,491 [2485.1, 2497.4]	4% [0.04, 0.042]	12% [0.12, 0.121]

Changes in NOx tons, changes in high-ozone events, and costs are expressed as averages per ozone season from 2016-2019. The differentiated prices are designed to reduce NOx emissions in proportion to their propensity to contribute to high-ozone events. Fewer reductions are therefore required to achieve the same outcome as a fixed price, as shown in panel (a), leading to lower overall costs as shown in panel (b). Bootstrapped standard errors using 1,000 draws of NOx-ozone coefficients from Equation (3.1) are shown in brackets.

These results differ from Fowlie and Muller, 2019 in that we find that estimates of the emissions-ozone relationship are sufficiently accurate that the differentiated emissions price yields cost savings. The difference likely arises from the fact that we consider temporally varying prices and we estimate large temporal variation in the marginal effect of NOx emissions on high-ozone events. However, similar to their analysis, our findings underscore the importance of accurately

estimating this relationship for each sub-region.

3.7 Conclusion

High-ozone events occur when ground-level ozone concentrations exceed the threshold the EPA uses to assess attainment with Clean Air Act air quality standards. Although emissions of ozone precursors, particularly NOx, have fallen over the past few decades, high-ozone events persist in much of the United States, including in the Northeast and mid-Atlantic.

In this context, we evaluate whether temporally or spatially differentiated emissions prices can cost-effectively reduce high-ozone events. To compare these policies with continuing the status quo of a uniform emissions price that does not vary by hour-of-day or region, we estimate (1) the statistical relationship between NOx emissions from power generation plants and local ozone formation, and (2) short-run marginal abatement costs for these generators. Unique to the literature, we estimate these relationships by geographic sub-region and hour-of-day to enable comparison of these policies.

We find rich variation over time and space in the marginal contribution of NOx to ozone. We also estimate short-run marginal abatement costs that negatively correlate with emissions rates of a sub-region's generation mix.

A central result of this analysis is that cost savings from a differentiated emissions price depends on the accuracy of the policy maker's prediction of the effect of emissions on high-ozone events. Future research may explore whether additional information beyond what we considered here, such as detailed weather forecasts or explicitly modeling air parcel movements over short time horizons (say, 48 hours), can improve accuracy of the key parameter estimates.

We end the paper with a brief discussion of policy implementation. Currently, EPA administers annual and seasonal emissions caps in which firms must submit allowances equal to emissions; each allowance covers one ton of emissions regardless of when or where the emissions occur.

At the beginning of each year, EPA could announce trading ratios such that one ton of emissions may count for more than one allowance, depending on whether the emissions are more likely to cause high-ozone events. The likelihood could be estimated using a methodology similar to the one used in this paper. EPA could also announce trading ratios multiple times during the ozone season based on updated information about atmospheric and economic conditions. However, it would be important for EPA to provide firms sufficient time to respond to changes in trading ratios. That the EPA already employs a trading ratio approach for penalizing states that exceed their NOx budgets may ease the burden of implementation. For example, CSAPR rules require generators to surrender three allowances for each ton of NOx emissions in excess of a specified level.

An alternative to trading ratios would be to replace the annual and seasonal emissions caps with spatially and temporally varying emissions prices, although this may require a more extensive regulatory process. Power generators in deregulated electricity markets (such as the NYISO and PJM systems, which cover our sample) submit supply offer curves reflecting their willingness to produce electricity at a schedule of prices for each hour of the day. If plant operators are aware of dynamic prices which change by hour-of-day, they can adjust their offer curves accordingly (the same is true for dynamic trading ratios). In this manner, a differentiated price scheme can be incorporated into existing systems to facilitate economically efficient dispatch of generation resources.

The research demonstrates that the accuracy of the price set on NOx emissions is important for achieving gains in cost effectiveness. Accuracy could be improved by using information on air parcel transport, meteorological conditions, and real-time ambient NOx and VOC concentrations.

One challenge lies in addressing formation of ozone from NOx which was released upwind, which can remain in the atmosphere for hours or days, thereby producing ozone far from its source. Our modeling suggests a significant and immediate sensitivity when it comes to NOx that is emitted in close proximity to monitors. Indeed, we show that significant reductions in ozone can be achieved when only contemporaneous NOx emissions are regulated. Additional modeling of air parcel transport would be necessary to determine both the extent to which NOx from non-local sources contributes to local ozone formation, and the precision with which these emissions can be traced back to their sources. If the contribution of non-local NOx is low and the uncertainty with respect to its contribution great, then a focus on contemporaneous, proximate, NOx may be preferable. If NOx contributions from outside sources can be measured with some degree of certainty, its contribution to non-local ozone formation can be reflected in an emissions price.

The EPA enforces NAAQS by setting different standards for each state. Although the marginal NOx contributions which we propose to use to define a differentiated emissions policy vary by location within states, implementation can still be performed at the state level. For example, under the emissions trading ratios approach, ratios would be allowed to vary locally but could be set in line with a state-level limit on high-ozone events (as is in place currently).

Appendix A: Chapter 1 Supplemental Materials

A.1 Data Set Construction Details

Generation and Marginal Costs

The main data set used for my analysis consists of hourly generation and marginal generation cost for fossil fuel-fired and wind power plants. Generation data for fossil generators is taken from the EPA's Continuous Emissions Monitoring System (CEMS), which tracks gross generation in MWh at the generation-unit level for all plants greater than 25 MW, excluding industrial and commercial cogeneration plants. I convert gross generation to net generation using conversion factors which I compute at the monthly level for each power plant using net generation data from EIA Form 923. The CEMS data set also tracks fuel consumption at the unit level, which I use to compute heat rates in MMBTU per MWh. To calculate marginal generation costs, I multiply heat rates by average fuel cost per MMBTU. I compute average monthly fuel costs for coal, natural gas, and oil using receipts and cost data tracked in EIA Form 923. Because these costs are only reported for a subset of plants (those operated by regulated entities), I compute the ERCOT-wide average cost for each fuel across all reporting plants and apply the system-wide average value in the marginal generation cost calculation.

For hourly wind generation, I use ERCOT's dispatch data, which includes telemetered wind

generation at the plant level starting in 2011. For pre-2011 data, I reconstruct plant generation data by combining site-specific wind profile data with monthly net generation data reported by EIA. ERCOT provides historic hourly wind profile data for 140 wind sites spanning 1980-2017. These data report a wind index value $w_{j,h}$ for each site j and hour h which I match to each wind plant. I convert these wind index values to pre-2011 generation $\hat{q}_{l,h}$, where l indexes plant, using a monthly plant-specific constant which I derive from EIA data. In particular:

$$\hat{q}_{l,h,m} = w_{j,h,m} k_{l,m}$$

and

$$k_{l,m} = \frac{q_{l,m}}{\sum_h w_{j,h,m}}$$

where m indexes month, and $q_{l,m}$ is plant l's net generation as reported in the EIA data, and plant l is matched to site j. That is, $k_{l,m}$ represents the monthly average ratio of the plant's net generation to its site wind index. This approach ensures that total quantity generated by each plant over the course of the month is equal to the value reported by EIA, and that generation is proportional to hourly wind strength. I assume zero marginal generation costs for wind plants.

Settlement Prices and Subsidies

ERCOT provides zonal settlement prices from 2001-2010 and nodal prices from 2011-2019. To assign settlement prices to power plants during the zonal period, I map each power plant to its geographically-defined zone using location coordinates provided by the EPA CEMS database and zonal delineations provided by ERCOT. During the nodal period, I map each gen-

eration resource in ERCOT to a settlement point using ERCOT's 60-Day Day-Ahead Market disclosure reports, as in Tsai and Eryilmaz, 2018. I construct a crosswalk which maps each generation resource (as designated by ERCOT) to its EIA plant identifier and CEMS identifier. This mapping allows an assignment of each nodal price to a particular power plant. I use the DSIRE database hosted by the North Carolina Clean Energy Technology Center at North Carolina State University to inform production tax credits used in my modeling.

Ownership Records

The Public Utility Commission of Texas requires that any operator of a power generation facility in Texas register as a Power Generation Company, and in so doing file a report which includes the ownership structure of the facility. Subsequent filings are submitted to record any changes in ownership. To assign ownership to each power plant, I reviewed records for each power plant individually, recording changes in ownership according to these filings. In many cases, the power plant was owned by a larger holding company. For each plant I record as the owner the company that is highest in the ownership structure and that has the largest share.

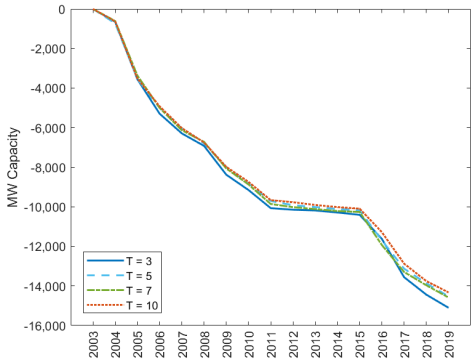
A.2 Sensitivity to Choice of T

In each period of the simulation (2003, ..., 2019) I solve the long-run model over a finite horizon with terminal period T. The actions taken in a given year reflect players' expectations about the industry evolution for T periods into the future; in period T the industry is assumed to remain static. Given a discount rate $\beta \in (0, 1]$, firm strategies converge to that of an infinite-period game with T sufficiently large because discounted profits earned in periods far into the future converge to zero. Moreover, there may exist a period in the future $\bar{T}$ beyond which present actions taken in period $t = 0$ have no meaningful impact, because profit-maximizing actions taken on behalf of players in periods $t = 1, 2, ..., \bar{T}$ serve to converge to the same equilibrium path in period $\bar{T} + 1$ and beyond, regardless of the actions taken in period $t = 0$. This property of "finite dependence," presented formally in Arcidiacono and Miller, 2011, allows that conditional choice probabilities be informed using payoffs that reflect only a small number of periods into the future; payoffs beyond period $\bar{T}$ are equivalent across choices and therefore "cancel out" when the player compares choice-specific value functions.

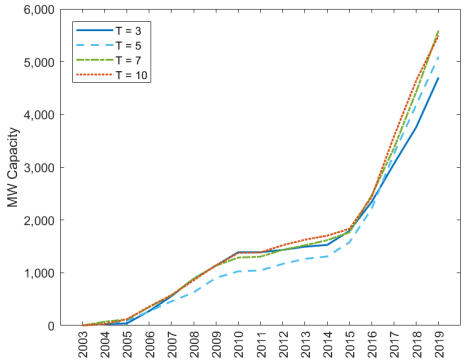
In this section I test whether the equilibrium path is sensitive to the choice of T. By choosing values of T such that discounted profits earned in period T and beyond may still vary significantly from zero, I am effectively testing whether the industry evolution exhibits finite dependence. In particular, I perform simulations of the baseline industry path for $T \in \{3, 5, 7, 10\}$. Figure A.1 illustrates the resulting industry paths at each choice of T and for $N = 50$ random draws of logit shocks. Equilibrium investment in new wind capacity increases as T increases; this effect accumulates to a disparity greater than 5,000 MW between the $T = 3$ path and the $T = 10$ path by the end of the 17-year period. Other disparities are smaller in magnitude; equilibrium

retirement and natural gas investment paths are comparable as T is varied.

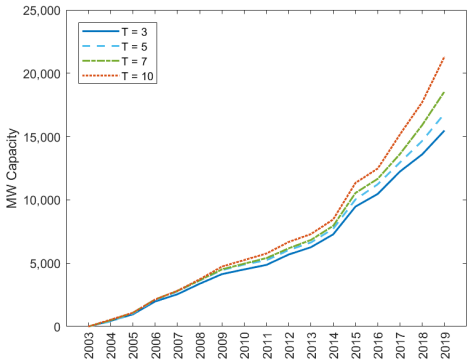
Figure A.1: Industry Path Sensitivity to Choice of T



(a) Fossil Retirements



(b) Fossil (Natural Gas) Investment



(c) Wind Investment

Appendix B: Chapter 2 Supplemental Materials: Counterfactual Modeling

Here I present a high-level overview of the procedure for performing the counterfactual exercise described in Section 2.3.2. I refer the reader to Holt, 2022a for a more detailed description of the structural model employed in my counterfactual analysis. The structural model I employ nests a short-run profit model which solves for hourly settlement prices given attributes of power plants installed on the system and demand. This profit model involves two components: solving for the hourly "uncongested" price, which results from equating an approximation of a system-wide aggregate supply function with demand, and an hourly "spike price." The spike price is the amount by which the observed settlement price exceeds the uncongested settlement price. This difference reflects both transmission constraints built into the nodal settlement price, as well as the system-wide scarcity price administered by ERCOT. I esimate the spike price component of the settlement price using a simple regression specification which expresses the price as a function of economic reserves. I define economic reserves as the difference between total fossil generation capacity, which is fixed by year, and net load. Net load varies by hour and includes total hourly demand on the system net of generation from renewable sources, nuclear, and a small subset of generation sources which are taken as exogenous (industrial and cogeneration plants, hydropower, biomass plants, and imports). In simulating the baseline outcome, I estimate a different set of reduced-form parameters before and after 2014, the year the SPM was introduced,

to capture the change in market design. As a result of the SPM, the latter period parameters will predict greater spike prices during periods of low reserves (greater demand). This is the same approach that is taken in the various counterfactual analyses performed in 2022a.

After building the short-run profit model, I use it to generate profits for 36,000 random states. Each random state is defined by a one-year period of hourly demand and a unique market composition in terms of installed capacity. States are chosen to reflect the range of possible outcomes of the industry's evolution under various policy counterfactuals of interest. I then use these results to fit a second model which describes plant-level profit as a linear function of state variables and interactions of state variables, resulting in a set of linear parameters λ. This approach is similar to the one taken in Sweeting, 2013. This step provides a means for expediently approximating short-run profits as a function of a set of state variables, nested within the long-run structural model. Since the structural model requires computing profits for many firms in many states of the world over many simulations, using the short-run model as-is would be computationally intractable.

To implement the counterfactual analysis, I repeat the process of estimating parameters λ but remove the structural break that occurs in 2014 from the short-run model. Profits that are reflected in the draws which are used to estimate λ therefore reflect an estimate of the spike price based on pre-2014 parameters (those which reflect pricing prior to the introduction of the SPM). In addition to introducing the SPM, ERCOT increased the system-wide offer cap, beginning in 2015, to \$9,000 from \$7,000. The system-wide offer cap is used as one of the state variables in the estimation of λ. In the counterfactual, I maintain the pre-2014 price cap of \$7,000 as the relevant value of this state variable when solving for firm profits in the stage game. This approach reflects that the choice of \$9,000 was made in tandem with the introduction of the SPM; increasing the

price cap is economically similar to imposing an administrative SPM because it allows prices to better reflect scarcity. The system-wide cap is typically not binding outside of periods during which the SPM is activated.

Appendix C: Chapter 3 Supplemental Materials

C.1 Derivation of Relationship Between Marginal Abatement Costs and Derivative of Emissions Rate With Respect to Allowance Price

In this section, we show that if we assume that aggregate marginal abatement costs are linear in abatement, the slope of the marginal abatement costs is inversely related to the derivative of the average emissions rate with respect to allowance price.

Suppressing hour and site subscripts for convenience, let $\bar{e}$ be the baseline emissions rate, which is the emissions rate when the allowance price equals zero. Let e be the emissions rate for a positive emissions price and q be the exogenous quantity of electricity generated. Recall that abatement is the change in emissions relative to emissions if the allowance price equals zero:

$$a \equiv -q(e - \bar{e}) \tag{C.1}$$

Multiplying Equation (3.2) by q implies $eq = \delta_0 q + \delta_1 qp$, where δ_0 includes the fixed effects and δ_1 is the coefficient on the allowance price. We assume that generators abate emissions such that marginal abatement cost m equals the emissions allowance price. Setting baseline emissions

equal to emissions at price $\bar{p} = 0$ implies:

$$\begin{aligned} a &= (\delta_0 q + \delta_1 qp) - (\delta_0 q + \delta_1 q\bar{p}) \\ &= \delta_1 qm \\ m &= -\frac{a}{\delta_1} \end{aligned} \tag{C.2}$$

Thus, the slope of the marginal abatement cost curve is the derivative of this equation with respect to a, or $-1/\delta_1$. In other words, the allowance price coefficient in Equation (3.2) is inversely proportional to the slope of the aggregate marginal abatement cost curve.

Note that we define abatement as tons of NOx, whereas the dependent variable in Equation (3.2) is the NOx emissions rate rather than the level of emissions. We can convert the change in emissions rate to abatement in tons if we assume total fossil fuel-fired generation to be exogenous, which is commonly done in the literature (e.g., Holland, Mansur, Muller, and Yates, 2016) and seems reasonable in our context, since non-fossil generation in the sub-regions we analyze is largely non-dispatchable and therefore cannot respond to changes in fossil fuel-fired generation. Because abatement is the reduction in emissions, we interpret the negative of the reciprocal of the hour-price interaction as the slope of the marginal abatement cost curve.

C.2 Regression Output

Table C.1 shows output from the full sample regressions. The marginal effect of NOx emissions in hour $h = 1$ (β_1 in Equation (3.1)) is equal to the NOx Tons coefficient reported in the first row of the table; the marginal effect of NOx emissions in hour $h = 2, ..., 24$ ($\beta_2, ..., \beta_{24}$

in Equation (3.1)) is equal to the sum of the NOx Tons coefficient and the coefficient associated with that hour's interaction term. For example, 2SLS results suggest that an increase in NOx emissions (within a 100-mile radius of the monitoring station) at 3pm (hour 15) in New York City increase the probability that ozone will exceed 70 ppb in that hour by 2.05 percent (equal to the sum of 0.0093 and 0.0112). These effects are illustrated graphically in Figure 3.4.

Temperature quartile indicator variables are used as a control. Temperature effects vary by sub-region. In all cases, temperatures in the highest quartile significantly increase the probability of a high-ozone event, but the magnitude of this effect varies by sub-region.

Table C.1: Ozone Regression Output, 2001-2019

	CT-RI-MA		NYC		Phil.-Wilm.		Wash.-Balt.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
NOx Tons	0.00722***	0.00539***	0.00731***	0.00925***	0.00754***	0.0130***	0.00924***	0.0190***
	(0.000364)	(0.000445)	(0.000293)	(0.000379)	(0.000194)	(0.000290)	(0.000232)	(0.000517)
NOx Tons * Hour 2	0.000599*	0.000531	0.000899***	0.000830***	0.000557***	0.000702***	0.00107***	0.00156***
	(0.000322)	(0.000367)	(0.000256)	(0.000241)	(0.000160)	(0.000144)	(0.000189)	(0.000192)
NOx Tons * Hour 3	0.00115***	0.000534	0.00176***	0.00135***	0.000910***	0.000865***	0.00150***	0.00223***
	(0.000296)	(0.000342)	(0.000238)	(0.000235)	(0.000145)	(0.000135)	(0.000181)	(0.000192)
NOx Tons * Hour 4	0.00138***	0.00118***	0.00237***	0.00205***	0.00108***	0.000900***	0.00158***	0.00233***
	(0.000286)	(0.000338)	(0.000239)	(0.000237)	(0.000144)	(0.000133)	(0.000146)	(0.000173)
NOx Tons * Hour 5	0.00132***	0.00131***	0.00197***	0.00157***	0.000972***	0.000577***	0.00101***	0.00107***
	(0.000294)	(0.000335)	(0.000226)	(0.000221)	(0.000140)	(0.000127)	(0.000135)	(0.000157)
NOx Tons * Hour 6	0.000487*	0.000588*	0.000953***	0.000300	0.000181	-0.000737***	-3.22e-05	-0.00115***
	(0.000285)	(0.000324)	(0.000215)	(0.000209)	(0.000137)	(0.000124)	(0.000129)	(0.000151)
NOx Tons * Hour 7	-0.000570**	-0.000412	-0.000340	-0.00124***	-0.000825***	-0.00220***	-0.000919***	-0.00292***
	(0.000281)	(0.000319)	(0.000213)	(0.000213)	(0.000137)	(0.000129)	(0.000130)	(0.000163)
NOx Tons * Hour 8	-0.00162***	-0.00142***	-0.00160***	-0.00270***	-0.00178***	-0.00346***	-0.00168***	-0.00416***
	(0.000291)	(0.000333)	(0.000217)	(0.000225)	(0.000140)	(0.000137)	(0.000133)	(0.000179)
NOx Tons * Hour 9	-0.00227***	-0.00152***	-0.00243***	-0.00358***	-0.00248***	-0.00440***	-0.00227***	-0.00502***
	(0.000333)	(0.000401)	(0.000242)	(0.000246)	(0.000147)	(0.000151)	(0.000149)	(0.000205)
NOx Tons * Hour 10	-0.00105**	-0.000166	-0.00172***	-0.00314***	-0.00120***	-0.00339***	-0.00224***	-0.00532***
	(0.000427)	(0.000473)	(0.000339)	(0.000336)	(0.000222)	(0.000230)	(0.000202)	(0.000254)
NOx Tons * Hour 11	0.00295***	0.00345***	0.00330***	0.00144***	0.00342***	0.000446	0.00336***	-0.00121***
	(0.000551)	(0.000575)	(0.000507)	(0.000500)	(0.000304)	(0.000300)	(0.000373)	(0.000382)
NOx Tons * Hour 12	0.00768***	0.00699***	0.00860***	0.00547***	0.00853***	0.00463***	0.00977***	0.00396***
	(0.000618)	(0.000642)	(0.000543)	(0.000546)	(0.000334)	(0.000337)	(0.000448)	(0.000464)
NOx Tons * Hour 13	0.0118***	0.0103***	0.0118***	0.00860***	0.0121***	0.00814***	0.0138***	0.00768***
	(0.000641)	(0.000677)	(0.000530)	(0.000556)	(0.000333)	(0.000352)	(0.000463)	(0.000499)
NOx Tons * Hour 14	0.0150***	0.0127***	0.0132***	0.0104***	0.0140***	0.0103***	0.0164***	0.0103***
	(0.000640)	(0.000686)	(0.000509)	(0.000549)	(0.000329)	(0.000357)	(0.000464)	(0.000508)
NOx Tons * Hour 15	0.0156***	0.0141***	0.0133***	0.0112***	0.0140***	0.0106***	0.0176***	0.0118***
	(0.000634)	(0.000693)	(0.000508)	(0.000555)	(0.000325)	(0.000357)	(0.000460)	(0.000509)
NOx Tons * Hour 16	0.0148***	0.0136***	0.0127***	0.0109***	0.0134***	0.0102***	0.0173***	0.0117***
	(0.000622)	(0.000683)	(0.000498)	(0.000544)	(0.000321)	(0.000353)	(0.000455)	(0.000503)
NOx Tons * Hour 17	0.0119***	0.0115***	0.0107***	0.00893***	0.0125***	0.00926***	0.0152***	0.00975***
	(0.000623)	(0.000681)	(0.000500)	(0.000535)	(0.000320)	(0.000348)	(0.000452)	(0.000493)
NOx Tons * Hour 18	0.00858***	0.00895***	0.00745***	0.00596***	0.0103***	0.00667***	0.0122***	0.00730***
	(0.000622)	(0.000678)	(0.000514)	(0.000531)	(0.000321)	(0.000340)	(0.000440)	(0.000470)
NOx Tons * Hour 19	0.00418***	0.00472***	0.00382***	0.00233***	0.00588***	0.00240***	0.00607***	0.00159***
	(0.000600)	(0.000642)	(0.000495)	(0.000497)	(0.000314)	(0.000315)	(0.000402)	(0.000415)
NOx Tons * Hour 20	0.00118**	0.00226***	-0.000356	-0.00157***	0.00130***	-0.00155***	0.00109***	-0.00307***
	(0.000545)	(0.000579)	(0.000415)	(0.000405)	(0.000276)	(0.000270)	(0.000326)	(0.000339)
NOx Tons * Hour 21	-0.000779	0.000394	-0.00282***	-0.00367***	-0.000740***	-0.00320***	-0.00152***	-0.00528***
	(0.000482)	(0.000510)	(0.000305)	(0.000319)	(0.000246)	(0.000242)	(0.000250)	(0.000284)
NOx Tons * Hour 22	-0.00190***	-0.000822*	-0.00299***	-0.00374***	-0.00151***	-0.00354***	-0.00182***	-0.00496***
	(0.000423)	(0.000477)	(0.000279)	(0.000284)	(0.000224)	(0.000215)	(0.000225)	(0.000245)
NOx Tons * Hour 23	-0.00177***	-0.00135***	-0.00230***	-0.00301***	-0.00125***	-0.00274***	-0.00164***	-0.00375***
	(0.000406)	(0.000441)	(0.000272)	(0.000259)	(0.000209)	(0.000184)	(0.000186)	(0.000206)
NOx Tons * Hour 24	-0.00120***	-0.00109***	-0.00120***	-0.00159***	-0.000598***	-0.00151***	-0.000603***	-0.00181***
	(0.000370)	(0.000401)	(0.000267)	(0.000255)	(0.000200)	(0.000170)	(0.000196)	(0.000198)
Temperature Quartile 2	0.00791***	0.00813***	0.00154***	0.000517	0.000275	-0.00265***	-0.00797***	-0.0134***
	(0.000324)	(0.000333)	(0.000410)	(0.000435)	(0.000248)	(0.000282)	(0.000311)	(0.000428)
Temperature Quartile 3	0.0185***	0.0198***	-0.00250***	-0.00430***	-0.00523***	-0.0120***	-0.0192***	-0.0310***
	(0.000638)	(0.000673)	(0.000768)	(0.000855)	(0.000453)	(0.000554)	(0.000574)	(0.000853)
Temperature Quartile 4	0.107***	0.113***	0.0587***	0.0570***	0.0394***	0.0260***	0.0236***	0.00225
		(0.00147)	(0.00142)	(0.00169)	(0.000816)	(0.00111)	(0.000940)	(0.00156)
Constant	-0.0290***	-0.0163***	-0.0829***	-0.0883***	0.0157***	-0.132***	-0.0998***	0.0241***
	(0.00280)	(0.00305)	(0.00315)	(0.00428)	(0.00111)	(0.00327)	(0.00330)	(0.00287)
Observations	596,734	596,734	361,650	361,650	914,284	914,284	554,407	554,407
R-squared	0.168	0.167	0.222	0.221	0.218	0.217	0.200	0.196

Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

First stage regression output is shown in Table C.2. Results suggest that our instrument—

simulated NOx tons—is a strong predictor of actual NOx emissions. There are 24 instrumented variables in our regressions (NOx and NOx interacted with hour-of-day). This table presents first stage results without hour-of-day interaction terms.

Table C.2: Ozone Regression Output, First Stage, 2001-2019

VARIABLES	(1) CT-RI-MA	(2) NYC	(3) Phil.-Wilm.	(4) Wash.-Balt.
Simulated NOx Tons	0.473***	0.587***	0.691***	0.841***
	(0.00195)	(0.00241)	(0.00179)	(0.00257)
Temperature Quartile 2	0.217***	0.418***	0.366***	0.485***
	(0.00359)	(0.00683)	(0.00467)	(0.00598)
Temperature Quartile 3	0.669***	1.121***	0.993***	1.126***
	(0.00583)	(0.0104)	(0.00612)	(0.00696)
Temperature Quartile 4	2.009***	3.067***	2.823***	2.495***
	(0.00956)	(0.0162)	(0.00900)	(0.00881)
Constant	3.785***	3.975***	-0.728***	-2.343***
	(0.0224)	(0.0448)	(0.0148)	(0.0188)
Observations	596,757	361,650	914,287	554,409
R-squared	0.832	0.848	0.857	0.888
RMSE	1.386	2.049	2.054	1.588

Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tables C.3 and C.4 present regression output for the *ex ante* (2011-2015) period, and Tables C.5 and C.6 present regression output for the *ex post* (2016-2019) period. These regression results are used to define the differentiated emissions prices, which vary by hour-of-day and sub-region.

Table C.3: Ozone Regression Output, *Ex Ante* Period (2011-2015)

VARIABLES	(1) NYC	(2) Phil.-Wilm.	(3) Wash.-Balt.
NOx Tons	0.00438***	0.0123***	0.00922***
	(0.00140)	(0.000662)	(0.00104)
NOx Tons * Hour 2	0.000312	0.000351	0.000620*
	(0.000663)	(0.000403)	(0.000337)
NOx Tons * Hour 3	0.00238***	0.000879**	0.000790**
	(0.000800)	(0.000367)	(0.000325)
NOx Tons * Hour 4	0.00633***	0.00137***	0.00189***
	(0.000788)	(0.000341)	(0.000387)
NOx Tons * Hour 5	0.00656***	0.000871***	0.00229***
	(0.000743)	(0.000333)	(0.000318)
NOx Tons * Hour 6	0.00467***	-6.55e-05	0.00145***
	(0.000719)	(0.000331)	(0.000316)
NOx Tons * Hour 7	0.00264***	-0.000899***	0.000713**
	(0.000738)	(0.000333)	(0.000313)
NOx Tons * Hour 8	0.000927	-0.00211***	-0.000470
	(0.000748)	(0.000337)	(0.000310)
NOx Tons * Hour 9	-0.00101	-0.00332***	-0.00151***
	(0.000767)	(0.000369)	(0.000350)
NOx Tons * Hour 10	0.00428**	-0.00259***	0.00175*
	(0.00184)	(0.000584)	(0.000907)
NOx Tons * Hour 11	0.0184***	0.00515***	0.0172***
	(0.00280)	(0.00109)	(0.00166)
NOx Tons * Hour 12	0.0320***	0.0140***	0.0287***
	(0.00259)	(0.00126)	(0.00195)
NOx Tons * Hour 13	0.0305***	0.0157***	0.0332***
	(0.00260)	(0.00125)	(0.00206)
NOx Tons * Hour 14	0.0304***	0.0151***	0.0358***
	(0.00250)	(0.00124)	(0.00203)
NOx Tons * Hour 15	0.0285***	0.0142***	0.0352***
	(0.00232)	(0.00120)	(0.00199)
NOx Tons * Hour 16	0.0243***	0.0136***	0.0322***
	(0.00222)	(0.00118)	(0.00202)
NOx Tons * Hour 17	0.0228***	0.0136***	0.0268***
	(0.00215)	(0.00117)	(0.00199)
NOx Tons * Hour 18	0.0184***	0.0120***	0.0215***
	(0.00216)	(0.00115)	(0.00191)
NOx Tons * Hour 19	0.0129***	0.00508***	0.0134***
	(0.00216)	(0.00102)	(0.00161)
NOx Tons * Hour 20	0.00605***	0.000780	0.00366***
	(0.00194)	(0.000866)	(0.00108)
NOx Tons * Hour 21	0.00306*	-0.00265***	-5.15e-05
	(0.00169)	(0.000667)	(0.000836)
NOx Tons * Hour 22	0.000413	-0.00340***	-0.00134**
	(0.00134)	(0.000544)	(0.000676)
NOx Tons * Hour 23	-0.000545	-0.00228***	-0.00122***
	(0.000795)	(0.000502)	(0.000412)
NOx Tons * Hour 24	0.00130*	-0.000727	-0.000132
	(0.000708)	(0.000522)	(0.000290)
Temperature Quartile 2	0.000727	-0.00456***	-0.00605***
	(0.000564)	(0.000346)	(0.000496)
Temperature Quartile 3	-0.00436***	-0.0134***	-0.0156***
	(0.00128)	(0.000761)	(0.00111)
Temperature Quartile 4	0.0120***	-0.00516***	-0.0163***
	(0.00355)	(0.00186)	(0.00258)
Constant	0.00652***	-0.0603***	-0.0386***
	(0.00197)	(0.00327)	(0.00395)
Observations	103,324	282,428	163,912
R-squared	0.185	0.154	0.178

Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C.4: Ozone Regression Output, *Ex Ante* Period (2011-2015), First Stage

VARIABLES	(1) NYC	(2) Phil.-Wilm.	(3) Wash.-Balt.
Simulated NOx Tons	0.590***	0.689***	0.792***
	(0.00637)	(0.00351)	(0.00484)
Temperature Quartile 2	0.0997***	0.136***	0.269***
	(0.00770)	(0.00611)	(0.00782)
Temperature Quartile 3	0.461***	0.562***	0.693***
	(0.0131)	(0.00821)	(0.00960)
Temperature Quartile 4	1.996***	2.159***	2.124***
	(0.0241)	(0.0131)	(0.0131)
Constant	1.534***	0.481***	-1.650***
	(0.0289)	(0.0263)	(0.0272)
Observations	103,324	282,428	163,912
R-squared	0.708	0.738	0.705
RMSE	1.397	1.565	1.285

There are 24 instrumented variables in our regressions (NOx and NOx interacted with hour-of-day). This table presents first stage results without hour-of-day interaction terms. Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C.5: Ozone Regression Output, *Ex Post* Period (2016-2019)

VARIABLES	(1) NYC	(2) Phil.-Wilm.	(3) Wash.-Balt.
NOx Tons	0.0226***	0.0107***	0.0163***
	(0.00366)	(0.000882)	(0.00128)
NOx Tons * Hour 2	0.00311*	0.000590*	-2.47e-05
	(0.00163)	(0.000353)	(0.000678)
NOx Tons * Hour 3	0.00219	0.00118***	0.000590
	(0.00155)	(0.000364)	(0.000686)
NOx Tons * Hour 4	0.00613***	0.00188***	0.000367
	(0.00162)	(0.000376)	(0.000697)
NOx Tons * Hour 5	0.00700***	0.00131***	-0.00150**
	(0.00144)	(0.000344)	(0.000662)
NOx Tons * Hour 6	0.00200	0.000470	-0.00253***
	(0.00137)	(0.000327)	(0.000657)
NOx Tons * Hour 7	-0.00350**	-0.000746**	-0.00332***
	(0.00156)	(0.000329)	(0.000662)
NOx Tons * Hour 8	-0.00778***	-0.00189***	-0.00414***
	(0.00169)	(0.000340)	(0.000670)
NOx Tons * Hour 9	-0.0111***	-0.00311***	-0.00513***
	(0.00182)	(0.000357)	(0.000700)
NOx Tons * Hour 10	-0.0126***	-0.00467***	-0.00701***
	(0.00220)	(0.000388)	(0.000798)
NOx Tons * Hour 11	-0.00319	-0.00180**	-0.00390***
	(0.00335)	(0.000908)	(0.00151)
NOx Tons * Hour 12	0.00641	0.00776***	0.00648***
	(0.00393)	(0.00143)	(0.00237)
NOx Tons * Hour 13	0.0155***	0.0113***	0.0126***
	(0.00470)	(0.00162)	(0.00259)
NOx Tons * Hour 14	0.0140***	0.0110***	0.0144***
	(0.00447)	(0.00160)	(0.00264)
NOx Tons * Hour 15	0.0131***	0.00915***	0.00995***
	(0.00428)	(0.00154)	(0.00243)
NOx Tons * Hour 16	0.00655	0.00873***	0.0113***
	(0.00407)	(0.00152)	(0.00242)
NOx Tons * Hour 17	-0.000240	0.00605***	0.00618***
	(0.00385)	(0.00145)	(0.00228)
NOx Tons * Hour 18	-0.00520	0.00293**	0.00314
	(0.00364)	(0.00132)	(0.00205)
NOx Tons * Hour 19	-0.00935***	-0.00271***	-0.00248
	(0.00333)	(0.000998)	(0.00170)
NOx Tons * Hour 20	-0.0136***	-0.00404***	-0.00409***
	(0.00295)	(0.000755)	(0.00140)
NOx Tons * Hour 21	-0.0130***	-0.00431***	-0.00551***
	(0.00281)	(0.000630)	(0.00101)
NOx Tons * Hour 22	-0.0111***	-0.00364***	-0.00446***
	(0.00250)	(0.000434)	(0.000867)
NOx Tons * Hour 23	-0.00885***	-0.00233***	-0.00307***
	(0.00198)	(0.000400)	(0.000809)
NOx Tons * Hour 24	-0.00556***	-0.00118***	-0.00215**
	(0.00153)	(0.000409)	(0.000837)
Temperature Quartile 2	0.00677***	0.00469***	-0.00436***
	(0.00101)	(0.000489)	(0.000676)
Temperature Quartile 3	0.00677***	0.00509***	-0.00874***
	(0.00199)	(0.000877)	(0.00114)
Temperature Quartile 4	0.0259***	0.0266***	0.00520***
	(0.00440)	(0.00173)	(0.00198)
Constant	-0.0133***	-0.0451***	-0.0113***
	(0.00278)	(0.00264)	(0.00349)
Observations	89,157	223,224	149,972
R-squared	0.093	0.080	0.075

Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C.6: Ozone Regression Output, *Ex Post* Period (2016-2019), First Stage

VARIABLES	(1) NYC	(2) Phil.-Wilm.	(3) Wash.-Balt.
Simulated NOx Tons	0.599***	0.752***	0.857***
	(0.00836)	(0.00390)	(0.00457)
Temperature Quartile 2	0.0753***	0.0847***	0.278***
	(0.00588)	(0.00430)	(0.00548)
Temperature Quartile 3	0.329***	0.327***	0.560***
	(0.00945)	(0.00561)	(0.00628)
Temperature Quartile 4	1.359***	1.257***	1.344***
	(0.0167)	(0.00825)	(0.00795)
Constant	0.0651***	-0.421***	-0.792***
	(0.0224)	(0.0243)	(0.0152)
Observations	89,157	223,227	149,974
R-squared	0.706	0.769	0.718
RMSE	0.860	0.907	0.796

There are 24 instrumented variables in our regressions (NOx and NOx interacted with hour-of-day). This table presents first stage results without hour-of-day interaction terms. Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C.7 shows the regression results from the MAC estimation procedure (Equation (3.2)). These coefficient values are depicted graphically in Figure 3.7. The dependent variable in these regressions is the average hourly emissions rate, in NOx lbs. per MWh electric power output, of all generators within a 100-mile radius of each AQS monitor. The right-hand-side variables of interest are the NOx allowance price interacted with hour-of-day.

Table C.7: Emissions Rate Regression Output, Dependent Variable = lbs. NOx Emissions per MWh

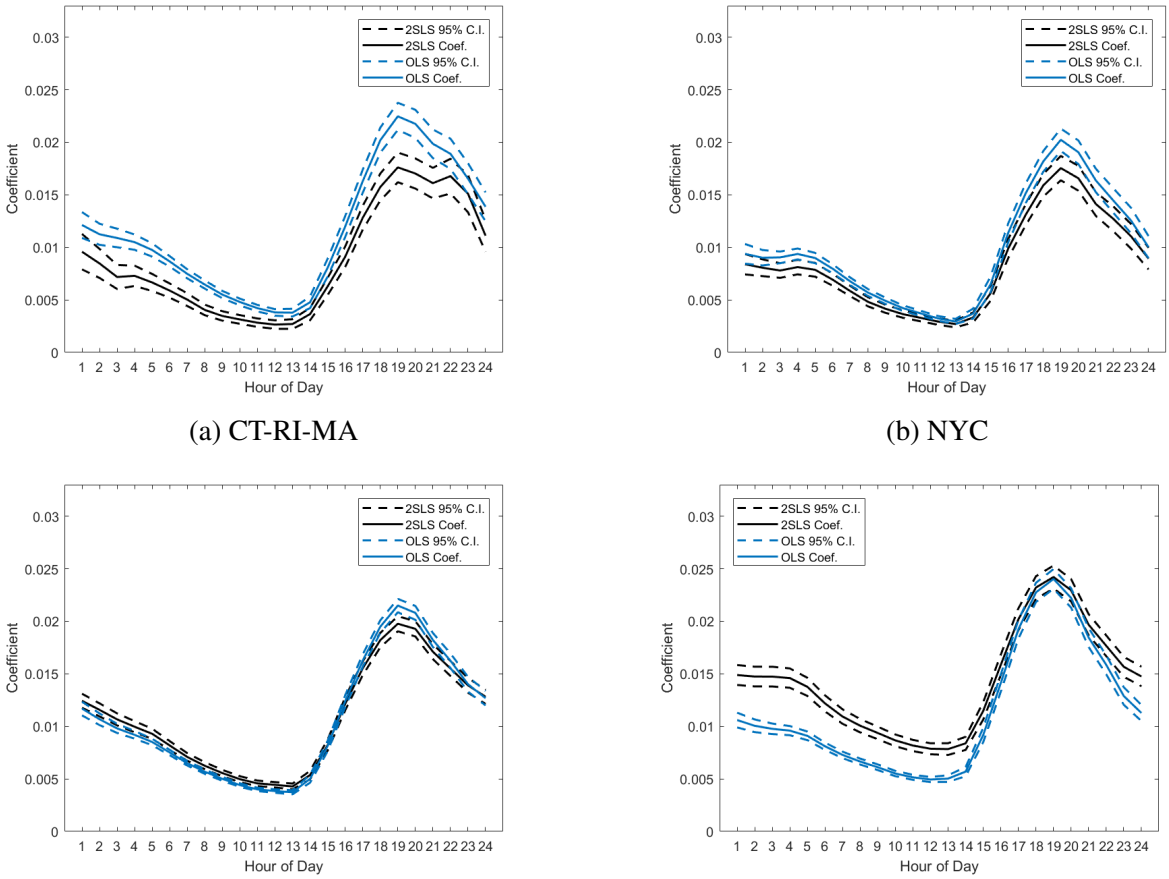
VARIABLES	(1) NYC	(2) Phil.-Wilm.	(3) Wash.-Balt.
NOx Price	-0.0000176*	-0.0003901***	-0.0008101***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 2	-0.0000017	0.0000141	0.0000454
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 3	-0.0000313**	0.0000037	0.0001037**
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 4	-0.0000217*	-0.0000092	0.0001108***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 5	-0.0000187	-0.0000056	0.0000583
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 6	-0.0000427***	-0.0000125	0.0000468
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 7	-0.0000405***	0.0000031	0.0000905**
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 8	-0.0000758***	-0.0000183	0.0000720*
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 9	-0.0000414***	-0.0000113	0.0000435
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 10	-0.0000387**	-0.0000407**	0.0000443
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 11	-0.0000963***	-0.0000849***	0.0000287
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 12	-0.0001075***	-0.0000932***	0.0000334
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 13	-0.0001130***	-0.0000854***	0.0000709*
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 14	-0.0001279***	-0.0000788***	0.0000823**
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 15	-0.0001352***	-0.0000692***	0.0000583
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 16	-0.0001276***	-0.0000565**	0.0000734*
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 17	-0.0001488***	-0.0000597***	0.0000971**
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 18	-0.0001756***	-0.0000810***	0.0001381***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 19	-0.0001887***	-0.0001135***	0.0001110***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 20	-0.0001482***	-0.0000682***	0.0001613***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 21	-0.0001752***	-0.0000812***	0.0002068***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 22	-0.0001310***	-0.0000563***	0.0001961***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 23	-0.0000532***	-0.0000038	0.0001583***
	(0.000)	(0.000)	(0.000)
NOx Price * Hour 24	-0.0000197	-0.0000184	0.0000030
	(0.000)	(0.000)	(0.000)
Constant	0.1706042***	0.7170494***	0.9174670***
	(0.005)	(0.006)	(0.011)
Observations	98,970	123,268	82,519
R-squared	0.356	0.512	0.383
RMSE	0.0859	0.105	0.172

Fixed effects by hour-of-day, day-of-week, month, year, station not shown. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.3 Coefficient Estimates Using Alternative Dependent Variable

Figure C.1 illustrates coefficient estimates when the dependent variable is defined as an indicator equal to one if the 8-hour moving average ozone concentration exceeds 70 ppb. Results are comparable to those shown in Figure 3.4, but the effect of NOx on 8-hr. ozone is lagged. The EPA defines attainment at the county level for each state as when the three-year average of the year's fourth-highest 8-hour moving average concentration level is no greater than 70 ppb.

Figure C.1: OLS and 2SLS Regression Results, 2001-2019, Dep. Var. = 1 if 8-hr Moving Average Ozone Concentration $>$ 70 ppb



(a) CT-RI-MA

(b) NYC

(c) Phil.-Wilm.

(d) Wash.-Balt.

The regressions are the same as Equation (3.1), except that the dependent variable uses the 8-hour moving average of the ozone level rather than the contemporaneous ozone level.

Bibliography

Aldy, J. E., Gerarden, T. D., & Sweeney, R. L. (2018). Investment versus output subsidies: Implications of alternative incentives for wind energy. *NBER Working Paper No. w24378*.

Arcidiacono, P., & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, *79*(6), 1823–1867.

Bajo-Buenestado, R. (2021). Operating reserve demand curve, scarcity pricing and intermittent generation: Lessons from the texas ercot experience. *Energy Policy*, *149*, 112057.

Baumol, W. J., Baumol, W. J., Oates, W. E., Bawa, V. S., Bawa, W., Bradford, D. F., Baumol, W. J., et al. (1988). *The theory of environmental policy*. Cambridge university press.

Bell, J. N. B., & Treshow, M. (2002). *Air pollution and plant life*. John Wiley & Sons.

Bell, M. L., Goldberg, R., Hogrefe, C., Kinney, P. L., Knowlton, K., Lynn, B., Rosenthal, J., Rosenzweig, C., & Patz, J. A. (2007). Climate change, ambient ozone, and health in 50 us cities. *Climatic Change*, *82*(1), 61–76.

Blaszczak, R. J. (1999). *Nitrogen oxides (nox): Why and how they are controlled* (tech. rep. EPA-456/F-99-006R). US Environmental Protection Agency.

Boiteux, M. (1960). Peak-load pricing. *The Journal of Business*, *33*(2), 157–179.

Borenstein, S., Bushnell, J., & Knittel, C. R. (1999). Market power in electricity markets: Beyond concentration measures. *The Energy Journal*, *20*(4).

Borenstein, S., Bushnell, J. B., & Wolak, F. A. (2002). Measuring market inefficiencies in california's restructured wholesale electricity market. *American Economic Review*, *92*(5), 1376–1405.

Brown, J., Bowman, C. et al. (2013). *Integrated science assessment for ozone and related photochemical oxidants* (tech. rep. EPA/600/R-20/012, 2020). US Environmental Protection Agency.

Buchanan, J. M. (1969). External diseconomies, corrective taxes, and market structure. *The American Economic Review*, *59*(1), 174–177.

Bushnell, J., & Ishii, J. (2007). An equilibrium model of investment in restructured energy markets. *Center for the Study of Energy Markets Working Paper No. 164*.

Carden, K., & Dombrowsky, A. K. (2021). *Estimation of the market equilibrium and economically optimal reserve margins for the ercot region for 2024* (tech. rep.). Astrapé Consulting.

Chao, H.-p., & Wilson, R. (1987). Priority service: Pricing, investment, and market organization. *The American Economic Review*, *77*(5), 899–916.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, *6*, 5549–5632.

Coulon, M., Powell, W. B., & Sircar, R. (2013). A model for hedging load and price risk in the texas electricity market. *Energy Economics*, *40*, 976–988.

Cramton, P. (2017). Electricity market design. *Oxford Review of Economic Policy*, *33*(4).
Cullen, J. A., & Reynolds, S. S. (2017). Market dynamics and investment in the electricity sector.
Davis, R. J., Holladay, J. S., & Sims, C. (2021). Coal-fired power plant retirements in the us. *NBER Working Paper No. 28949*.
Davison, A., & Barnes, J. (1998). Effects of ozone on wild plants. *New Phytologist*, *139*(1), 135–151.
EIA. (2020). Energy and the environment explained [https://www.eia.gov/energyexplained/energy-and-the-environment/where-greenhouse-gases-come-from.php].
Ericson, R., & Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of economic studies*, *62*(1), 53–82.
Fell, H., & Linn, J. (2013). Renewable electricity policies, heterogeneity, and cost effectiveness. *Journal of Environmental Economics and Management*, *66*(3), 688–707.
Fowlie, M., & Muller, N. (2019). Market-based emissions regulation when damages vary across sources: What are the gains from differentiation? *Journal of the Association of Environmental and Resource Economists*, *6*(3), 593–632.
Fowlie, M., Reguant, M., & Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, *124*(1), 249–302.
Grainger, C., & Schreiber, A. (2019). Discrimination in ambient air pollution monitoring? *AEA Papers and Proceedings*, *109*, 277–82.
Grainger, C., Schreiber, A., & Chang, W. (2018). Do regulators strategically avoid pollution hotspots when siting monitors? evidence from remote sensing of air pollution. *University of Wisconsin unpublished manuscript*.
Grewe, V., Dahlmann, K., Matthes, S., & Steinbrecht, W. (2012). Attributing ozone to nox emissions: Implications for climate mitigation measures. *Atmospheric environment*, *59*, 102–107.
Hogan, W. W. (2005). On an "energy only" electricity market design for resource adequacy.
Hogan, W. W. (2013). Electricity scarcity pricing through operating reserves. *Economics of Energy & Environmental Policy*, *2*(2), 65–86.
Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2016). Are there environmental benefits from driving electric vehicles? the importance of local factors. *American Economic Review*, *106*(12), 3700–3729.
Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2020). Decompositions and policy consequences of an extraordinary decline in air pollution from electricity generation. *American Economic Journal: Economic Policy*, *12*(4), 244–74.
Holt, C. (2022a). Long-run dynamics of industry transition: A structural model of the texas electricity market 2003-2019. *Job market paper*.
Holt, C. (2022b). Scarcity pricing and imperfect competition in wholesale electricity markets.
Igami, M. (2017). Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. *Journal of Political Economy*, *125*(3), 798–847.
IMM. (2020). *2019 state of the market report for the ercot electricity markets* (tech. rep.). Potomac Economics, Independent Market Monitor for ERCOT.
Ito, K., De Leon, S. F., & Lippmann, M. (2005). Associations between ozone and daily mortality: Analysis and meta-analysis. *Epidemiology*, *16*(4), 446–457.

Johnsen, R., Lariviere, J., & Wolff, H. (2019). Fracking, coal, and air quality. *Journal of the Association of Environmental and Resource Economists*, *6*(5), 1001–1037.

Johnston, J., Henriquez-Auba, R., Maluenda, B., & Fripp, M. (2019). Switch 2.0: A modern platform for planning high-renewable power systems. *SoftwareX*, *10*, 100251.

Joskow, P., & Tirole, J. (2007). Reliability and competitive electricity markets. *The RAND Journal of Economics*, *38*(1), 60–84.

Karaduman, O. (2020). Economics of grid-scale energy storage. *Job market paper*.

Kreps, D. M., & Scheinkman, J. A. (1983). Quantity precommitment and bertrand competition yield cournot outcomes. *The Bell Journal of Economics*, *14*(2), 326–337.

Levin, T., & Botterud, A. (2015). Electricity market design for generator revenue sufficiency with increased variable generation. *Energy Policy*, *87*, 392–406.

Linn, J. (2008). Technological modifications in the nitrogen oxides tradable permit program. *The Energy Journal*, *29*(3), 153–176.

Linn, J., & McCormack, K. (2019). The roles of energy markets and environmental regulation in reducing coal-fired plant profits and electricity sector emissions. *The RAND Journal of Economics*, *50*(4), 733–767.

Montgomery, W. D. (1972). Markets in licenses and efficient pollution control programs. *Journal of economic theory*, *5*(3), 395–418.

Muller, N. Z. (2012). The design of optimal climate policy with air pollution co-benefits. *Resource and Energy Economics*, *34*(4), 696–722.

Myatt, J. (2017). Market power and long-run technology choice in the us electricity industry. *Job market paper*.

Palmer, K., & Burtraw, D. (2005). Cost-effectiveness of renewable electricity policies. *Energy economics*, *27*(6), 873–894.

Pigou, A. C. (1924). *The economics of welfare*. Macmillan.

Raimi, D. (2017). *Decommissioning us power plants: Decisions, costs, and key issues* (tech. rep.). Resources for the Future.

Reguant, M. (2014). Complementary bidding mechanisms and startup costs in electricity markets. *The Review of Economic Studies*, *81*(4), 1708–1742.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, *80*(3), 1019–1061.

Schmalensee, R. (2019). On the efficiency of competitive energy storage. *Available at SSRN 3405058*.

Sherlock, M. (2020). *The renewable electricity production tax credit: In brief* (tech. rep. R43453). Congressional Research Service.

Sillman, S. (1999). The relation between ozone, nox and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment*, *33*(12), 1821–1845.

Stoft, S. (2002). *Power system economics: Designing markets for electricity*. IEEE Press, Piscataway.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica*, *81*(5), 1763–1803.

Sweeting, A. (2015). A model of non-stationary dynamic price competition with an application to platform design. *Available at SSRN 2672027*.

Tietenberg, T. H. (1990). Economic instruments for environmental regulation. *Oxford review of economic policy*, *6*(1), 17–33.

Tsai, C.-H., & Eryilmaz, D. (2018). Effect of wind generation on ercot nodal prices. *Energy Economics*, *76*, 21–33.

Turvey, R., & Anderson, D. (1977). *Electricity economics: Essays and case studies*. Johns Hopkins University Press, Baltimore.

Weintraub, G. Y., Benkard, C. L., Jeziorski, P., & Roy, B. V. (2010). Nonstationary oblivious equilibrium. *The Johns Hopkins University, Department of Economics, Working Paper No. 568*.

Weintraub, G. Y., Benkard, C. L., & Van Roy, B. (2008). Markov perfect industry dynamics with many firms. *Econometrica*, *76*(6), 1375–1411.

Woerman, M. (2019). Market size and market power: Evidence from the texas electricity market. *Job market paper*.

Xiao, X., Cohan, D. S., Byun, D. W., & Ngan, F. (2010). Highly nonlinear ozone formation in the houston region and implications for emission controls. *Journal of Geophysical Research: Atmospheres*, *115*(D23).