Dynamic Incentives and Permit Market Equilibrium in Cap-and-Trade Regulation*

Yuta Toyama†

May 14, 2023

Abstract

This paper develops and estimates a dynamic structural model of emissions abatement, investment, and permit trading with banking under cap-and-trade regulation. The model accounts for forward-looking behavior and transaction costs in the permit market, which determine the temporal and geographical distribution of emissions in equilibrium, and, thus, the welfare implications of the regulation. The model is applied to the US Acid Rain Program to evaluate the role of regulatory designs. Permit banking mitigates inefficiencies arising from transaction costs and modifies the timing of emissions. An emissions tax policy could achieve an outcome close to dynamic cap-and-trade without transaction costs.

Key words: *cap-and-trade regulation, dynamic equilibrium model, gains from trade, permit banking, transaction costs, electricity industry, sulfur dioxide emissions, non-uniformly mixed pollutants*
JEL Code: D22, L94, Q52, Q58

*This paper is based on Chapter 1 of my Ph.D. dissertation at Northwestern University. I would like to thank Rob Porter and Mar Reguant for their guidance and encouragement throughout the project. I am grateful to Toshi Arimura, Vivek Bhattacharya, Igal Hendel, Gaston Illanes, Daiya Isogawa, Koichiro Ito, Kei Kawai, Zhentong Lu, Mark Roberts, Lola Segura, Max Tabord-Meehan, Hidenori Takahashi, Yuta Takahashi, Nicholas Vreugdenhil, Naoki Wakamori, Yasutora Watanabe, Anthony Lee Zhang, and the seminar participants at many institutions and conferences for their helpful comments and discussions. I also thank the Co-Editor, John Asker, and four anonymous referees whose comments and suggestions significantly improved the paper. Rentaro Utamaru provided superb research assistance. I gratefully acknowledge the financial support issued by the Research Institute for Environmental Economics and Management, a Waseda University Grant for Special Research Projects (2018S-223) and JSPS KAKENHI Grant Number 19K13696. The previous version was circulated under the title "Incentives to Invest, Storable Permits, and Transaction Costs in Market-Based Environmental Regulation." All errors are my own.

†Graduate School of Economics, Waseda University, 1-6-1 Nishi-Waseda, Tokyo, Japan. E-mail: yuta.toyama@gmail.com

1 Introduction

Designing the allocation mechanism for scarce public resources is a key issue in broad fields of economics. This issue is particularly relevant in environmental economics because the optimal allocation of natural resources depends on their externalities. Since the seminal work of Coase (1960), economists have advocated a market-based allocation mechanism that utilizes economic incentives to achieve efficient resource allocation. This idea has been adopted in various settings, including fishery quota, spectrum license, water rights, and permits for air pollutants. While the literature has focused on such mechanisms from a static perspective, it has put less emphasis on dynamic aspects of the regulatory design and the agent's behavior.[1] This paper therefore provides an empirical study on the dynamic aspects of cap-and-trade regulation, a canonical example of market-based resource allocation schemes.[2]

In cap-and-trade regulation, the regulator uses *emissions permits*, the tradable rights for producing emissions, as a regulatory tool. Regulated firms can reallocate such permits through trading so that they can achieve the target level of aggregate emissions in a flexible and efficient manner. Cap-and-trade is now widely adopted in the air pollution regulations of the United States and the European Union.

While the theoretical framework originally proposed by Coase (1960) was a static one, its implementation in practice involves several dynamic aspects. First, a cap-and-trade program spans multiple periods, during which the regulatory standard becomes stricter (i.e., the emissions cap decreases). The regulator often allows for the inter-temporal reallocation of emissions permits, which is referred to as a permit banking system. Considering the dynamic nature of the regulatory framework, firms should make abatement decisions in a forward-looking manner. Such dynamic consideration is particularly crucial when making investment decision on abatement technology, which can be a major margin of emissions reduction. Although these dynamic features prevail in many cap-and-trade programs, empirical studies in the literature focus primarily on static decisions in the steady-state (see, e.g., Carlson, Burtraw, Cropper, and Palmer 2000, Fowlie 2010b, and Chan 2015).

The goal of this study is to propose an empirical framework that evaluates the welfare consequences of cap-and-trade programs by accounting for the dynamic nature of their regulatory design and the forward-looking incentives of regulated firms. To do this, I develop and estimate a new model of emissions abatement and permit market equilibrium in a dynamic setting. I apply the model to evaluate the US Acid Rain Program, a federal cap-and-trade program designed to reduce sulfur dioxide emissions from fossil-fuel power plants. Through counterfactual simulations based on the estimated model, I examine the consequences of the Acid Rain Program with an emphasis on the dynamic aspects of firms'

[1]A recent exception is Weyl and Zhang (2018), who study the tradeoff between allocative efficiency and investment incentives in the property right design.

[2]Cap-and-trade regulation is also referred to as the emissions trading program. I use the term "cap-and-trade" throughout the paper.

behavior and regulatory design.

In a cap-and-trade program, regulated firms must surrender emissions permits to offset their emissions. To meet this regulatory requirement, firms face a "make-or-buy" decision problem: reduce emissions or trade (buy) permits. Dynamic incentives play a role in both decisions. Investment in clean, but costly technology is an important margin for reducing emissions. The trading of emissions permits is also a forward-looking decision because firms can save (bank) emissions permits across periods.

Motivated by these considerations, I construct a dynamic equilibrium model in which price-taking firms make decisions regarding the abatement of emissions and the trading (and banking) of emissions permits. Equilibrium permit prices are determined by market clearing conditions, which also determine how firms comply with the regulation in an equilibrium. In addition, the model incorporates two important factors that determine how the market mechanism affects the welfare effects of a cap-and-trade program.

First, the model allows for rich firm heterogeneity in terms of abatement costs and the initial allocation of emissions permits. Given the firm heterogeneity in abatement cost, the initial allocation of permits may not be the cost-effective allocation, i.e., the outcome under which the total abatement cost is minimized to achieve a given level of emissions. In such a scenario, the trading of permits might lead to a more cost-effective distribution of emissions.[3] Thus, it is crucial to incorporate firm heterogeneity when evaluating cap-and-trade. However, allowing for heterogeneity in a dynamic equilibrium framework can be computationally prohibitive. To circumvent this issue, I introduce a tractable framework for estimation and policy simulations. In particular, following the literature on the estimation of dynamic structural models (e.g., Aguirregabiria and Mira, 2010), my estimation approach avoids the computation of dynamic competitive equilibrium in estimation.

Secondly, I incorporate the transaction costs of permit trading as a wedge in the permit market. Given that no centralized trading exchanges exist for many cap-and-trade programs, how well the permit market works is an empirical question. Transaction costs capture the wedge between the market price and the firm-level shadow value of permits, which affects abatement and trading decisions. The transaction costs also play an important role in modeling dynamics. A theoretical study by Rubin (1996) shows that, without any frictions in the permit market, all firms should have the same shadow value of permits that coincides with the market price. Furthermore, the equilibrium permit price should increase in line with the interest rate, which is known as the Hotelling rule. These predictions are relatively strong in an empirical analysis. By including transaction costs, the shadow value of emissions permits is determined by trading and banking decisions and is thus no longer constant across firms.

Stavins (1995) is the first theoretical study to investigate how transaction costs discourage

[3] I also discuss the implication of trading in terms of the heterogeneity in environmental damages. This point is particularly relevant to the SO_2 emissions, which are known as non-uniformly mixed pollutant.

permit trading and thereby lead to inefficient outcomes in a static setting. Concerns related to transaction costs have been noted in practice. Previous studies document that many firms tend not to trade emissions permits, and instead choose to comply with the regulation using their allocated permits (see, e.g., Jaraitė-Kažukauskė and Kažukauskas, 2015 for the EU Emissions Trading Scheme). In my empirical analysis, I introduce two types of transaction costs: (1) a sunk cost associated with participation (entry) in the permit market, and (2) variable costs that depend on the trading volume. I argue that these costs can be identified from a firm's optimal decisions and estimate these costs in my empirical analysis.

I apply my empirical framework to study the first nine years (1995–2003) of the US Acid Rain Program, a cap-and-trade program for regulating sulfur dioxide (SO_2) emissions in the US electricity industry.[4] The aim of the Acid Rain Program is to reduce the aggregate SO_2 emissions from coal power plants to half of their 1980 levels. The regulator distributed emissions permits to existing generation facilities, and these facilities were required to surrender sufficient permits to offset their emissions each year.[5] Regulated sources could choose how to comply with the regulation. For example, they could switch to cleaner coal, invest in abatement equipment, or obtain additional permits from the market. Rich data on production and abatement by power plants and the trading of emissions permits are available from this program.

The Acid Rain Program is an interesting example of a cap-and-trade program in which dynamic incentives play an important role in compliance decisions. While the regulation started in 1995, the US EPA (i.e., regulator) announced the permit allocation schedule in 1990. The allocation is generous in the first five years of the regulation (1995–1999, referred to as Phase I) but then decreases by almost half in the period after 2000 (Phase II). Casual observation suggests that firms took this schedule into account. Specifically, regulated firms saved a significant amount of permits in the first five years, and then started using them once the cap became tighter after 2000. While this observation implies the importance of banking in firms' compliance strategies, some are concerned about the excessive banking of permits (e.g., Smith, Platt, and Ellerman, 1998). In addition, the environmental and health consequences of permit banking are ambiguous and have therefore proved controversial (e.g., Burtraw and Mansur (1999), Burtraw (2000), Ellerman, Joskow, and Harrison Jr (2003)). My framework analyzes how such dynamic incentives affect firms' compliance decisions and the welfare consequences of the cap-and-trade program.

An econometric analysis poses a challenge in terms of computation. My model belongs to a class of dynamic competitive equilibrium models with multiple heterogeneous firms. A full solution approach (i.e., solving a dynamic competitive equilibrium for each evaluation

[4]I choose the terminal period of my analysis to be 2003 because of the proposal of the Clean Air Interstate Rule in December 2003, which had a major impact on the regulatory environment for SO_2 emissions. See Section 2.2 for details.

[5]Emissions permits are called emissions "allowances" in the Acid Rain Program because the term "permit" has another meaning in US environmental law. Because "permit" is the standard terminology in the economics literature, I use the term "permit" in this paper.

of the model parameters) is computationally prohibitive. To circumvent the computation costs, I use the observed permit prices as a sequence of equilibrium prices, instead of solving equilibrium permit prices in the evaluation of the objective function. This trick is similar in spirit to the two-step estimators in dynamic Markov games (e.g., Aguirregabiria and Mira, 2007). Because each heterogeneous firm faces a different optimization problem, this approach significantly reduces the computation costs.[6]

In estimation, I first use rich data on coal purchases at the plant level to construct the hedonic function of fuel price, thus providing the marginal abatement cost via the purchase of cleaner coal. I also use the detailed information on production and the engineering estimate of the scrubber, allowing me to obtain the per-period cost function and the investment cost. These primitives are embedded into the dynamic decision problem to estimate other model primitives, including the transaction cost function. The estimates imply that the variable transaction costs from permit trading are substantial. The median of the marginal transaction cost is estimated to be \$32.1, whereas the permit prices range between \$100 and \$200 in the sample period. This result suggests that the dispersion of the shadow value of emissions across firms is large and thus the distribution of emissions may not be efficient.

Using the estimated model, I conduct counterfactual simulations to investigate the welfare consequences of the regulation. The welfare outcomes that I focus on include abatement costs (i.e., fuel costs and investment costs) and health and environmental damages. While the abatement cost serves as a primary measure in the evaluation of the cap-and-trade in the literature, the health and environmental damages are particularly relevant in this context. First, SO_2 emissions are known as non-uniformly mixed pollution; health and environmental damages depend on the location of the source of emissions. Thus, the geographic distribution of emissions matters in terms of welfare. In addition, permit banking has an ambiguous influence on health and environmental outcomes. Permit banking leads to earlier abatement and thus greater benefits in the early period, though emissions would be larger in later periods. My framework can analyze the net-benefit of the policy. To quantify the net-benefit, I use estimates of health and environmental damages from the APEEP model provided by Muller and Mendelsohn (2009b).

I conducted three counterfactual simulations. In my first counterfactual experiment, I examine the effect of eliminating all transaction costs. This simulation quantifies the outcome with the lowest abatement costs under the cap-and-trade. While previous works, including Carlson, Burtraw, Cropper, and Palmer (2000) and Gollop and Roberts (1985), conduct a similar exercise using a static framework, this paper analyzes the implications of transaction costs on dynamic decisions such as abatement investment and permit banking.[7] I find that

[6]The two-step approach used to estimate single-agent dynamic models (e.g., Hotz and Miller 1993; Aguirregabiria and Mira 2002) is not suitable in my setting because firms are heterogeneous in many dimensions, such as permit allocation, characteristics of power plants, and fuel costs. Given that the optimal decisions depend on these factors, estimating policy functions from the data in a flexible way is quite difficult due to the curse of dimensionality.

[7]Carlson, Burtraw, Cropper, and Palmer (2000) and Gollop and Roberts (1985) quantify the abatement

eliminating the transaction costs would lead to a more dispersed distribution of emissions, reflecting more active trading of emissions permits. The total production and abatement costs would decrease by $814 million in total in the cost-effective outcome, implying that "unrealized" gains from trade are significant in the sample period.

The health and environmental damage, meanwhile, decreases by $2.4 billion without transaction costs. This decrease is due to the change in the inter-temporal and cross-sectional (i.e., geographic) distribution of emissions. The former refers to the shift in the timing of emissions. Without the transaction costs, firms bank more permits in Phase I and use them in Phase II, implying the lower (higher) emissions in Phase I (Phase II). Overall, the discounted value of the total damages decreases. In addition, the geographic distribution of SO_2 emissions would change in the absence of transaction costs. Since the damage from SO_2 emissions differs across locations (Muller & Mendelsohn, 2009b), the total damage would also change. The primary cause of the greater aggregate damage is the change in the geographic distribution of emissions (i.e., emissions decrease in the region where the damage is greater).

In the second counterfactual simulation, I investigate the impacts of permit banking. I simulate the equilibrium outcome of cap-and-trade when permit banking between Phase I and II is not allowed while fixing the transaction cost as estimated. I find that permit banking reduces the total production and abatement cost by $325 million. These results suggest that permit banking mitigates the negative effect of transaction costs on abatement costs. In terms of the benefits of emissions abatement, the total damage of SO_2 emissions decreases because the total emissions in Phase I falls, leading to the lower discounted value of health and environmental damages.

Lastly, I simulate emissions tax policy as an alternative to the cap-and-trade program. I set the constant emissions tax rate that achieves the same aggregate amount of emissions. Interestingly, the cost-effectiveness of the simple emissions tax policy is relatively close to that of the cap-and-trade without transaction costs. Under the emissions tax policy, all firms face the same shadow value of emissions in the cross-sectional sense, which is equivalent to eliminating the transaction costs in the cap-and-trade. The difference between these two outcomes is due to the inter-temporal smoothing of abatement costs. This finding may seem to suggest that the contribution of permit banking to cost-effectiveness is limited. However, as I discussed in the previous paragraph, the permit banking improves the cost-effectiveness significantly when transaction costs exist. My interpretation of these findings is that the permit banking system can be particularly effective in mitigating the negative impact of transaction costs in cap-and-trade regulation.

The paper proceeds as follows. I begin by briefly reviewing the related literature. Section 2 then outlines the institutional background of the Acid Rain Program and provides a descriptive analysis of the data. Motivated by the descriptive findings, I introduce the

pattern when the marginal abatement costs of emissions are equalized across coal power plants. Without transaction costs, all firms would face the same shadow value of emissions given by the market price of permits, and thus the marginal abatement costs should be equalized across firms.

model in Section 3. I then present the estimation methodologies and the estimates of model primitives in Section 4. Section 5 presents the counterfactual experiments, through which I evaluate the consequences of the Acid Rain Program. Section 6 discusses several caveats and possible extensions of the paper, and, finally, Section 7 provides a conclusion.

Related Literature The study is related to three strands of literature. First, it contributes to the empirical literature on dynamic structural models in industrial organization (see, e.g., Bajari, Benkard, and Levin, 2007, Ryan, 2012, Collard-Wexler, 2013, and Kalouptsidi, 2014 in an oligopolistic setting, and Rust, 1987 and Aguirregabiria and Mira, 2002 in a single-agent setting). The empirical setting employed here is unique in that the investment in technology is substitutable with other production inputs, namely emissions permits. With frictions in the permit market, firms' decisions are cast as a make-or-buy problem in a dynamic setting. Although my model is tailored to cap-and-trade regulation, it can be used to analyze firms' dynamic incentives when they are subject to frictions or imperfections in input markets. Methodologically, my paper relates to the structural estimation of dynamic equilibrium models in which aggregate outcomes (e.g., permit prices in my application) are endogenously determined in equilibrium (see, e.g, Lee, 2005; Lee and Wolpin, 2006; Gillingham, Iskhakov, Munk-Nielsen, Rust, and Schjerning, 2015, 2019).

Second, my study is related to the empirical literature on cap-and-trade programs. While much of this literature tests the qualitative predictions of models of permit trading, a few recent works adopt a structural approach to measure the welfare implications of cap-and-trade programs (Fowlie 2010b, Ryan 2012, Fowlie, Reguant, and Ryan 2016, and Dardati 2016).[8] A distinctive feature of my paper is to model trading and banking decisions in a nonstationary dynamic equilibrium framework. Existing studies assume frictionless permit markets and a stationary regulatory setting. In such a setting, cap-and-trade is equivalent to imposing a Pigouvian tax. My model describes how firms make decisions regarding investment, trading, and banking when transaction costs exist and the regulatory standard (i.e., the permit allocation) changes over time. My framework can be used to study how the regulatory design of permit trading, such as the availability of permit banking and alternative allocation rules for emissions permits, affects firms' abatement decisions.

In contemporaneous work, Chen (2018) structurally estimates firms' beliefs regarding future permit prices using a single-agent dynamic model of emissions abatement and permit trading in the context of the Acid Rain Program. Although my paper and that of Chen (2018) are similar in terms of their empirical setting and modeling approach, they address different research questions and consequently take a different approach to model permit prices. On the one hand, Chen (2018) estimates flexible beliefs regarding future permit prices without imposing any equilibrium restrictions. The framework can demonstrate the

[8]The literature has examined the independence of outcomes from the initial allocation (Reguant and Ellerman, 2008 and Fowlie and Perloff, 2013) and the internalization of emissions costs (Kolstad and Wolak, 2008, Fowlie, 2010a, and Fabra and Reguant, 2014).

discrepancy between the estimated beliefs and the one implied by rational expectations. On the other hand, my study provides a dynamic equilibrium framework in which permit prices are determined endogenously. The framework can conduct various policy simulations to evaluate the program and counterfactual regulatory designs, which is the primary purpose of my paper.

Finally, my study provides new insights for the evaluation of the Acid Rain Program by studying the inter-temporal aspects of abatement decisions and the regulatory design. One approach adopted in the literature is to calculate the cost saving that results from permit trading by estimating a cost function and a discrete choice model for abatement choices (see, e.g., Ellerman, Joskow, Schmalensee, Montero, and Bailey, 2000, Carlson, Burtraw, Cropper, and Palmer, 2000, Keohane, 2006, and Chan, 2015). Researchers found that adopting a permit trading program led to significant cost savings as compared with traditional command-and-control approaches, although the actual cost did not reach the least-cost solution. Another approach analyzes aggregate variables to discuss the efficiency of the permit market (Joskow, Schmalensee, and Bailey 1998, Helfand, Moore, and Liu 2006, and Ellerman and Montero 2007).[9] This study complements previous research by empirically examining the dynamic aspects of compliance and abatement decisions under a cap-and-trade program. In particular, this study is one of the first to quantify the role of the permit banking system.[10] For this purpose, I construct and estimate an equilibrium model of the cap-and-trade program that enables me to simulate the outcome when permit banking is not allowed.

2 Empirical Setting and Descriptive Analysis

2.1 The Acid Rain Program

Fossil-fuel electricity plants, especially coal-fired plants, produce sulfur dioxide (SO_2) emissions as a byproduct of electricity generation. SO_2 is known to have detrimental effects on human health and the environment. Although the federal government introduced command-and-control-type regulations in the Clean Air Act Amendments of 1970, such regulations have not been effective in reducing SO_2 emissions.[11] The failure of the previous regulations led to the introduction of the Acid Rain Program (ARP) as part of Title IV of the 1990 Clean Air

[9] Joskow, Schmalensee, and Bailey (1998) find that prices in the spot market and the EPA auction are very similar, concluding that "a relatively efficient private market" had developed by mid-1994. Helfand, Moore, and Liu (2006) use monthly permit prices for the period 1994 to 2003 to test whether the price path follows the Hotelling $r-$percent rule for inter-temporal arbitrage. They reject the Hotelling rule, which suggests there is inefficiency in the market. Ellerman and Montero (2007) argue for an efficient market of permits by comparing the actual and theoretically predicted volume of aggregate banking.

[10] Previous studies (e.g., Ellerman, Joskow, Schmalensee, Montero, and Bailey, 2000) note the importance of permit banking as a source of cost-effectiveness in cap-and-trade. Arimura (2001) conducts numerical simulation to quantify the cost savings brought about by permit banking in a frictionless setting.

[11] Ellerman, Joskow, Schmalensee, Montero, and Bailey (2000) provide a brief history of the regulation on SO_2 emissions.

Act Amendments.

The ARP is a cap-and-trade program for SO_2 emissions that began in 1995. The aim of the regulation is to regulate SO_2 emissions from electricity generating units (EGUs) that use fossil fuels and have an output capacity greater than 25 megawatts. The regulation was implemented in two phases. In Phase I (1995–1999), a subset of eligible EGUs fell under the regulation. These units included 263 EGUs, called the "Table 1" group, that were notably dirty (i.e., they produced a large amount of emissions) and old before the regulation, as well as an additional 182 EGUs from the Non-Table 1 group to serve as substitution or compensating units. In Phase II (begun in 2000), all eligible EGUs were mandated to comply with the regulation.

The ARP aims to reduce SO_2 emissions from generation facilities to half of their 1980 levels, based on which the total number of emissions permits is determined each year. Most emissions permits are allocated for free to existing units. The EPA adopts a rule that determines the unit-level allocation of emissions permits based on the characteristics of a unit.[12] The allocation is determined primarily by the average heat input during the period 1985–1987 and the target emissions rate of fuel (i.e., emissions per fuel input) for each phase. Specifically, the target emissions rate of fuel for Phase I is 2.5 pounds (lbs) per 1 million British thermal unit (MMBtu), and the rate for Phase II is 1.2 lb/MMBtu. Some units are allocated additional permits based on technical and political considerations (Joskow & Schmalensee, 1998). It is worth emphasizing that the provisions in the 1990 legislation include detailed rules for permit allocations. Thus, generation facilities were aware of the schedule of permit allocation before the program started in 1995.

Under the ARP, emissions permits are tradable among participants. Firms can sell or buy permits with other firms, including financial companies or brokers that do not own any generating units. Although the EPA also holds an annual auction to distribute around 2.7% of the yearly allocation, a centralized trading exchange does not exist. Bilateral trading, which is often mediated by brokers, is the primary means to trade emissions permits with other participants.

The operation of each regulated unit, especially emissions levels of SO_2, is recorded through the Continuous Emissions Monitoring System.[13] At the end of each calendar year, the annual level of SO_2 emissions is finalized, and each regulated unit is required to surrender emissions permits within a grace period of 60 days.[14] One unit of emissions permit is needed to offset one ton of SO_2 emissions. The remaining permits are carried over to the next year,

[12]The unit-level allocation depends only on past information, and there is no update on the permit allocation based on actual output or emissions. See U.S. Environmental Protection Agency (1993b, 1993a) for the details.

[13]There should be no concern about manipulating the measurement of emissions because the operators are required to perform periodic performance evaluations of the monitoring system. See U.S. Environmental Protection Agency (2009) for the details.

[14]If an affected unit does not hold sufficient permits to offset the emissions at the end of the compliance deadline, unit operators are required to pay a penalty of USD 2000 per SO_2 ton. However, compliance was nearly 100% during the period of my analysis.

referred to as the banking of emissions permits. There is no expiration date for banked permits. As I discuss in Section 2.3, regulated firms banked a significant number of permits in Phase I, when the annual allocation was more generous than it was in Phase II.

Although existing units obtained their initial allocation of permits with no charge, most of them still needed to reduce their emissions in order to comply with the regulation. The regulated units were able to reduce emissions by reducing their utilization (output) or reducing their emissions per input (emissions rate of fuel).[15] The latter option of reducing the emissions rate of fuel was the primary margin of abatement, as I will explain in Section 2.3.2.

2.2 Data Sources

This subsection describes the data sources used for this study. I combine the transaction data of emissions permits and various data on electricity generation and emissions abatement. The study focuses on the period between 1995 and 2003. Although the ARP continued after 2004, the proposal of the Clean Air Interstate Rule, announced in December 2003, had a large impact on regulated firms' expectations over the future regulatory environment.[16] The proposed regulation aimed to strengthen the stringency of the SO_2 regulations from 2010 within the framework of the ARP. After the announcement, the permit price started to rise dramatically, primarily because the value of emissions permits issued before 2010 would be higher than those issued after 2010, according to the proposed regulation.[17] Firms also started to invest in scrubbers in anticipation of a stricter regulation.[18] Thus, I do not include data after 2004, focusing instead on those periods when the regulatory environment for SO_2 emissions was stable.[19]

The data on permit transactions are taken from the Allowance Tracking System (ATS) provided by U.S. Environmental Protection Agency (1993–2013a). The EPA operates the ATS to manage permit allocations and to track private transactions and the surrendering of permits for compliance. The ATS data are available to the public. Each transaction record in the tracking system contains the account names of a transferor and a transferee, vintage of permits, quantity of transferred permits, and confirmation date of the transaction.[20] I

[15] The ratio of output to input is the design parameter for generating units. Therefore, firms are not able to increase this aspect (i.e., improve fuel efficiency) to reduce emissions.

[16] The announcement was made by the EPA on December 17, 2013. The newly proposed rule was referred to as the Interstate Air Quality Rule (U.S. Environmental Protection Agency, 2003).

[17] See Federal Register (2004) for the details of the Clean Air Interstate Rule.

[18] See Schmalensee and Stavins (2013) for a detailed review on how the regulatory environment for SO_2 emissions has changed since 2004.

[19] One might be concerned that the regulator (i.e., the EPA) is able to modify the regulation at will, leading to regulatory uncertainty. However, the EPA does not have the authority to modify the regulatory rules of the Acid Rain Program, such as tightening the overall cap or changing the permit allocation. To do these, new legislation would need to be passed by the Congress.

[20] The confirmation date must lag behind the actual transaction date to some extent, although the prompt recording of private trading was considered the rule rather than the exception, according to EPA staff and industry experts. See Joskow, Schmalensee, and Bailey 1998 for details.

constructed the transaction data at the firm and year levels from the database.[21] Specifically, I aggregated the account-level data into firm-level data using ownership information constructed by various sources, including the eGrid database (U.S. Environmental Protection Agency, 1996-2010) and EIA-860 (Energy Information Administration, 1990-2013). The final data set includes (1) permit holding at the beginning of the year, (2) annual allocation, (3) volume of permit trading (net purchase of emissions permits), and (4) banking volume.[22]

The ATS does not collect information on transaction prices. I therefore use the market-price index of SO_2 permits provided by Cantor Fitzgerald, one of the major brokers in the SO_2 permit market (BGC Environmental Brokerage Services, 2011).[23] The frequency of the price data is monthly.

Regarding production and emissions information, I compiled data from various sources including the EPA and the US Energy Information Administration (EIA). The EPA makes publicly available the unit-level operation data of the generating units, collected by the Continuous Emissions Monitoring System (CEMS). The CEMS data include gross generation (in MWh), heat input (in MMBtu), and SO_2 emissions (U.S. Environmental Protection Agency (1993–2013b)). In addition, the EIA conducts various surveys on the operation of power plants. Specifically, the Form EIA-767 "Steam-Electric Plant Operation and Design Report" (Energy Information Administration (1985–2005)) provides information on fuel usage (sulfur content, ash content, heat inputs), net generation, and generation capacity at the unit and monthly level. In addition, the Form FERC No. 423 (EIA-423) "Monthly Report of Cost and Quality of Fuels for Electric Plants" (Energy Information Administration (1990–2003)) provides plant-level and monthly-level information on fuel procurement, including fuel type, sulfur content, heat content, and purchase costs.

2.3 Descriptive Analysis

In this subsection, I provide a descriptive analysis of the data. I highlight various aspects of the ARP, including the banking of emissions permits, the abatement behavior of regulated sources, and the market for emissions permits. These descriptive findings provide the motivation for the modeling approach introduced in Section 3.

[21] Note that permit transactions between power plants (or generating units) within the same firm are considered to be reallocation of permits within a firm. The trading of emissions permits is defined as a transaction with another firm or broker.

[22] Emissions allowances issued under the Acid Rain Program have a vintage, that is, the year that the allowance was issued. Firms can use a permit with a vintage that is either current or older (i.e., permit banking) for compliance purposes. In principle, firms can trade the emissions permits of future vintages, although the trading volume for such permits is relatively small. Therefore, I focus on the trading of permits with a current or old vintage and construct a data set from these transactions.

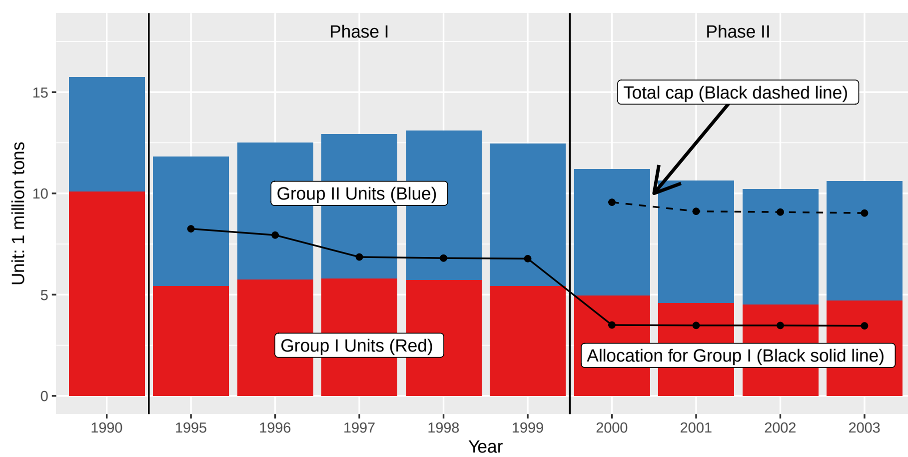
[23] Cantor Fitzgerald was acquired by BGC Partners, L.P. in 2011 and became BGC Environmental Brokerage Services, L.P.

2.3.1 Banking of Emissions Permits

Figure 1 shows the aggregate SO_2 emissions level and emissions caps under the ARP from 1990 to 2003. The bars indicate the aggregate emissions each year, while the dashed lines represent the emission caps, which are equal to the total amount of emissions permits allocated by the EPA in that year. As mentioned in Section 2, the timing of the regulation varied across EGUs. I denote those units regulated since 1995 as Group I units, and those regulated since 2000 as Group II units. The blue bar in the figure corresponds to the emissions of Group I units, and the orange bar corresponds to those of Group II units. The blue dashed line shows the allocation for Group I units, and the black dashed line from 2000 shows the total cap of emissions, including both Group I and II units.

The figure demonstrates that once Phase I started in 1995, Group I units reduced their emissions by almost half compared with their 1990 levels. The emissions before 1999 were significantly lower than the emissions cap, implying that those units bank their permits. While both Group I and II units reduced their emissions further in 2000 (i.e., the first year of Phase II), Group I units did not reduce their emissions to the same extent as they did in 1995. They began using the banked permits to ensure compliance. Given that the regulator had already announced the allocation schedule in 1990, Figure 1 suggests that regulated firms behaved in a forward-looking manner. This observation motivates the use of a dynamic structural model in my analysis.

Figure 1: Aggregate Volume of SO_2 Emissions and Caps (1990–2003)



Notes: The blue (orange) bar corresponds to emissions from Group I (Group II) units. The blue dashed line indicates the permit allocation for Group I units, while the black dashed line (from 2000) shows the total cap, including the allocations for both Group I and II units.

2.3.2 Abatement Strategy for Coal Units

Emissions from electricity generation can be reduced either by (1) reducing output or by (2) reducing the emissions rate of fuel (i.e., emissions per input).[24] However, the former strategy was not a major option for coal units regulated under the Acid Rain Program. To demonstrate this, in Appendix A.1, I estimate the effect of the regulation on the output (utilization rate) of EGUs in a difference-in-differences (DID) framework by exploiting the variation of the timing of the regulation across units. I found that the utilization rate decreased by only 0.6–3.8 percentage points after the introduction of the Acid Rain Program. In addition, I examined whether the regulated firms retired coal units, a potential option for emissions abatement. The data show that this margin is small. Among the 263 EGUs in the "Table 1" group, only seven units retired before 1995, and two additional units retired before 2003. Of the other coal units, around 6% of EGUs retired between 1990 and 2003. In this subsection, I explain the abatement strategy of adjusting the emissions rate of fuel of EGUs whose primary fuel type is coal.[25]

Two common options are available to reduce the emissions rate of fuel for coal units. The first is called fuel switching. An operator of coal units can switch the type of coal from dirty (e.g., high-sulfur bituminous coal) to cleaner (e.g., subbituminous coal or low-sulfur bituminous coal). The fuel costs of cleaner coal are higher than those of dirty coal. In addition, switching fuel types requires retrofitting the boiler to make it compatible with the new type of coal, which incurs a fixed cost. Another abatement option is to install flue-gas desulfurization equipment (a scrubber). This equipment is installed at the stack of a generation unit and eliminates more than 80% of SO_2 emissions. However, this option incurs a large investment cost, as well as a long lead time (two to three years, on average).

Figure 2 shows the distribution of the unit-level SO_2 emissions rate of fuel (measured in pounds per MMBtu) for each group. The left panel shows the distribution for Group I sources. The emissions rates of fuel for these sources fell between 1990 and 1995, the beginning of Phase I. The rates then stayed almost constant during Phase I, before falling further in 1999 in anticipation of the beginning of Phase II. The emissions rates of fuel for generating units in Group II did not change until 1999, but then decreased in 2000, the first year of the cap-and-trade program for these units. These observations imply that firms adjusted their emissions rate of fuel at the beginning of each phase, but that the rate then remained almost constant within the phase.

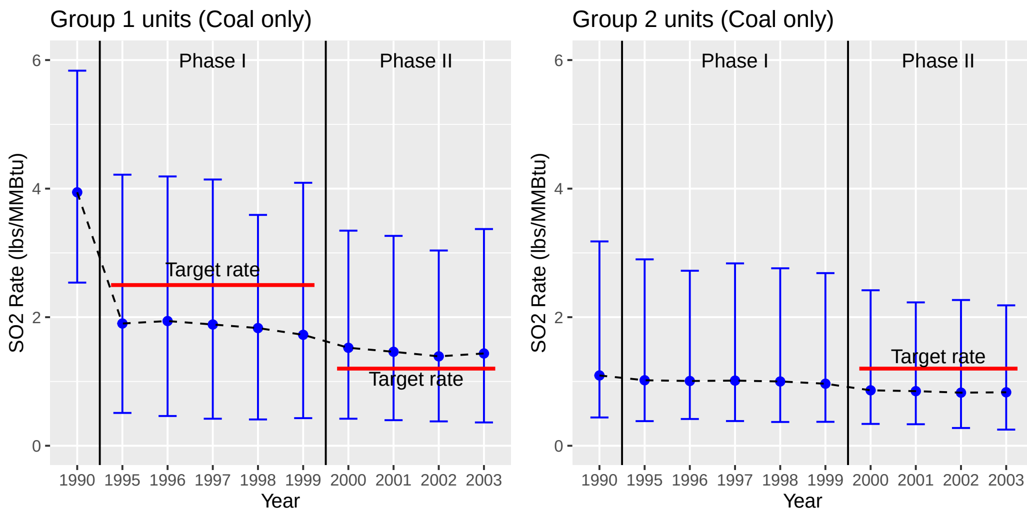
Another important finding to be gleaned from Figure 2 is the flexibility of the compliance patterns. The red horizontal lines indicate the target emissions rates of fuel in each phase. Conditional on the level of heat input (i.e., the average heat input during the period 1985–1987) that is used to calculate the initial allocation of permits, generating units would need

[24] The ratio of output to input is a fixed design parameter of generating units. Thus, firms cannot improve their fuel efficiency as a way to reduce emissions.

[25] Although the target of the ARP includes all types of fossil fuel units (coal, gas, and oil), SO_2 emissions from gas and oil units are quite small. In the analysis I treat these emissions as exogenous.

to achieve this target rate if they did not trade emissions permits with other units. The figure indicates that some units achieved a greater emissions reduction than necessary, while others did not reduce their emissions rates and therefore needed to secure additional permits. This implies that the trading of permits played an important role in compliance decisions.

Figure 2: Distribution of Unit-level SO_2 Emissions Rate of Fuel



Notes: The blue dots show the weighted average of emissions rates of fuel. The upper (lower) bars correspond to the 10th (90th) percentile of the distribution.

2.3.3 Heterogeneity of Regulated Firms

The heterogeneity of regulated firms is a key factor in the evaluation of a cap-and-trade program, as this is the source of the gains from trade: firms with higher (lower) costs of abatement can buy (sell) emission permits by trading with other firms. As a result, the pattern of emissions is more cost-effective than that in an autarky, where no emissions permits are traded.[26]

Table 1 shows the descriptive statistics for the characteristics of the regulated firms. The table shows that regulated firms differ substantially in terms of firm size (measured by the number of regulated units and the total capacity of those units), the share of generating units with a scrubber. These factors affect the firms' abatement and trading decisions. For example, plants that did not have a scrubber needed to make more effort to comply, either by reducing their emissions or by buying permits. Firms with a higher initial allocation, conditional on other factors being fixed, are more likely to be sellers of permits in the market. The model introduced in Section 3 incorporates the observed heterogeneity across firms.

[26]Given the heterogeneous environmental damages caused by SO_2 emissions (Muller and Mendelsohn, 2009b), the implication of permit trading on the net-benefit of the program is an empirical question. I investigate this point in simulation analysis in Section 5 .

Table 1: Firm Heterogeneity

	N	Mean	St. Dev.	25 Percentile	Median	75 Percentile
# of Coal units in Group 1	138	2.43	4.76	0	0	3
# of Coal units in Group 2	138	3.86	5.38	1	2	5
# of Gas and Oil units	138	5.40	7.44	0	2.5	7.8
Total Capacity of Coal units in Group 1	138	701.72	1,614.71	0	0	800
Total Capacity of Coal units in Group 2	138	1,124.80	1,616.05	51.8	515.5	1,550.8
Share of Scrubbed Units in Group 1 in 1990	50	0.07	0.20	0.00	0.00	0.00
Share of Scrubbed Units in Group 2 in 1990	119	0.27	0.39	0.00	0.00	0.51
Firm size	138	2,295.60	2,882.71	400.8	1,242	3,358.2
Firm size: Owning Group 1	50	3,111.69	3,313.77	1,039.25	2,117.50	3,738.50
Firm size: Owning only Group 2	88	1,831.91	2,510.17	97.00	663.00	2,894.00
Initial Allocation	138	56,054.51	85,132.48	8,049.67	28,147.62	66,920.44

Notes: There are 138 firms in the sample. The unit for generation capacity is megawatts. The unit for initial allocation is one emissions permit, which is required to offset one ton of SO_2 emissions.

2.3.4 Firm-level Trading Information

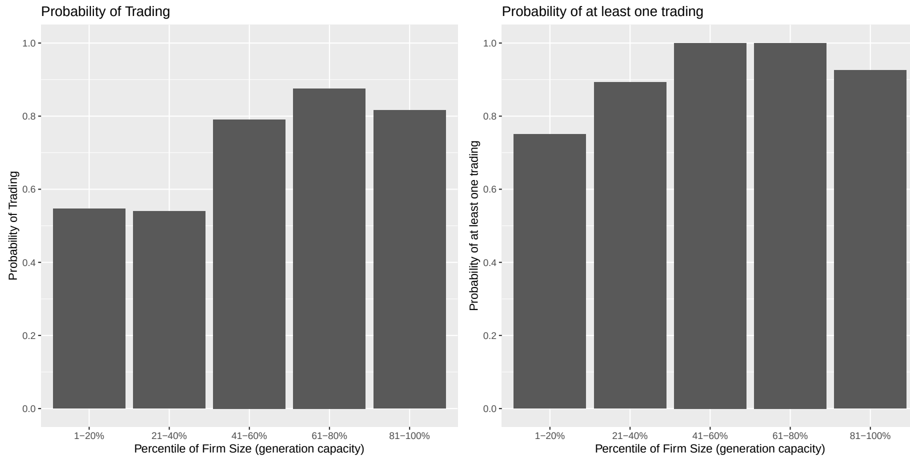
This subsection investigates the trading pattern of emissions permits. The U.S. Environmental Protection Agency (2004) reports that transactions of emissions permits between related entities (i.e., power plants and generating facilities under the same ownership) have been active since the beginning of the program. Therefore, I focus on trading with other firms (e.g., other affected firms and financial brokers) in the market.

Figure 3 reports the trading pattern at the firm-level. Specifically, the left panel shows the unconditional probability of a market transaction at the firm-year level, and the right panel shows the trading experience in the sample period at the firm level. Overall, the panels implies the positive correlation between the trading decisions in the permit market and firm size, measured by the sum of the nameplate capacity of units under the ARP.

Furthermore, the left panel shows that firms did not necessarily trade every year. The unconditional probability of conducting permit trading was 73%. The trading probability was positively correlated with firm size. This observation is also found in the context of the EU–ETS scheme (see, e.g., Jaraitė-Kažukauskė and Kažukauskas, 2015). Although this finding can be interpreted as suggestive evidence of fixed transaction costs, it should be noted that firms do not need to conduct a transaction in every period, as they can bank emissions permits. In the right panel, meanwhile, I show the firm-level experience of market trading during the sample period. As can be seen, 91% of firms had at least one experience of trading with another firm in the sample period, although some firms, most of which were small, did not trade at all.[27]

[27] In the Online Appendix, I have added figures regarding trading patterns. Figure A5 shows the aggregate volume of trading over time, while Figure A7 shows the frequency of trading normalized by the firm size.

Figure 3: Trading Pattern at Firm Level



2.3.5 Price of Emissions Permits

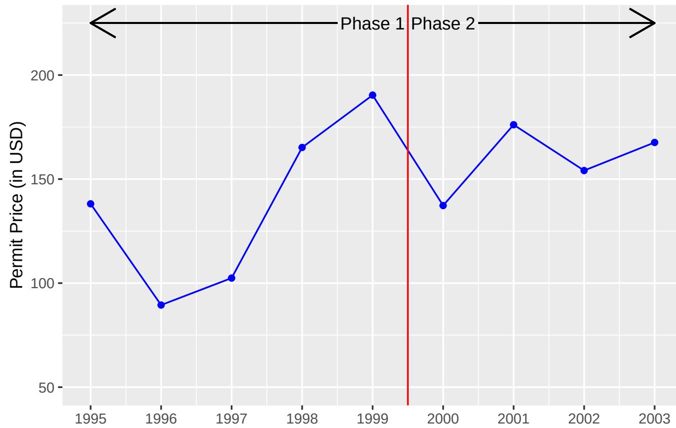
As discussed in Section 2.1, there are no centralized trading exchanges for emissions permits under the Acid Rain Program. Although regulated firms must conduct bilateral trade with other firms, brokers act as intermediaries for these transactions. Brokers also provide information about permit prices. Figure 4 shows the price information provided by Cantor Fitzgerald, a major broker in this market. I use the monthly SO_2 price index as a price measure. Cantor Fitzgerald constructs the index using various trading data, including allowance bids (to buy), allowance offers (to sell), and actual trade prices, and publishes it on the company website every month. I aggregate the monthly price index by taking the volume-weighted mean for each year.[28] Note that the price is normalized to the 2000 level using the producer price index (U.S. Bureau of Labor Statistics, 1984–2015).

The price at the beginning was around \$150, falling to below \$100 in 1996 and 1997. It then rose to \$200 in 1999, before fluctuating in the range \$120–\$200 after 2000. The figure suggests that the market price reflects the presence of permit banking. In the absence of permit banking, I would expect to see a spike in the permit price at the beginning of Phase II because the regulatory intensity in Phase II is much stricter. Such a patter is not found in Figure 4. [29]

[28] Figure A6 in the Online Appendix presents the comparison of the volume-weighted mean and (unweighted) median prices. These two measures are quite close to each other.

[29] A key theoretical prediction regarding permit prices is the Hotelling rule: Permit prices should increase with the risk-free interest rate if the market is efficient and there are no transaction costs. Helfand, Moore, and Liu (2006) test the Hotelling rule using monthly prices of emissions permits for the same period. They reject the rule after controlling for structural changes and market shocks.

Figure 4: Price of Emissions Permits by a Broker



Notes: Prices are normalized to January 2000 prices using the producer price index. The prices are the weighted mean across months in each year. The weight is the aggregate trading volume of permits.

3 Model

3.1 Overview of the Model

This section introduces a model of abatement and trading decisions and permit market equilibrium under the cap-and-trade program. The model incorporates the institutional features and descriptive findings in the previous section, including the non-stationary nature of the regulations (i.e., the changing permit allocations), dynamic decisions on permit banking and investment, frictions in permit trading, and firm heterogeneity.

The unit of decision maker is a firm (or equivalently a utility company in this context) indexed by $i = 1, \dots, N$. A firm owns multiple power plants, each of which may hold multiple generating units that are regulated by the Acid Rain Program. Firm i owns J_{it} units of the regulated generating units, each of which is indexed by j.

The model is set as a nonstationary and finite-horizon model. Each discrete decision period corresponds to one compliance year. Time is indexed by $t = 1995, \dots, 2003 (\equiv T)$. Firms have a common discount factor of $\beta \in (0, 1)$.

Figure 5 provides an overview of the firm-level decision problem. It has two building blocks: (i) investment in a scrubber at the beginning of each phase (1995 and 2000), and (ii) decisions regarding coal quality (emissions rate of fuel), permit trading, and permit banking in each year. Note that I do not model the entry/exit decision of generating units because the exit of coal generating units was limited in my sample period, as discussed at the beginning of Section 2.3.2.[30]

At the beginning of each phase (i.e., 1995 and 2000), a firm makes a decision regarding

[30]See also Section 3.8 for a further discussion.

scrubber investment that determines the removal rate of scrubber $\alpha_i^l \in [0,1]$ in Phase $l \in \{1,2\}$. The removal rate is assumed to be fixed within each phase. I will discuss this in greater detail in Section 3.5.

Given the scrubber investment in each phase, firm i makes decisions on coal quality, permit trading, and banking. The timeline of decisions in each period is as follows:

1. Firm i holds permits that are carried over from the previous period, denoted by h_{it}. A firm also receives an annual allocation of permits, denoted by a_{it}.

2. Participation decision: Denote firm i's experience of market trading by I_{it}; i.e., $I_{it} = 1$ if a firm has experience in market trading and 0 otherwise. If $I_{it} = 0$, a firm can pay the one-time sunk cost F_{it} to participate.

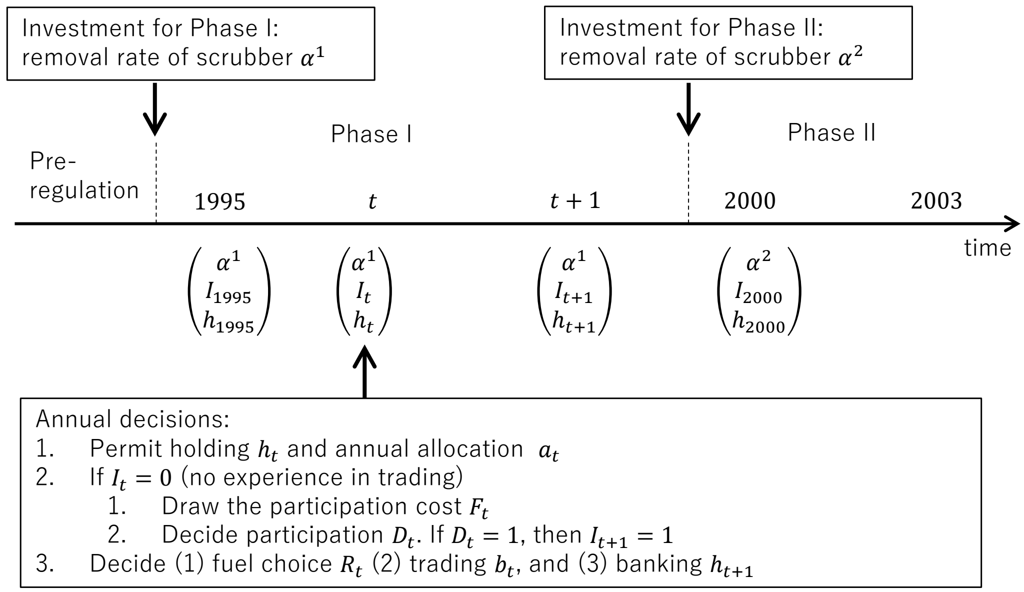
3. A firm chooses (1) the emissions rate of fuel R_{jt} for each generating unit $j = 1, \cdots, J_{it}$, (2) the net volume of trading b_{it} if a firm is already participating in the market, and (3) the banking of permits $h_{i,t+1}$. When determining the net-purchase of emissions permits b_{it}, each firm is a price-taker and treats the market price of emissions permits P_t as given.

4. A firm obtains profits from electricity generation and pays the costs of permits (or obtains the revenue from selling permits).

5. Move to the next period with permit holding $h_{i,t+1}$.

To formally define the optimization problem, I first specify in Section 3.2 the gross-profit from electricity production and associated SO_2 emissions. Section 3.3 introduces permit trading. These two components are incorporated into the optimization problem introduced in Section 3.4. To close the model, the equilibrium prices of the emissions permits P_t are determined by market clearing conditions in each period, which I introduce in Section 3.6.

Figure 5: Overview of the Firm-level Decision Problem



3.2 Electricity Production and SO_2 Emissions

This subsection introduces gross profit from electricity production and associated SO_2 emissions. Note that the gross profit defined here does not include the costs associated with permit trading. I formally define the optimization problem in Section 3.4.

Consider a firm that produces electricity and emits SO_2 as a byproduct. Firm i owns J_{it} units of the regulated generating units, each of which is indexed by j. I define the gross profit from electricity production by the negative of the total cost of generation. The gross profit π_{it} is thus given by

$$\pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) = -\sum_{j=1}^{J_{it}} p_{jt}^{fuel}(R_{jt}) \cdot HR_j \cdot q_{jt} \tag{3.1}$$

where $p_{jt}^{fuel}(R_{jt})$ denotes the unit-specific fuel price per 1 MMBtu of fuel input. The fuel price is a function of the emissions rate of fuel R_{jt}, the unit of which is pounds per 1 MMBtu (lbs per MMBtu). The unit-specific heat rate HR_j is the inverse of the production efficiency measure, which represents how much fuel input (in MMBtu) is needed to produce one unit of output (MWh of electricity). The heat rate HR_j is a design parameter of generating units and, therefore, is assumed to be exogenous. The unit-level production is denoted by q_{jt}.

The total cost is based on the fuel costs, which account for around 75% of the total operating expenses for fossil-fuel power plants (see EIA, 2012). The specification does not consider other types of costs such as the start-up costs of electricity generation. I will further discuss this point in Section 3.8.

Electricity generation is associated with SO_2 emissions. Firm-level emissions are given by

$$e_{it}\left(\{q_{jt}, R_{jt}, \alpha_{jt}\}_{j=1}^{J_{it}}\right) = \sum_{j=1}^{J_{it}} (1-\alpha_{jt}) R_{jt} \cdot HR_j \cdot q_{jt}, \tag{3.2}$$

where α_{jt} is the removal rate of the scrubber. α_{jt} can take a value between 0 and 1. The removal rate is endogenously determined by the investment decision.

As discussed in Section 2, firms can adjust the emissions rates of fuel for coal units. On the other hand, I assume that the emissions rate of fuel R_{jt} for gas and oil units are fixed and exogenous in the model. This is because gas and oil units have relatively low SO_2 emissions rate of fuel and are not able to reduce it.

Before moving to the structure of permit trading, I discuss two caveats in the model of electricity production.

3.2.1 Aggregation of Unit-Level Removal Rate of Scrubber at the Firm Level

To address the dimensionality issue in investment decision problem, I aggregate the unit-level removal rate of scrubber at the firm level. Specifically, I assume that each firm chooses a single removal rate of a scrubber that is common across coal units within the same firm (i.e., $\alpha_{jt} = \alpha_{it}$, for $j = 1, \cdots, J_{it}$). This assumption reduces the number of state variables related to the scrubber decision to just one.

I adopt this approach in order to maintain the tractability of the model. A key issue in modeling dynamic investment decisions is the dimensionality of the state space. If I model the investment decision at the generation unit level, the state variable should include the installation status of a scrubber at each generation unit. Because the average number of coal units in each firm in my sample is seven, such a modeling approach would be subject to the curse of dimensionality and therefore be intractable for the analysis.[31]

3.2.2 Exogenous Electricity Production

In the structural model of this paper, I assume exogenous electricity production; that is, unit-level electricity generation q_{jt} is fixed at the observed level. This assumption is motivated by the observation in Section 2.3.2 that the reduction of output is not a major margin of emissions abatement (see Appendix A.1 for more detail). This approach also follows previous studies including Carlson, Burtraw, Cropper, and Palmer (2000), Fowlie (2010b), and Chan (2015), in which the major margin of abatement is changing coal quality or adopting abatement technology. Yet, this assumption excludes reallocation of production across generating units (especially across coal and gas units) as a margin of emissions abatement. In Section 6.1, I

[31] This approach is, however, not completely innocuous. The firm would prefer to install a scrubber to the dirtiest plant, as it would be easier to abate emissions from those dirty units. Thus, the model would under-evaluate the marginal abatement costs under this assumption.

extensively argue this point by justifying the approach and discussing potential issues due to the assumption.

3.3 Permit Trading with Transaction Costs

A firm receives an annual allocation of permits a_{it} in each period. Since the allocation plan was announced before the regulation, the sequence of $\{a_{it}\}_{t=1995}^{2003}$ is exogenous in the model. The firm also holds the emissions permits that are carried over from the previous period, denoted by h_{it}. A firm chooses emissions level e_{it}, which is determined by production quantity and coal quality $\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}$ (see Equation (3.2)), net purchase volume b_{it}, and banking volume $h_{i,t+1}$. Net-purchase b_{it} is positive (negative) if firm i is a buyer (seller), implying that she is buying (selling) $|b_{it}|$ units of permits.

The transition of a permit holding is given by

$$e_{it} + h_{i,t+1} = a_{it} + h_{it} + b_{it}, \tag{3.3}$$

$$h_{i,t+1} \geq 0. \tag{3.4}$$

Equation (3.4) is the non-negativity constraint of banking, and it excludes the possibility of borrowing permits from a future allocation. I assume that firms achieve perfect compliance because the compliance rate under the Acid Rain Program was nearly perfect.

I model the permit market as a competitive market with transaction costs. The competitive market was motivated by the observation that the Acid Rain Program was a federal-level regulation in which many electric utilities and financial companies participated. Exercising market power in the permit market was limited.[32] Firms are thus price-takers in the permit market and face the market price P_t.

I incorporate the transaction cost to capture how well the permit market works in a reduced-form way.[33] Given that there were no centralized exchanges for emissions permits and the majority of permit transactions were bilateral, quantifying the degree of inefficiency in the permit market is an empirical question.[34] Furthermore, the transaction costs capture inefficiency that is not inherent to the market structure of emissions permits. Such inefficiency includes managerial frictions (e.g., lack of experience in the early periods of the program). Although I can consider different micro-foundations for transaction costs, I aim to be agnostic

[32] Liski and Montero (2011) examined how the four biggest electric utilities (in terms of initial allocation) traded in the permit market. They found that firms' behavior is not consistent with the model of market power in a storable commodity market.

[33] In the model, permit trading refers to the transaction of permits *across firms*. This model implicitly assumes that there are no costs associated with transactions among generating units *within the same firm*. This type of transaction is a reallocation of inputs within a firm and thus should not incur any transaction costs. As mentioned in Section 2.3.4, the trading of permits between related entities (reallocation) has been active since the implementation of the regulation (see, e.g., U.S. Environmental Protection Agency, 2004).

[34] Ideally, I should incorporate bilateral trading of emissions permits across participants into the model. However, such a model could prove to be substantially complicated and difficult to solve because emissions permits are divisible objects and, moreover, the model features dynamic investments in abatement technology and permit banking.

about their interpretations and instead focus on its implications for regulated firms and welfare outcomes.

The model incorporates two types of transaction costs. First, when a firm trades for the first time, it pays a sunk cost of participation F_{it}. This cost is motivated by the observation that some firms did not rely on permit trading in their compliance strategy. An interpretation of F_{it} includes the costs associated with setting up a trading desk at the company and hiring a financial trading expert. I specify F_{it} as $F_{it} = F + \sigma_F \epsilon_{it}$, where F is the mean participation cost, ϵ_{it} is an idiosyncratic cost shock that follows the type-I extreme value distribution $G(\cdot)$, and σ_F is the standard deviation parameter.

Second, firms pay variable transaction costs associated with the net purchase of permits b_{it} (Stavins, 1995). I define the variable transaction cost by $TC(|b_{it}|)$, which is a differentiable and strictly convex function.[35] The cost depends on the volume of trading given by $|b_{it}|$.[36] The convexity of the variable transaction costs is crucial to ensuring that the model is well-behaved. I will elaborate on this point in Section 3.7.1. Variable transaction costs might include monetary costs such as brokerage commissions and bid-ask spreads.[37] The convex nature of the cost function captures the difficulty of large-scale transactions of emissions permits. Suppose that a firm wants to buy a certain quantity of permits, but its trading partner cannot meet the demand. The firm would then need to find a different trading partner, which would entail a costly search process in a bilateral market.

The convex assumption of the variable transaction cost is also motivated by the finance literature, which has extensively studied frictions and transaction costs in the financial market. In the theoretical analysis, the convex transaction costs are often employed (e.g., Gârleanu and Pedersen, 2013, and Dávila and Parlatore, 2017). Such an assumption is motivated by empirical findings in the stock market (see, e.g., Breen, Hodrick, and Korajczyk, 2002, Lillo, Farmer, and Mantegna, 2003, and Robert, Robert, and Jeffrey, 2012). While the direct application of the findings from the stock market requires careful consideration, the trading structure of emissions permit is similar to a typical financial commodity. Brokers and financial institutions are involved in the transaction of emissions permits, and they often work as an intermediary between electricity companies regulated by the Acid Rain Program. Specifically, according to the transaction data of emissions permits, the fraction of emissions

[35] While I take the absolute value of b_{it} in the transaction cost function, the functional form of $TC(\cdot)$ I adopt in an empirical application assures the smoothness at $b = 0$. See Equation (4.2) in Section 4.2 for the details.

[36] The model assumes that variable transaction costs are dependent on the net volume of trading, $|b_{it}|$, rather than the gross volume of trading. The net volume may not align with the gross volume in cases where a firm simultaneously sells and buys permits within a given period (i.e., a year). To evaluate the plausibility of this assumption, I calculate the discrepancy between the net volume and the gross volume using the formula $\frac{(grossvolume)_{it} - |b_{it}|)}{(grossvolume)_{it}}$, for firm i in year t. The variable $(grossvolume)_{it}$ is constructed by summing the trading volume across transactions, which can involve either selling or buying permits, at the firm and year levels. This measure is equal to zero when a firm only sells or buys permits within a year. I find that 50.7% of the observations report a value of zero, suggesting that roughly 50% of the firms only sell or buy permits within a given period. The unconditional mean of this measure is 28.1%.

[37] In the model, both buyers and sellers pay the variable transaction costs $TC(|b|)$. In estimation, I allow the variable transaction cost to be different for buyers and sellers. See Equation (4.2) in Section 4.2.

permits traded by financial companies is approximately 34.5%.

Net-purchase by Fringe Firms The sample does not include all firms participating in permit trading. For example, financial companies or brokers do not have generation facilities, but they can still trade permits. In addition, some electricity companies were excluded from the sample in the process of data cleaning. I denote these firms as fringe firms in the permit market. To accommodate the presence of fringe firms, I introduce the net demand for permits by firms outside my sample; the total net purchase by fringe firms is denoted as $\bar{B}_t^{fringe}(P_t)$. I explain the specification and estimation of the fringe demand function in the Online Appendix B.

3.4 Optimal Choices of Coal Quality, Trading, and Banking

I now introduce the optimization problem. A firm makes both discrete (participation) and continuous decisions related to coal quality (emissions rate of fuel), trading, and banking. I first explain the decision problems conditional on the status of trading participation. These problems characterize the values from participation and nonparticipation, which determine the optimal participation decision.

Let V_{it}^1 and V_{it}^0 be the optimal values when a firm participates in trading ("trader") and does not participate ("nontrader"), respectively. The Bellman equation for a trader is given by

$$
\begin{aligned}
V_{it}^1\left(h_{it}, \alpha_{it}\right) = \quad & \max_{\{R_{jt}\}_{j \in J_{it}}, b_{it}, h_{i,t+1}} && \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) - \left(P_t b_{it} + TC(|b_{it}|)\right) + \beta EV_{i,t+1}(h_{i,t+1}, 1, \alpha_{it+1}) \\
& \text{s.t.} && e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha_{it}\right) + h_{i,t+1} = a_{it} + h_{it} + b_{it}, \\
& && h_{i,t+1} \geq 0.
\end{aligned} \tag{3.5}
$$

Here, $EV_{it}(h_{it}, I_{it}, \alpha_{it})$ denotes the ex-ante value function for firm i in period t when the firm holds h_{it} units of emissions permits, the trading experience is I_{it}, and the removal rate of scrubber is α_{it}. Recall that $I_{it} = 1$ if firm i has participated in the market previously, thus paying the participation cost. In such a case, $EV_{it}(h_{it}, 1, \alpha_{it}) = V_{it}^1(h_{it}, \alpha_{it})$. When a firm is a nontrader (i.e., $I_{it} = 0$), the trading volume b_{it} is not the choice variable. The Bellman equation in this case is similarly given.

Note that the value functions $V_{it}^0(\cdot)$ and $V_{it}^1(\cdot)$ are indexed by firm i and time t. The former is due to the firm heterogeneity, while the latter is due to the nonstationary nature of the decision problem. These indices implicitly subsume all state variables, except for h_{it}, I_{it}, α_{it}, and ϵ_{it}. I assume perfect foresight over the state variable in the next period, except for the shock to the participation cost ϵ_{it}.[38]

The right-hand side of the Bellman equation (3.5) is a constrained optimization problem.

[38]See Section 6.2 for a detailed discussion of this assumption.

A firm optimally chooses unit-level coal quality (emissions rate of fuel) $\{R_{jt}\}_{j=1}^{J_{it}}$, the trading volume b_{it}, and the banking $h_{i,t+1}$ subject to the transition of permit holding and the non-borrowing constraint.[39] The optimality conditions for the traders are given by[40]

$$\lambda_{it}(1-\alpha_{it})HR_j q_{jt} = -\frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}} HR_j q_{jt} \tag{3.6}$$

$$\lambda_{it} = P_t + \frac{dTC(|b_{it}|)}{db_{it}} \tag{3.7}$$

$$\lambda_{it} = \beta \frac{dEV_{i,t+1}(h_{i,t+1}, I_{i,t+1}, R_{i,t+1})}{dh_{i,t+1}} + \mu_{it}, \tag{3.8}$$

$$\mu_{it} h_{it+1} = 0, \mu_{it} \geq 0, h_{i,t+1} \geq 0, \tag{3.9}$$

and the transition of permit holding given by Equation (3.3). λ_{it} denotes the Lagrange multiplier on the transition of permit holding (3.3), while μ_{it} denotes the Lagrange multiplier on the non-borrowing constraint (3.4). The Lagrange multiplier λ_{it} is interpreted as the shadow value of emissions permits for firm i.

Equation (3.6) is obtained from the optimal coal quality decision. The term $(1-\alpha_{jt})HR_j q_{jt}$ in the left-hand side is the marginal abatement of emissions when a firm adjusts coal quality. Multiplying the term with the shadow value λ_{it}, the left-hand side is the value of marginal abatement through coal quality choice. The right-hand side is the marginal cost of coal purchase. Rewriting this equation, I get

$$\lambda_{it} = \frac{1}{(1-\alpha_{it})} \frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}}, \tag{3.10}$$

implying that the marginal price of fuel after adjusting the removal rate of scrubber is equal to the shadow costs of permits λ_{it}. This equation will be key in identifying the transaction cost parameter, which I discuss in Section 4.2.

Equations (3.7)–(3.9) determine the shadow value λ_{it} from the trading and banking decisions. Equation (3.7) states that the shadow value is equal to the sum of the market price and the marginal transaction costs $\frac{dTC(|b_{it}|)}{db_{it}}$. In other words, the marginal transaction cost serves as the wedge between the market price and the firm-level shadow value of emissions.

Equations (3.8) and (3.9) constitute the Euler equation: the shadow value of an emissions permit today is equal to the sum of the discounted marginal value of holding an additional permit tomorrow and the shadow value of borrowing (when it is binding). These conditions, along with the transition equation of permit holdings (3.3), determine the optimal choices for coal type $\{R_{jt}\}_{j=1}^{J_{it}}$, trading b_{it}, and banking $h_{i,t+1}$.

[39] In numerical analysis, I restrict R_{jt} between 0.1 lbs per MMBtu and 5.0 lbs per MMBtu.

[40] The Lagrangian for this constrained optimization problem is $\mathcal{L} = \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) - (P_t b_{it} + TC(|b_{it}|)) + \beta EV_{i,t+1}(h_{i,t+1}, 1, \alpha_{i,t+1}) + \lambda_{it}\left(a_{it} + h_{it} + b_{it} - e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha_{it}\right) - h_{i,t+1}\right) + \mu_{it} h_{i,t+1}$. Taking the first-order-conditions for each choice variable, I obtain the optimality conditions (3.6), (3.7), (3.8), and (3.9). See the Online Appendix D.1 for the detailed derivation.

The optimality conditions for non-traders are the same as those above, except $b_{it} = 0$ and I do not have Equation (3.7). These conditions imply that the shadow value of an emissions permit is not directly related to today's permit price. Rather, the shadow value is given by the discounted marginal value from Equation (3.8).

Given the optimal value V_{it}^0 and V_{it}^1, I consider the participation decision. If a firm has no prior trading experience (i.e., $I_{it} = 0$), it can choose whether to participate in the market by paying $F_{it}(= F + \sigma_F \epsilon_{it})$. A firm participates in the market if $V_{it}^1(h_{it}, \alpha_{it}) - (F + \sigma_F \epsilon_{it}) > V_{it}^0(h_{it}, \alpha_{it})$. Thus, this optimal decision leads to the following participation probability:

$$\mathbb{P}_{it}(h_{it}, \alpha_{it}) = \int \mathbf{1}\left\{V_{it}^1(h_{it}, \alpha_{it}) - (F + \sigma_F \epsilon_{it}) > V_{it}^0(h_{it}, \alpha_{it})\right\} dG(\epsilon_{it}).$$

Since I assume the type-I extreme value distribution of ϵ_{it}, the participation probability is given by the well-known logit formula.[41] If a firm has already participated in trading (i.e., $I_{it} = 1$), it does not have to pay the participation costs.

Based on the optimal choices for traders and non-traders, I now provide the value function. Let $V_{it}(h_{it}, I_{it}, \alpha_{it}, \epsilon_{it})$ be the value function after observing the random draw of the participation costs. The value function is given by

$$V_{it}(h_{it}, I_{it}, \alpha_{it}, \epsilon_{it}) = \begin{cases} \max\left\{V_{it}^0(h_{it}, \alpha_{it}), V_{it}^1(h_{it}, \alpha_{it}) - (F + \sigma_F \epsilon_{it})\right\} & if \;\; I_{it} = 0 \\ V_{it}^1(h_{it}, \alpha_{it}) & if \;\; I_{it} = 1. \end{cases}$$

Finally, the ex-ante value function $EV_{it}(h_{it}, I_{it}, \alpha_{it})$ (before observing ϵ_{it}) are given by the integral of V_{it} with respect to ϵ_{it}.[42]

Continuation Value at the Terminal Period My model is a finite-period model and the terminal period T is 2003, which is the last period of my sample. Although the Acid Rain Program continued after 2003, I choose the year 2003 as the terminal period for the reasons discussed at the beginning of Section 2.2. To model the incentive to bank permits at the terminal period 2003, I include the reduced-form continuation value function $CV_{T+1}(h_{i,T+1})$ in the model. In Section 4.2, I provide the functional form of $CV_{T+1}(h_{i,T+1})$, and estimate it along with other parameters.[43]

3.5 Investment Decisions on Scrubbers

This subsection introduces the investment decision problem regarding scrubbers. The model assumes that a firm makes a scrubber investment at the beginning of each phase (in 1995 and 2000). This assumption implies that the removal rate of a scrubber α_{it} is phase-specific,

[41] See the Online Appendix D.2 for more details.

[42] See the Online Appendix D.2 for the details of the ex-ante value function $EV_{it}(h_{it}, I_{it}, \alpha_{it})$.

[43] See, e.g., Keane and Wolpin (2001) and Jørgensen and Tô (2020) for an empirical analysis that takes a similar approach.

i.e., $\alpha_{it} = \alpha_i^1$ for $t \in \{1995, \ldots, 1999\}$ and $\alpha_{it} = \alpha_i^2$ for $t \in \{2000, \ldots, 2003\}$. While this assumption simplifies the investment problem, this is motivated by the data and several other considerations. First, the data show the lumpy investment pattern, as found in Figure A8 in the Online Appendix.[44] Much of the scrubber adoption occurs between the announcement of the ARP and the beginning of regulation. Such lumpy investment might reflect the fact that the installment of a scrubber requires several years of lead time. As a result, firms make a long-term decision to install a scrubber, rather than making a new decision every year based on the economic environment at the time. An alternative approach would be to incorporate the time-to-build nature of scrubber investment and to allow firms to make an annual decision. Such a model would be significantly complicated, though. Given these considerations, I choose to adopt a simple approach to model scrubber adoption.[45]

Firm i determines the install rate of a scrubber in each Phase l, denoted by g_i^l for $l = 1, 2$. The install rate g_i^l denotes the share of the total capacity of generating units to which the scrubber is installed, and thus can take a value between 0 and 1. The removal rate of a scrubber α_i^l is determined by $\alpha_i^l = 0.9 g_i^l$, implying that the maximum removal rate of a scrubber is 90%.[46] I assume that the install rate of a scrubber is a continuous choice variable. I denote the cost function of scrubber installment by $\Gamma\left(\left(g - \bar{g}\right) k_i\right)$, where g is the install rate chosen by a firm, $\bar{g}$ is the install rate before the investment decision, and k_i is the total capacity of coal units owned by firm i. Note that the capacity size of units to which a scrubber is installed is denoted by $\left(g - \bar{g}\right) k_i$.[47]

The investment problem for Phase I is given by

$$\max_{g_i^1} \; EV_{i,1995}(h_{i,1995}, I_{i,1995}, \underbrace{0.9 g_i^1}_{=\alpha_i^1}) - \Gamma\left(\left(g_i^1 - g_i^0\right) k_i\right) \quad \text{s.t. } g_i^1 \geq g_i^0, \tag{3.11}$$

where g_i^0 is the install rate of a scrubber in 1990. I choose 1990 as the baseline year because the Acid Rain Program was announced in 1990. I incorporate the irreversibility of investment by allowing g_i^0 to affect both the investment cost and the choice set of the install rate of a scrubber g_i^1. The problem for Phase II is defined similarly, except that the investment cost now depends on g_i^1, which is determined endogenously at the beginning of Phase I.

[44] For Group 1 units that are regulated from 1995, the number of units with a scrubber increased from 26 in 1990 to 52 in 1995. After that, it only increased from 52 in 1995 to 58 in 2003. Regarding the Group 2 units that are regulated from 2000, the number increased from 122 in 1990 to 148 in 2000. Although the number increased further to 171 in 2003, the increase in the total capacity of units with a scrubber was not as large. From 1990 to 2000, the total capacity increased by 9,012 MW (52,225 MW to 61,273 MW), while it increased by 4,136 MW from 2000 to 2003. See Figure A8 in the Online Appendix for more details.

[45] See, e.g., Kalouptsidi (2014), whose work incorporates the time-to-build nature of investment decisions in dynamic structural estimation.

[46] This number is taken from Table 9.3 of Ellerman, Joskow, Schmalensee, Montero, and Bailey (2000).

[47] The firm-level generation capacity k_i is given by the sum of the generation capacity of coal generating units a firm owns. The capacity is a fixed characteristic of generating units and is assumed to be exogenous in this analysis. In other words, the abatement choice for SO_2 emissions does not affect the generation capacity.

3.6 Dynamic Competitive Equilibrium of Permit Trading

To close the model, I define an equilibrium of the permit market. I assume that firms have perfect foresight over the future environment and that the only stochastic shock is the participation cost ϵ_{it}. I discuss the importance and implications of this assumption in Section 6.2.

Definition (Dynamic Competitive Equilibrium)**.** *In a finite-period competitive equilibrium with perfect foresight, a sequence of permit prices* $\{P_t\}_{t=1995}^{2003}$ *is determined such that*

1. *[Optimization] Each firm* i *optimally chooses* $\left\{\{R_{jt}^*\}_{j=1}^{J_{it}}, b_{it}^*, h_{i,t+1}^*\right\}_{t=1995}^{2003}$ *and* $\{g_i^{1*}, g_i^{2*}\}$*, given a sequence of permit prices, and*

2. *[Market Clearing]* $\sum_{i=1}^N b_{it}^*\left(\{P_t\}_{t=1995}^{2003}\right)+\bar{B}_t^{fringe}(P_t) = 0$ *holds for all* $t = 1995, \cdots, 2003$.

To solve the equilibrium, I repeat the following procedure: (i) Given a candidate of permit prices $\{P_t\}_{t=1995}^{2003}$, solve the individual optimization problem using backward induction for each firm, and (ii) calculate the aggregate level of net purchases to determine whether the market-clearing conditions are satisfied in all periods. I use a heuristic rule of updating the price vector in each iteration, which successfully identifies a unique vector of equilibrium permit prices satisfying the market-clearing conditions.[48] I elaborate on how to numerically compute a vector of equilibrium prices in the Online Appendix E.3.

3.7 Model Implications

This subsection discusses the implications of the model. Most notably, I argue that the role played by transaction costs, $TC(\cdot)$ and F_{it}, in firms' dynamic decisions and equilibrium outcomes is significant.[49] To discuss this point, I first show the benchmark case in which no transaction costs exist; namely, $TC(\cdot) = 0$ and $F_{it} + \epsilon_{it} = 0$. In this case, the optimality conditions (3.6)–(3.9) can be summarized as[50]

$$\frac{1}{(1-\alpha_{jt})}\frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}} = P_t. \tag{3.12}$$

$$P_t = \beta P_{t+1} + \mu_{it},\ \mu_{it} \geq 0 \perp h_{i,t+1} \geq 0. \tag{3.13}$$

I discuss the roles of transaction costs in terms of (1) avoiding the indeterminacy of trading/banking decisions and (2) the cost-effectiveness of the cap-and-trade.

[48] Although I lack a formal proof for the uniqueness of the equilibrium, I used several different initial values for permit prices and confirmed that these initial values converge to the same equilibrium prices.

[49] A discussion on how the presence of transaction costs breaks the independence property of the initial allocation(i.e., Coase, 1960 theorem) is available upon request.

[50] Equation (3.7) implies that $\lambda_{it} = P_t$ holds for all i. Using the envelope theorem, Equation (3.8) implies that $\frac{dEV_{i,t+1}(h_{i,t+1}, I_{i,t+1}, R_{i,t+1})}{dh_{i,t+1}} = P_{t+1}$. See the Online Appendix D.2 for the derivation.

3.7.1 Avoiding the Indeterminacy of Trading/Banking Decision

Equation (3.13) implies that the equilibrium permit prices P_t should increase over time at the rate β^{-1} as long as the aggregate banking is positive and no transaction costs exist. This property is known as the Hotelling $r-$percent rule: the price of an exhaustible resource should increase at a rate equal to the interest rate, or the inverse of the discount factor (see, e.g., Rubin, 1996).

More importantly, the model without transaction costs suffers from the indeterminacy of individual optimal decisions; that is, it does not identify the individual optimal behavior for trading b_{it} and banking $h_{i,t+1}$ in the absence of transaction costs. This is because the discounted marginal value from banking is constant (and given by βP_{t+1}), which is equal to the current shadow value P_t in equilibrium. Thus, the marginal values of net purchases b_{it} and banking $h_{i,t+1}$ are always the same. Therefore, any choices regarding b_{it} and $h_{i,t+1}$ are equivalent for a firm, as long as it can produce the level of emissions given by the optimality condition on coal quality (3.12).

I now consider the case in which transaction costs are present. Combining optimality conditions (3.7) and (3.8) and using the envelope theorem, I obtain the following condition:

$$P_t + \frac{dTC(|b_{it}|)}{db_{it}} = \beta \left\{ P_{t+1} + \frac{dTC(|b_{it}|)}{db_{it}} \right\} + \mu_{it}.$$

This condition implies that the permit price does not necessarily increase at the rate of β^{-1}. Without convex transaction costs, the price path in which $P_t > \beta P_{t+1}$ (or $P_t < \beta P_{t+1}$) cannot be an equilibrium because firms have an incentive to sell their permit holding (or buy an infinite number of permits) in period t. Intuitively, the presence of transaction costs prevents firms from engaging in complete inter-temporal arbitrage.

The model now identifies the optimal decisions for both net purchases b_{it} and banking volume $h_{i,t+1}$ because the marginal values of the two are no longer constant. The marginal cost from net purchases increases due to the convex transaction costs $TC(|b_{it}|)$. The discounted marginal value from banking, given by $\beta \left(P_{t+1} + \frac{dTC(|b_{it}|)}{db_{it}} \right)$, decreases in h_{t+1} because holding additional permits in period $t+1$ (i.e., higher $h_{i,t+1}$) leads to lower $b_{i,t+1}$ (selling more permits) and, thus, a lower marginal value. In other words, the marginal revenue from selling is shown to fall when firms sell additional permits because they have to pay transaction costs.

3.7.2 Cost-effectiveness of Cap-and-Trade

One of the virtues of a cap-and-trade regulation is that the equilibrium allocation of emissions, given the emissions cap, is cost-effective in the absence of transaction costs. Equation (3.12) implies that the marginal abatement cost is equalized across firms at the level of permit price P_t. The key mechanism is that all firms face the same shadow value given by the market price (i.e., $\lambda_{it} = P_t, \forall i$).

I now examine how the transaction cost affects the shadow costs of emissions permits and leads to an inefficient outcome of cap-and-trade. Consider two types of firms: a buyer (i.e., $b_{buyer,t} > 0$), and a seller (i.e., $b_{seller,t} < 0$). Equation (3.6) implies that

$$\lambda_{buyer,t} > P_t > \lambda_{seller,t}.$$

The inequalities hold because $\frac{dTC(|b_{it}|)}{db_{it}} > 0$ for $b > 0$, and $\frac{dTC(|b_{it}|)}{db_{it}} < 0$ for $b < 0$.[51] Intuitively, in the presence of variable transaction costs, buyers incur additional costs to purchase emissions permits. In contrast, the revenue from selling a unit of emissions permits is the market price minus the marginal transaction costs. Thus, the marginal cost of emissions for the buyer is strictly higher than that for the seller. In other words, buyers emit less and sellers emit more than the efficient level at which the marginal abatement costs are equalized across firms.

The heterogeneity of the shadow value has an important impact on abatement decisions and permit banking. The return on investment is determined by the marginal abatement of emissions, given by $\sum_{j=1}^{J_{it}} R_{jt} \cdot HR_{jt} \cdot q_{jt}$, evaluated at the shadow value of λ_{it}.[52] Because the shadow value for buyers is higher than that for sellers, buyers have a greater incentive to invest, whereas sellers have a lower incentive.

With regard to permit banking, buyers have a lower incentive to bank permits because the transaction costs lead to a higher shadow value today. By contrast, sellers prefer to bank additional permits because of the lower shadow value. Thus, the aggregate level of permit banking can be higher or lower than the case without transaction costs. I will examine the impact of transaction costs on aggregate permit banking in Section 5.

3.8 Discussion of Modeling Assumptions

In this subsection, I briefly discuss modeling assumptions. However, I defer detailed discussions on several issues, including (1) model of production decision, (2) perfect foresight assumption, and (3) public utility regulation to Section 6.

Entry/Exit Decisions The model does not incorporate entry and exit decisions. As I discussed at the beginning of Section 2.3.2, the retirement of the coal unit was relatively limited in the sample period I employed, indicating that emissions abatement through the unit retirement was not a major option. Regarding entry decision, I observe virtually no entry of coal units. Although I observe the entry of gas units, those units served as marginal units in the merit order of electricity generation due to the higher marginal costs of production and smaller generation capacity.[53] Therefore, firms would not replace coal units with gas

[51] To see this, $\frac{dTC(|b_{it}|)}{db_{it}} = \frac{dTC(|b_{it}|)}{d|b_{it}|} \frac{d|b_{it}|}{db_{it}}$. The first term is positive and the second term is positive (negative) when $b > 0$ ($b < 0$).

[52] See the Online Appendix D.3 for the derivation of the marginal returns from an abatement investment.

[53] Although gas units have recently become competitive with coal units in terms of production costs due to the shale boom since the late 2000s, gas units exhibited much higher marginal production costs in my sample

units as an abatement strategy in my study period. Given these observations, I chose not to incorporate entry and exit decisions in my model. See Ryan (2012), Fowlie, Reguant, and Ryan 2014, Dardati (2016), and Cook and Lin Lawell (2018), which empirically study the implications of environmental regulations for entry and exit decisions.

Start-up Costs The production decision in the model excludes the dynamic incentive that arises due to the presence of the start-up costs in electricity generation. Such a dynamic incentive would matter when I model the production decision at a high frequency (e.g., daily or hourly). Instead, this paper focuses on longer-horizon decisions (namely year-level decisions) under cap-and-trade. See Mansur (2008) and Reguant (2014) for an empirical study that explicitly considers the start-up costs in electricity generation in a high-frequency (daily) setting.

Environmental Regulation on NO_x Emissions The analysis focuses on SO_2 emissions and does not consider NO_x emissions from coal power plants. The NO_x budget program, which is a cap-and-trade program on NO_x emissions regulation, started in 2003. There are 56 firms (out of a total of 138 firms) that own generating units that are affected by this program. However, the effects of NO_x regulation on SO_2 abatement are limited. First, the NO_x control technologies, including low NO_x burners and selective catalytic reduction, do not affect SO_2 emissions, which is primarily abated by fuel switching and scrubbers. Second, the abatement of these two types of emissions might be interrelated through the reduction of electricity generation. If the generation unit faces a shadow cost of NO_x emissions due to the presence of cap-and-trade on NO_x, it might reduce electricity generation due to the higher production costs, which contributes to the abatement of SO_2 emissions. However, Fowlie (2010b) showed that the abatement of NO_x emissions through production reduction is limited. For these reasons, I do not consider NO_x emissions in this paper.

Interpretation of Continuation Value Function $CV_{T+1}(\cdot)$ The continuation value function $CV_{T+1}(h_{T+1})$ captures the incentive to bank emissions permits at the terminal period (i.e., 2003). More precisely, I assume that $CV_{T+1}(\cdot)$ captures a firm's incentive to bank under the expectation that the Acid Rain Program continues after 2004 without any additional regulations. This is a reasonable assumption, given that the Clean Air Interstate Rule was announced in the last month of 2003 (i.e., December 2003), implying that, in 2003, firms were expecting the same regulatory environment to continue after 2004.[54]

period of 1995–2003.

[54]One might think that an alternative approach is to model the problem in the terminal period as a stationary and infinite-period dynamic programming problem. I believe, however, that the stationary assumption is not adequate in Phase II. Even though permit allocation does not change during Phase II (i.e., after 2000), the aggregate stock (banking level) of emissions permits decrease over time. Thus, equilibrium permit prices still change depending on the aggregate level of permit holding in each year (see, e.g., Rubin (1996) and Schennach (2000) for theoretical analysis). Therefore, I assume a non-stationary environment in Phase II. This assumption requires me to model the continuation value function in the terminal period.

4 Estimation

This section explains the estimation of model primitives and results. In Section 4.1, I obtain the gross profit function $\pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right)$ in Equation (3.1) by estimating the hedonic function for coal price $p_{jt}^{fuel}(R_{jt})$ without solving the dynamic decision problem. Using the estimated profit function, I estimate the remaining model parameters in Section 4.2, including the variable transaction costs $TC(|b|)$, the distribution of the fixed transaction costs F_{it}, and the continuation value at the terminal period $CV_{T+1}(h_{i,T+1})$. Identification of these parameters is also discussed. To estimate these parameters, I use a simulated nonlinear least squares approach, in which I numerically solve the individual dynamic decision problems to match the model prediction with its empirical counterpart. Note that I fix the annual discount factor at $\beta = 0.95$ throughout this paper. Moreover, I use engineering estimates of the install costs of a scrubber.[55]

4.1 Hedonic Regression of Coal Price

The gross profit function given by Equation (3.1) depends on the fuel price function $p_{jt}^{fuel}(R_{j,t})$. I estimate the fuel price hedonic function using the fuel procurement data.[56] The data are taken from the Form FERC No.423 (EIA-423) "Monthly Report of Cost and Quality of Fuels for Electric Plants". These data report plant- and monthly-level information on fuel procurement, including fuel type, sulfur content, heat content, and purchase costs.

I consider the following hedonic function, which describes the coal price $p_{k,l,m}^{fuel}$ in fuel delivery k for plant n in month-year m:

$$\log(p_{k,n,m}^{fuel}) = (\phi_r + \phi_1^{age}\log(age_{nm}))\log(R_{k,n,m}) + \phi_0^{age}\log(age_{nm}) + \phi_s + \phi_m + u_{k,n,m}, \quad (4.1)$$

where $p_{k,n,m}^{fuel}$ is the coal price including shipping costs, measured in cents per MMBtu, and $R_{k,n,m}$ is the SO_2 emissions rate of fuel, measured in lbs per MMBtu. Although other coal characteristics (e.g., ash content) are available in the dataset, I only include the emissions rate of fuel in this function because those characteristics depend on the SO_2 emissions rate of fuel.[57]

In the above specification, I incorporate a rich set of control variables that captures the plant-level heterogeneity. First, ϕ_m and ϕ_s are year-and-month dummies and state-of-plant

[55] Estimation of the fringe demand $\bar{B}_t^{fringe}(P_t)$ is discussed in the Online Appendix B.

[56] I need to estimate the fuel price hedonic function because the choice of emissions rate of fuel is continuous in my model. This approach stands in contrast to the approach adopted in other papers where the choice of coal type is modeled as the discrete decision (e.g., Keohane, 2006; Chan, 2015).

[57] The primary role of the fuel price function is to predict how the choice of an SO_2 emissions rate of fuel affects the coal price. Because other product characteristics could change with the choice of the SO_2 emissions rate of fuel, it would be erroneous to predict how the coal price would change with respect to the SO_2 rate, *holding other characteristics fixed.* For example, the distance between coal plants and coal mines is an important determinant of coal price. However, if a plant changes its emissions rate of coal, it also changes the coal mine from which it buys coal, thus affecting the distance. Therefore, I only include plant-level characteristics, not coal-specific characteristics, in the fuel price function.

dummies. The geographic location of power plants is an important factor in coal procurement, as it involves a large shipping cost. The elasticity parameter ϕ_r is allowed to be different across regions because plants located closer to the West and the Powder River Basin are able to buy low-sulfur coal at the cheaper prices due to lower shipping costs.[58] I also include the variable age_{nm} defined by the average age of coal units owned by plant n. The age variable captures the plant-level heterogeneity in abatement costs through fuel procurement. For example, the newer plants might find it easier to switch from dirtier to cleaner coal.

Table 2 reports the results of coal price regression. Region group 1 consists of Northeast and South, while region group 2 consists of Midwest and West. The estimated coefficients imply that the price of coal for the given emissions rate of fuel is cheaper in region group 2, which is consistent with the fact that Midwest and West regions are closer to the Powder River Basin and thus have better access to coal. Note that the estimated coefficient on the log of emissions rate of fuel, $\phi_r + \phi_1^{age} \log(age_{nm})$, is negative for all observations. Thus, the marginal fuel price increases as the emissions rate of fuel decreases, implying that the marginal cost of abatement by coal quality changes is higher for lower emissions rate of fuel.

Table 2: Hedonic Regression of Coal Price

Model:	log(fuel price) (1)
Variables	
log(emissions rate of fuel)*(Region Group 1)	-0.1496 (0.0042)
log(emissions rate of fuel)*(Region Group 2)	-0.0686 (0.0043)
log(emissions rate of fuel)*log(age)	0.0114 (0.0013)
log(age)	0.0161 (0.0008)
Year-and-Month fixed effect	Yes
State of Plant fixed effect	Yes
Fit statistics	
R^2	0.49
Observations	298,378

Notes: Estimates of month-and-year and state dummies are omitted. Robust standard errors are reported.

4.2 Estimation of Transaction Costs and Continuation Value Function

With the gross profit function $\pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right)$ estimated in Step 1, I estimate the remaining parameters, including the transaction costs and continuation value. As noted above, I fix the

[58] I use the definition of US regions provided by the US Census Bureau (U.S. Census Bureau, 2011). There are four regions: Northeast (CT, ME, MA, NH, RI, VT, NJ, NY, PA), Midwest (IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD), South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX), and West (AZ, CO, ID, MT, NV, NM, UT, WY, AK, CA, HI, OR, WA).

annual discount factor at $\beta = 0.95$ throughout the analysis. I use engineering estimates of the install costs of a scrubber. The install costs of a scrubber is taken from Ellerman, Joskow, Schmalensee, Montero, and Bailey (2000) and set to \$239 per kilowatt hour of capacity in 1994 USD.[59]

My model features two types of transaction costs: variable costs and participation costs. The variable transaction cost function $TC(|b|)$ is specified as follows:

$$TC(|b|) = \frac{1}{\eta_2 + 1} \exp\left(\eta_0^{buy} \mathbf{1}\{b > 0\} + \eta_0^{sell} \mathbf{1}\{b < 0\} + \eta_1 \log(size_i)\right) |b|^{\eta_2+1}, \tag{4.2}$$

where $size_i$ denotes firm i's size, measured by the sum of the generation capacity of firm i. I allow a firm's size to affect the magnitude of transaction costs to determine whether bigger firms face lower transaction costs. An example of such a possibility is investing human capital of financial traders. Larger firms, which also trade larger volumes of permits, may need to invest in more labor, ultimately leading to lower variable transaction costs. In addition, I allow transaction costs to differ depending on whether the firm is buying or selling permits. Lastly, to estimate the variable transaction cost function in a flexible way, I incorporate the parameter η_2, which measures the curvature of the variable transaction cost function. The parameter is assumed to satisfy $\eta_2 > 0$ to maintain the strict convexity and the smoothness at $b = 0$.[60]

The participation cost F_{it} is specified as

$$F_{it} = F + \sigma_F \epsilon_{it}, \tag{4.3}$$

where ϵ_{it} follows an i.i.d. type-I extreme value (Gumbel) distribution, with standard deviation parameter σ_F.

I consider the following specification of the continuation value in the terminal period:

$$CV(h_{i,T+1}) \quad = \quad \exp\left(\psi_0 + \psi_1 \log(size_i) + \psi_2 \alpha_i^2 + \psi_3 I_{i,T}\right) (h_{i,T+1})^{\psi_4}. \tag{4.4}$$

The coefficient depends on the firm size, $size_i$, and on the removal rate of the scrubber in Phase II, α_i^2. These variables capture the heterogeneity in the incentives to bank permits in the terminal period (i.e., 2003). The parameters estimated in this step are summarized as $\boldsymbol{\theta} = (\eta_0^{buy}, \eta_0^{sell}, \eta_1, \eta_2, F, \sigma_F, \psi_0, \psi_1, \psi_2, \psi_3, \psi_4)$.

The estimation procedure builds on the literature of estimation of dynamic structural

[59] I use estimates from ICF90 and EPRI93 reported in Table 9.3 of Ellerman, Joskow, Schmalensee, Montero, and Bailey (2000). I take the average of initial capital costs of ICF90 and EPRI93. Because these reported numbers are measured in 1994 USD, I converted them into 2000 USD in the analysis.

[60] Alternatively, I could adopt a flexible specification that uses both quadratic and linear terms such as $TC(|b|) = \eta_1 |b| + \eta_2 |b|^2$. However, a linear component creates a kink at $b = 0$, making numerical computation difficult. The specification of the transaction cost $TC(|b|)$ adopted for estimation is a smooth function with respect to b and keeps a flexibility by a parameter η_2.

models in industrial organization and labor economics.[61] I use a simulated nonlinear least squares approach to estimate the model parameters. For a given candidate of parameter $\boldsymbol{\theta}$, I solve the model to obtain the prediction of the choice variables and match it with the data. The procedure for obtaining the model prediction is as follows:

1. Fix a candidate of parameter $\boldsymbol{\theta}$ and the observed permit prices $\{P_t\}_{t=1995}^{2003}$.
2. For each firm i, solve the optimization problem using backward induction and obtain the policy functions.
3. Using the policy functions, simulate the optimal decisions for each pattern of participation in permit trading. Denote the year of participation by $s \in \{\emptyset, 1995, \cdots, 2003\}$, where $s = \emptyset$ means that the firm does not trade in that period. Denote the optimal decision for pattern s by $\hat{x}_{it}(s)$.
4. Calculate the probability that each pattern of participation timing is realized. Denote this probability by $Prob_{it}^{enter}(s)$.
5. The prediction for firm i in year t is given by

$$\hat{x}_{it} = \sum_{s \in \{\emptyset, 1995, \cdots, 2003\}} Prob_i^{enter}(s) \hat{x}_{it}(s). \tag{4.5}$$

Note that I do not have to solve for a dynamic competitive equilibrium to obtain a model prediction because I can use the observed prices of emissions permits as a sequence of equilibrium prices. Given the observed permit prices, I solve the single-agent optimization problems, which are much easier to solve than the dynamic competitive equilibrium.[62]

Using the simulated choices, I calculate the objective function. The objective function measures the distance between the prediction and the data at the firm and year levels:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left(\mathbf{x}_i^{data} - \hat{\mathbf{x}}_i(\boldsymbol{\theta}) \right)' \Omega_i \left(\mathbf{x}_i^{data} - \hat{\mathbf{x}}_i(\boldsymbol{\theta}) \right),$$

where $\mathbf{x}_i^{data}$ is a vector of endogenous variables, and $\hat{\mathbf{x}}_i(\boldsymbol{\theta})$ is the corresponding vector for the model prediction, given parameter $\boldsymbol{\theta}$. The vector $\mathbf{x}_i^{data}$ includes emissions e_{it}, net purchases b_{it}, trading volume $|b_{it}|$, permit banking h_{it}, the install rate of a scrubber (g_i^1, g_i^2), as well as a dummy variable that indicates whether firm i participates in the permit market. The weighting matrix Ω_i is a diagonal matrix used to adjust for differences in scaling. Specifically, I use the inverse of the variance of each choice variable in the dataset as a weight.

[61] See Aguirregabiria and Mira (2010) for a survey of this literature.

[62] This empirical strategy is similar in spirit to that in the two-step estimation of a dynamic game, in which the equilibrium objects are recovered directly from the observed data. For example, Aguirregabiria and Mira (2007) estimate players' beliefs over other players' policies from the observed data. They then solve the optimal response of a player, given the estimated beliefs, in order to construct the pseudo-likelihood function.

Standard errors are calculated using the bootstrap method at the firm-history level. I randomly draw samples from 114 firms, with replacement, and construct 40 bootstrap samples.

Identification I first discuss the identification of the variable transaction costs. The optimality conditions (3.10) and (3.7) imply that, without variable transaction cost $TC(\cdot)$, the marginal price of fuel (i.e., the left-hand side of Equation (3.10)) should be equalized across firms and made equal to the permit price. Since I obtain the marginal price of fuel by estimating the hedonic function, I can identify the variable transaction costs using these two conditions. Intuitively, the variable transaction costs are identified from how the marginal fuel price vary with the trading volume.[63]

My identification strategy relies on the assumption that the variation in transaction volume is uncorrelated with unobserved factors that may influence the fuel price. Thus, it is crucial to estimate the fuel price function by controlling for potential heterogeneity. In estimating the fuel hedonic function, I incorporate a rich set of fixed effects that capture aggregate shocks and geographic difference in coal access. Any remaining heterogeneity the specification does not capture might be attributed to the estimated transaction costs, implying the overestimation of the costs.

Other primitives are also identified in a similar manner. Parameters on the distribution of the sunk participation costs (F, σ_F) are identified using the information on the firm-level participation in permit trading. Note that I observe the level of payoffs, namely the gross profit (i.e., the negative of the total production cost) $\pi_{it}(\cdot)$, which allows me to identify the scale parameter σ_F. Lastly, the continuation value at the terminal period $CV_{T+1}(h_{i,T+1}, R_i^2)$ is identified by the optimality condition for the banking in the terminal period $h_{i,T+1}$.[64]

Estimation Results Table 3 presents the parameter estimates of $\boldsymbol{\theta}$. With regard to the variable transaction costs, the coefficient on the firm size is negative but small. This result implies that, although bigger firms tend to have lower transaction costs, the heterogeneity across firms is negligible. Based on the parameter estimates, I calculate the marginal transaction cost, given by $\exp\left(\eta_0^{buy}\mathbf{1}\{b>0\} + \eta_0^{sell}\mathbf{1}\{b<0\} + \eta_1 \log(size_i)\right)|b|^{\eta_2}$. The mean of the costs is \$46.1, while the median is \$32.1 Considering that the permit prices range between \$100 and \$200, as reported in Figure 4, the estimated transaction costs are sizeable. This estimate indicates the large dispersion of the shadow value of emissions across firms, implying an

[63]With the transaction-level price, which is not available in this empirical context, I could measure transaction costs at the transaction level by comparing the shadow value of emissions and the actual transaction price. The granular data would allow me to estimate transaction costs at the buyer-seller combination level. See Hagerty (2019), who uses the transaction-level price of water rights to estimate transaction costs in the context of water markets in California.

[64]The FOC from the banking decision in the terminal period $h_{i,T+1}$ implies $\lambda_{i,T} = \frac{\partial CV_{T+1}(h_{i,T+1}, R_i^2)}{\partial h_{i,T+1}} = \psi_4 \cdot \exp\left(\psi_0 + \psi_1 \log(size_i) + \psi_2 \alpha_{i,T}^2 + \psi_3 I_{i,T}\right)(h_{i,T+1})^{\psi_4 - 1}$. Since $\lambda_{i,T}$ is obtained by the marginal fuel price (i.e., Equation (3.10)), I can use this equation to identify the parameters in the continuation value function.

inefficient outcome of the cap-and-trade. The mean and the standard deviation of the participation costs is around \$0.64 million and \$3.57 million, respectively. However, these parameters are rather imprecisely estimated because the participation rate in permit trading is relatively high (91%), as seen in Section 2.3.4. The estimated parameters in the continuation value function at the terminal period have intuitive signs. Installment of a scrubber and participation in permit trading have a positive effect on continuation value.

Table 3: Parameter Estimates of Transaction Costs and Continuation Value Function

	Parameter	Description	Estimate	Standard Errors
Variable Costs $TC(\cdot)$	η_0^{buy}	Constant for buying	−3.015	(1.189)
	η_0^{sell}	Constant for selling	−4.005	(1.392)
	η_1	Firm size	−0.008	(0.004)
	η_2	Curvature	0.461	(0.208)
Participation Costs F_{it}	F	Mean (\$1 million)	0.639	(0.357)
	σ_F	Std. Dev. (\$1 million)	3.574	(0.266)
Continuation Value $CV_{T+1}(\cdot)$	ψ_0	Constant	1.332	(0.358)
	ψ_1	Firm Size	0.019	(0.004)
	ψ_2	Install rate of a Scrubber	0.184	(0.013)
	ψ_3	Trading Status	0.831	(0.131)
	ψ_4	Curvature	0.115	(0.048)

5 Policy Simulation

Using the estimated model, I conduct a series of counterfactual simulations to investigate the implications of transaction costs and alternative policy design. Table 4 summarizes the simulation designs. I first quantify the outcome with the lowest abatement costs under cap-and-trade by eliminating transaction costs (Section 5.1). I then evaluate the impact of the permit banking system in Section 5.2. As an alternative to cap-and-trade regulation, I simulate the outcome under emissions tax policy in Section 5.3. As I will explain later, the emissions tax policy can be considered to be the policy that eliminates both transaction costs and permit banking. The Online Appendix F explains how to simulate the equilibrium outcome in each case. All simulation outcomes are compared against the baseline outcome, which I solve using the estimated parameter. I calculate standard errors of simulation results via parametric bootstrap of model estimates.[65]

[65] I use parameter estimates based on bootstrap samples obtained in Section 4.2. Note that I fix all other parameters, including the fuel price hedonic function. I do this to quantify the standard errors of simulations associated with the key model parameters, namely transaction costs, that affect the welfare implications.

Table 4: Simulation Design

	With banking	No banking
With transaction costs	Baseline	No permit banking between Phases (Section 5.2)
Without transaction cost	Cost-effective (Section 5.1)	Constant emissions tax (Section 5.3)

5.1 The Potential Gains from Trade by Eliminating Transaction Cost

I first simulate the outcome when all transaction costs are eliminated.[66] This outcome is the one that minimizes the total abatement costs. By comparing this outcome with the baseline, I evaluate how well the cap-and-trade program works in terms of both abatement costs and environmental externality (i.e., health and environmental damages).

As discussed in Section 3.7.2, transaction costs lead to an inefficient outcome in a cap-and-trade program. The estimates of the model parameters suggest that the variable transaction costs are substantial, implying the large dispersion of the shadow value across firms. In this simulation, I set both participation and variable transaction costs to zero.[67] It is also noted that I fix the level of permit banking in 2003 and the fringe supply at the level of baseline equilibrium outcome. By doing so, I fix the total amount of emissions across two different cases.[68]

This simulation quantifies the distortion of the firm's decisions due to the presence of transaction costs in permit trading. I first explain its effect on permit banking. Figure 6 plots the aggregate level of permit banking in each year, with and without transaction costs. Figure 6 indicates that firms bank fewer permits on average in the baseline case compared with the cost-effective case. The transaction costs have a heterogeneous impact on the firm's incentive to bank permits. Consider a firm that has a higher abatement cost and thus wants to buy permits. In the presence of transaction costs, this firm saves fewer permits in Phase I instead of buying permits from the market. By contrast, consider firms that want to sell permits in the frictionless setting. These firms may prefer to save permits due to the frictions. The aggregate impact of the transaction cost on banking is thus ambiguous. The result in Figure 6 shows that the former channel is stronger. Many firms save fewer permits in Phase

[66] One might interpret this counterfactual simulation as the case in which the regulator introduces a fully efficient centralized auction to allocate emissions permits.

[67] I utilize the Hotelling rule to solve the market equilibrium without transactions costs. See the Online Appendix F.1 for more details.

[68] An alternative approach is to enable firms to endogenously bank permits in the terminal period while still fixing the fringe supply at the baseline level. This approach ensures that the total quantity of emissions permits available during the sample period is fixed (not the aggregate emissions level). I also conducted a simulation using this configuration and found that the results were close to those presented in Column (2) of Table 5. Specifically, the total emissions amount to 53.13 million tons, with permit banking at the terminal period amounting to 2.45 million tons. The total cost is \$93,025 million, while the total damage amounts to \$40,008 million.

I due to the transaction costs.

Figure 7 plots the distributions of emissions rates of fuel in the case without transaction costs and the baseline case. The distribution is more dispersed in the absence of transaction costs than in the baseline. Eliminating transaction costs makes firms trade more actively, which in turn makes them more flexible in their compliance. Firms that find it costly to reduce their own emissions are more likely to purchase emissions permits, whereas other firms invest more because their revenue from selling permits increases once transaction costs are removed.

How do these distortions translate into efficiency measures? Table 5 presents the efficiency measures of the equilibrium outcome with and without transaction costs. With regard to the total costs of electricity generation and abatement for firms, the table shows that the costs would be reduced by \$814 million.

In addition to the cost-effectiveness, I investigate the net-benefit of the program by calculating health and environmental damages in Table 5. To do this, I use the estimates of the marginal damages from SO_2 emissions constructed by Muller and Mendelsohn (2009b, 2009a).[69] Health and environmental damage decreases by \$2.44 billion in the absence of transaction costs.

To determine the source of reduction in health and environmental damage, I consider the following decomposition exercise. I denote the total discounted damage by $D = \sum_{t=1995}^{2003} \tilde{d}_t E_t$ where $E_t = \sum_{i=1}^{N} \sum_{j=1}^{J_{it}} e_{jt}$ is the aggregate emission in year t and $\tilde{d}_t$ is the (discounted) average damage from emissions in year t. The latter is defined by $\tilde{d}_t = \beta^{t-1995} \left(\sum_{i=1}^{N} \sum_{j=1}^{J_{it}} \frac{e_{jt}}{E_t} d_j \right)$ where e_{jt} is emissions from unit j in year t and d_j is the health and environmental damage from emissions produced by unit j. This measure is interpreted as the weighted average of health and environmental damages, where the weight is given by the amount of SO_2 emissions.[70] Then, the change in the aggregate damages from the baseline D^{base} to the counterfactual

[69] Muller and Mendelsohn (2009b) use the APEEP model, an integrated assessment model, to calculate marginal damages from SO_2 emissions at different levels of height at the county level. They also report the plant-level marginal damage for some plants. This is reported as the damage at the point sources with an effective height being defined as higher than 500 meters. I use the plant-level damages where possible. However, if the plant-level damage is not reported, I use the marginal damages from point sources with an effective height of higher than 250 meters and less than 500 meters (denoted as high point sources). Following Muller and Mendelsohn (2009b), I assume that damages are linear in SO_2 emissions. The emissions damage from a particular county is given by the product of the marginal damage and the total SO_2 emissions from electricity plants located in the county. While the recent and updated version of the model (namely AP2 and AP3 models) are available, I choose to use the marginal damages from the original APEEP model because the sample period of my analysis (i.e., 1995 to 2003) is closer to when the APEEP model was developed. For the papers that use the AP2 model for calculating the health and environmental damages of air pollutants, see, e.g., Fowlie and Muller (2013) and Chan, Chupp, Cropper, and Muller (2015).

[70] For the detailed derivation of the discounted average damage $\tilde{d}_j$, see the Online Appendix D.4.

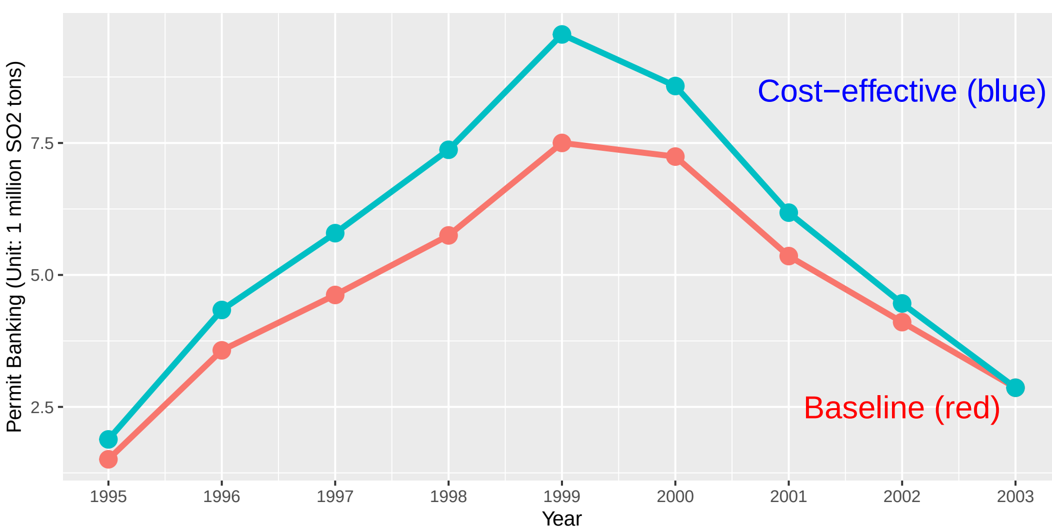
scenario D^{cf} can be decomposed as follows:

$$\begin{aligned} D^{cf} - D^{base} &= \sum_{t=1995}^{2003} \tilde{d}_t^{\,cf} E_t^{cf} - \sum_{t=1995}^{2003} \tilde{d}_t^{\,base} E_t^{base} \\ &= \sum_{t=1995}^{2003} \left(\tilde{d}_t^{\,cf} - \tilde{d}_t^{\,base} \right) E_t^{cf} + \sum_{t=1995}^{2003} \left(E_t^{cf} - E_t^{base} \right) \tilde{d}_t^{base}. \end{aligned} \tag{5.1}$$

The first term $\sum_{t=1995}^{2003} \left(\tilde{d}_t^{\,cf} - \tilde{d}_t^{\,base} \right) E_t^{cf}$ is the change caused by the intra-temporal (cross-sectional) distribution of emissions. Since $\tilde{d}_t$ is the weighted average of damages across generating units, it depends on the geographic distribution of emissions in each scenario.[71] Meanwhile, the second term $\sum_{t=1995}^{2003} \left(E_t^{cf} - E_t^{base} \right) \tilde{d}_t^{base}$ reflects the change in the inter-temporal distribution (i.e., aggregate emissions E_t^{cf} and E_t^{base} in each year).

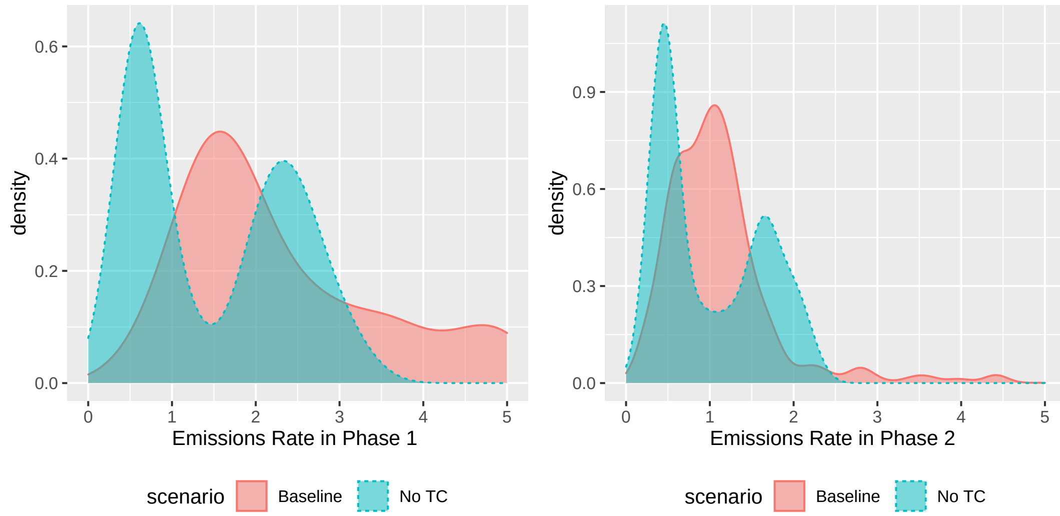
Table 6 reports the results of the decomposition exercise. The first term in Equation (6) accounts for \$1.990 billion and the second term \$447 million. Figure 8 shows the aggregate emissions in each year across different scenarios. Without transaction costs, the early abatement (i.e., lower emissions in Phase I) is achieved and thus contributes to the lower discounted damages from SO_2 emissions, as shown in Figure 8. Another key channel is the change in geographic distribution of SO2 emissions. SO_2 emissions are known as non-uniformly mixed pollution; health and environmental damages depend on the location of the emissions' source. The result suggests that more active trading of emissions permits leads to greater emissions in regions where the health damage is lower.

Figure 6: Permit Banking with and without Transaction Cost



[71]The difference between $\tilde{d}_t^{\,cf}$ and $\tilde{d}_t^{\,baseline}$ comes from the weight term $\frac{e_{jt}}{E_t}$, which reflects the cross-sectional distribution of emissions across generating units.

Figure 7: Distribution of Emissions Rate with and without Transaction Costs



Notes: Emissions rate of fuel is measured by lbs per MMBtu.

Table 5: Welfare Implications from Policy Simulations

	(1) Baseline		(2) Cost-effective		(3) No Banking		(4) Emissions Tax	
Emissions (1 million tons)	52.71	(1.27)	52.71	(1.27)	52.38	(1.07)	52.71	(1.27)
Banking at the terminal (1 million tons)	2.86	(1.12)	2.86	(1.12)	2.49	(0.86)	n.a.	n.a.
Total costs ($1 million)	93882	(210)	93068	(137)	94207	(210)	93109	(137)
Change from baseline ($1 million)			-814	(170)	325	(24)	-773	(170)
Total damages ($1 million)	42131	(1038)	39694	(962)	43054	(867)	39026	(946)
Change from baseline ($1 million)			-2437	(287)	923	(308)	-3105	(290)
Average damage ($1 per 1 ton)	799	(5)	753	(0)	822	(4)	740	(0)

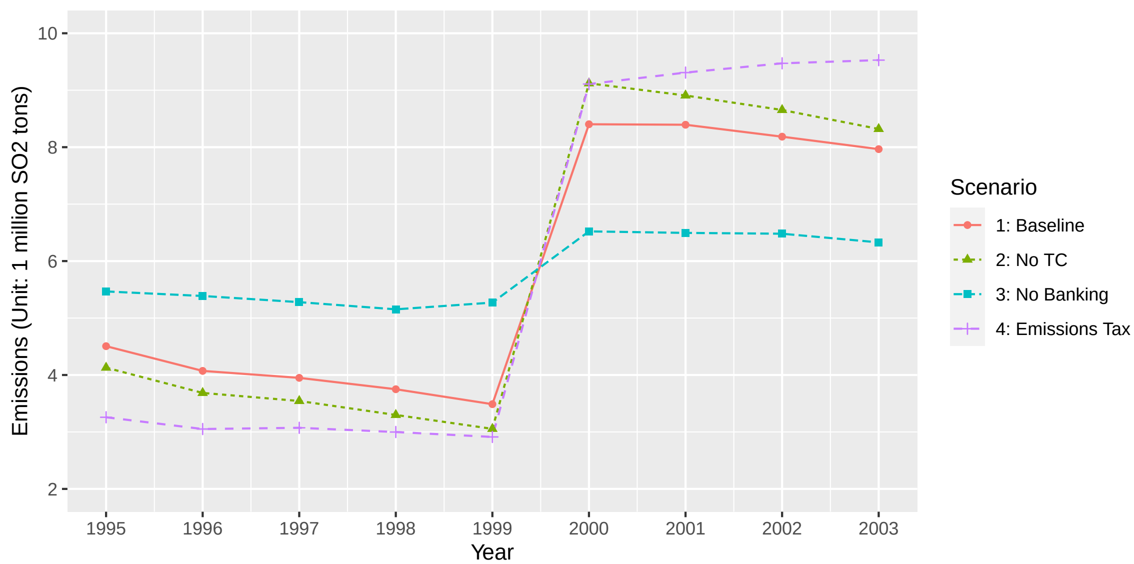
Notes: The numbers are the totals from 1995 to 2003. The units for emissions, left-over permits, and banking at the terminal period are 1 million SO_2 tons. Standard errors of simulation results are reported in parentheses.

Table 6: Decomposition of the Change in Health and Environmental Damages

	Cost-effective	No Banking b.w. Phases	Emissions Tax
(1) Intra-temporal effect: Average damage	-1990	-325	-1937
(2) Inter-temporal effect: Timing of emissions	-447	1247	-1169
Total	-2437	923	-3105

Notes: The first and second rows refer to $\sum_{t=1995}^{2003}\left(\tilde{d}_t^{\,cf}-\tilde{d}_t^{\,base}\right)E_t^{cf}$ and $\sum_{t=1995}^{2003}\left(E_t^{cf}-E_t^{base}\right)\tilde{d}_t^{base}$ in Equation (5.1), respectively. The unit is $1 million.

Figure 8: Path of Aggregate Emissions across Different Scenarios



5.2 Implications of the Permit Banking

A key feature of the Acid Rain Program is the permit banking system and the per-announced schedule of permit allocation, which decreased from Phase II (2000). To investigate how the permit banking system affects equilibrium outcomes, I simulate the case in which permit banking between Phases I and II is not allowed. In this simulation, firms cannot carry over emissions permits from 1999 to 2000, but they can bank in other years. This setting mimics the institutional setting of the EU–ETS.[72] Again, I fixed the level of fringe supply in this simulation.

Table 5 shows that the permit banking reduces the total production cost by \$325 million. The larger welfare gain is achieved through the reduction in environmental damages. The permit banking reduces the damages by \$923 million. The change in the damage is mostly due to the shift in the timing of emissions. Applying the decompression of Equation (5.1), I find that the decrease in health and environmental damages stems from the change in the first term in Equation (5.1). Permit banking achieves the earlier reduction of emissions (i.e., the lower level of emissions in Phase I), and thus, the discounted damage is lower.[73] The average health damage does not vary drastically from the baseline, indicating that the geographic distribution of SO_2 emissions is not greatly affected by the permit banking.[74]

[72]In the EU–ETS, permit banking between Phase 1 (2005–07) and Phase 2 (2008–12) was not allowed.

[73]An important caveat regarding this analysis is the assumption of calculating the SO_2 damages. Specifically, I assume that the damage from SO2 is linear with respect to emissions by following the assumption of Muller and Mendelsohn (2009b). If I consider the nonlinear effects of damages from SO2 emissions, the benefit of early reduction would be larger. The benefit of the early abatement of emissions in my analysis is a result of the discounting of the damages. Thus, my calculation of the benefit of early abatement should be taken as a lower bound. I believe this would be a fruitful extension of the analysis.

[74]The total amount of emissions over the period is slightly smaller in this simulation than that of other

5.3 Emissions Tax as an Alternative Policy

In this subsection, I simulate emissions tax policy as an alternative to cap-and-trade regulation. Instrument choice in environmental regulation has been an important topic in policy debate and academic literature since the work of Weitzman (1974). In the context of this paper, emissions tax has both advantages and disadvantages over cap-and-trade. On the one hand, emissions tax is not subject to the presence of transaction costs. As long as firms fully consider emissions tax as the shadow cost of emissions, the marginal abatement cost would be equalized across firms. On the other hand, emissions tax does not necessarily allow for the inter-temporal smoothing of abatement costs. Under an emissions tax policy with a fixed tax rate, firms make their abatement decisions so that the marginal abatement cost remains constant over time and is equal to the tax rate. However, when permit banking is permitted, as demonstrated in Equations (3.8) and (3.9) in Section 3.4, the marginal abatement cost increases at the rate of the inverse of discount factor (i.e., β^{-1}). This indicates that firms can spread out their abatement costs across periods and achieve a more gradual increase in their abatement costs.[75]

In this exercise, I simulate the impact of an emissions tax policy. To make a fair comparison between cap-and-trade and emissions tax, I set the emissions tax that achieves the same level of aggregate emissions in the baseline cap-and-trade policy.[76] One important caveat is that this counterfactual exercise does not aim to address the original "price vs. quantity" argument proposed by Weitzman (1974).[77] Instead, the primary goal of my analysis is to compare the welfare outcomes of these two policies that are not equivalent due to transaction costs and permit banking.

Table 5 shows that the total cost under the constant emissions tax policy is \$93.10 billion. This number is comparable the one under cost-effective outcome (\$93.07 billion). This result

simulations. Specifically, the discrepancy in the total amount of emissions between these two simulations is 0.6%. Therefore, I believe it would not affect the qualitative conclusion of the analysis. There are two reasons why this is the case. First, some emissions permits expire between Phase 1 and Phase 2 because I do not allow for the banking between two phases in this setting. Specifically, the remaining permits must expire if a firm does not participate in permit trading (i.e., a firm cannot sell permits). Even though the firm participates, the marginal revenue from selling permits could be less than zero, owing to transaction costs. Firms do not sell all of their remaining permits in such a case. Another reason is that the amount of banked permits at the terminal period (i.e., 2003) is different between the baseline and the non-banking simulation.

[75] Note that this argument depends on the schedule of emissions tax. In the emissions tax simulation, I consider the constant tax rate over time. The simulation of eliminating transactions costs in Section 5.1 can be interpreted as a time-varying emissions tax policy in which the tax rate increases at the rate of β^{-1} (i.e., the Hotelling rule).

[76] In the simulation, the continuation value function is needed because it depends on the scrubber rate and thus it affects the scrubber decision. Therefore, I need to set the level of permit banking in 2003 that appears in the continuation value function. Here I set the level at the one in the baseline case.

[77] The key focus of Weitzman (1974) is the role of the uncertainty in abatement costs and benefits, which makes (static) cap-and-trade (i.e., quantity instrument) and emissions tax (i.e., price instrument) no longer equivalent. Specifically, Weitzman (1974) shows that the relative slopes of the marginal abatement cost curve and the marginal benefit curve determine the comparative advantage of price instrument over quantity instrument. As discussed in 6.2, the current model has assumed away the uncertainty in the abatement cost, so I cannot directly address this point in the existing framework. I believe this exercise would be valuable as a future extension.

has two implications. First, the emissions tax policy is a useful tool that can achieve a similar degree of cost-effectiveness as that produced under cap-and-trade without transaction costs. The second implication pertains to the value of permit banking. Theoretically, the difference between the emissions tax and the cost-effective cap-and-trade is reduced to the presence of permit banking. The above result suggests that permit banking does not significantly improve the cost-effectiveness. However, as I discuss in Section 5.2, the permit banking improves the cost-effectiveness significantly when transaction costs do exist. These findings imply that the permit banking system mitigates the negative effect of transaction costs in cap-and-trade.

Regarding health and environmental damages, the emissions tax policy yields the smallest damage of \$39.03 billion among the four scenarios. As shown in Table 6, this reduction in damages is primarily attributed to the inter-temporal channel. Under the emissions tax policy, aggregate emissions during Phase I (1995-1999) were the lowest, while those during Phase II (2000-2003) were the highest, resulting in a lower discounted value of the aggregate damages over the sample period.

6 Caveats and Limitations

This section discusses some of the caveats and limitations of the paper. In particular, I discuss (1) modeling of production decision for power plants, (2) assumption of perfect foresight, (3) implications of the rate-of-return regulation on my analysis, and (4) policy design of permit allocation rule.

6.1 Model of Production Decision

This subsection provides further justifications for the exogenous electricity assumption and its drawbacks and implications for the analysis. As discussed in Section 3.2.2, I assume exogenous electricity production, implying that the unit-level electricity generation q_{jt} is fixed at the observed level. This assumption is based on the discussion in Section 2.3.2 that the reduction of output is not a major margin of emissions abatement. This approach also follows that adopted in previous studies including Carlson, Burtraw, Cropper, and Palmer (2000), Fowlie (2010b), and Chan (2015), where the primary margin of abatement is reducing the emissions rate by changing coal quality or adopting abatement technology.

However, this approach suffers from some drawbacks that are worthy of discussion. First, the exogenous production assumption does not consider the reallocation of productions across generating units as an abatement option. Specifically, introducing cap-and-trade could increase the marginal cost of generation by penalizing its SO_2 emissions, leading to lower utilization of coal units and higher utilization of gas units. Secondly, my model implicitly assumes away competition in the output (electricity) market.[78] This implies that

[78]Some papers have studied the interaction between market competition in the output market and environmental regulation (Mansur, 2007; Fowlie, 2009; Fowlie, Reguant, & Ryan, 2014). Incorporating

the electricity price does not change with the introduction of cap-and-trade.

To assess the importance of these concerns and the validity of my approach, I conduct an extensive descriptive analysis. The details and results can be found in the Online Appendix A. First, in the Online Appendix A.1, I run a difference-in-differences regression to estimate the impact of the introduction of cap-and-trade on the utilization rate of generating units. I found that the utilization rate decreased by only 0.6–3.8 percentage points after a cap-and-trade program was introduced. This finding is consistent with the previous argument that the primary method of emissions abatement centers around the reduction of the emissions rate.

Secondly, I examine how cap-and-trade could affect electricity prices. In the Online Appendix A.2, I conduct a regression analysis to determine the effects of the Acid Rain Program on output prices. The effects are both economically and statistically insignificant (see Table A2). This finding might be because the increase in the marginal cost due to cap-and-trade is limited. Table A3 in the Online Appendix shows that the permit cost (i.e., the cost of SO_2 emissions evaluated at the observed price of emissions permits) accounts for only 6.2–7.9% of the marginal cost of coal units.[79]

Lastly, I examine the reallocation of electricity generation across units in the Online Appendix A.3. Following the literature (e.g., Borenstein, Bushnell, and Wolak (2002), Bushnell, Mansur, and Saravia (2008), and Asker, Collard-Wexler, and De Loecker (2019)), I construct a simple model of the production decision and conduct a simulation analysis. In the model, a generating unit has a constant marginal cost and can produce output up to its capacity. The firm utilizes its generation units from the lowest to the highest cost of generation. The marginal cost of electricity generation is the sum of the fuel cost and the cost of SO_2 emissions. The latter depends on the shadow value of emissions permits. I explore how the shadow value of emissions permits affects the allocation of production across units. I find that the share of coal and gas units is almost constant with the shadow values (see Figure A2). This result is a consequence of the significant difference in the marginal cost across coal and gas units (see Table A3). On the other hand, the reallocation within coal units (from dirtier coal units to cleaner coal units) may occur to some extent when the

imperfect competition in the output market would complicate the analysis in several aspects. First, I would need to consider the impact of environmental regulation on consumer surplus in the electricity market. As Buchanan (1969) has observed theoretically and Fowlie, Reguant, and Ryan (2014) have discussed empirically, the introduction of corrective tax (i.e., Pigouvian taxation or equivalently cap-and-trade regulation) exaggerates the welfare loss that already exists due to market power. Secondly, market structure affects a firm's abatement incentive. With market power, firms may prefer to abate emissions by reducing output because such output reduction is partially mitigated by an increase in output price. Furthermore, imperfect competition in the output market will affect abatement incentives through coal quality choice and a scrubber. However, on aggregate, the implications of imperfect competition on aggregate emissions are ambiguous, as noted by Mansur (2007). The aggregate impact depends on technology substitution between dominant firms and fringe firms. I leave this issue open for future work.

[79]I also construct a state-level merit-order curve to see the probability of being marginal units. The percentage of coal units being marginal is around 81%. While this finding might concern that the introduction of the cap-and-trade could affect the output price determined by the marginal unit on a merit-order curve, its magnitude is limited given that the permit cost is relatively small compared to the fuel cost.

shadow value of permits increases (see Figure A3).

While this descriptive analysis supports my modeling approach regarding production decisions, it is worth examining how relaxing this assumption would affect my results. Since this assumption excludes one abatement option (i.e., decreasing electricity generation), the abatement cost implied from my model is larger. In the context of the Acid Rain Program and SO_2 abatement, however, I believe this bias should be limited because the major abatement option is fuel switching and installing a scrubber. On the other hand, if I consider applying my framework to CO2 regulations, I would need to incorporate a richer production decision model into the framework because the production reallocation is an essential margin of emissions reduction in the CO2 emissions.

6.2 Perfect Foresight and No Aggregate Uncertainty

My framework assumes away the aggregate uncertainty regarding profit π_t, namely, electricity demand and production costs.[80] Under this assumption, the permit prices P_t are deterministic objects in equilibrium and therefore firms should have perfect foresight over permit prices. Thus, the state variables that appear in a dynamic decision problem are permit holding h_{it}, trading experience I_{it}, and the firm-level removal rate of a scrubber α_{it}.[81]

Although perfect foresight is certainly a strong assumption, it is often imposed to keep tractability of the analysis in dynamic structural estimation (see, e.g., Lee, 2005; Conlon, 2012; Igami, 2017). This subsection discusses the difficulty in incorporating aggregate uncertainty in my model and the potential implication for my analysis.

The difficulty in incorporating a stochastic transition of demand, costs, and permit prices can be attributed to the dimensionality of the state space. In addition to $(h_{it}, I_{it}, \alpha_{it})$, I would have to consider the transition of the profit function π_{it} and the permit price P_t.[82] Because my framework incorporates the rich heterogeneity of regulated firms, I must solve the dynamic optimization problem separately for each firm.[83] Therefore, expanding the state

[80] I also assume perfect foresight with regard to the permit allocation a_{it}. This assumption reflects the fact that the permit allocation schedule was announced before the Acid Rain Program started and did not change during my sample period.

[81] The value function V_t is indexed by time script t and subsumes the exogenous state variables.

[82] If I rewrite the profit function as $\pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) \equiv \pi_i\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}; D_t, C_t\right)$, where D_t is the aggregate demand shock and C_t is the aggregate cost shock, the additional state variables are (D_t, C_t, P_t), yielding six state variables.

[83] A theoretical study by Schennach (2000) analyzes cap-and-trade programs in a dynamic environment with uncertainty by adopting a social planner's approach. When permit trading is allowed without any frictions (i.e., transaction costs and information asymmetry), the decentralized outcome of cap-and-trade is equal to the social planner's solution that allocates emissions permits efficiently and minimizes the total abatement costs. Thus, the paper can reduce the dynamic decision problems of heterogeneous firms into a single social planner problem. Moreover, the Coase theorem implies that the distribution of initial allocation and permit holding does not affect the abatement outcomes. Thus, the distribution of permit holding is not needed to be tracked as a state variable and the dimensionality of the state space is lower. This approach, however, cannot be applied once I depart from the basic assumptions in Schennach (2000). In my paper, transaction costs do not allow me to follow such a social planner's approach. See also Cantillon and Slechten (2018), who show that the social planner approach may not apply when there is asymmetric information about abatement costs.

space would make it extremely difficult to compute and estimate the model.

Another issue relates to how to model the transition of equilibrium permit prices, which is conceptually and computationally difficult. With the aggregate uncertainty of demand and costs, the equilibrium permit prices become random variables. Thus, firms must form an expectation regarding future equilibrium permit prices. In a rational expectation, firms must track all information that forecasts the future permit prices P_{t+1}. Permit prices are determined by the trading decisions of all firms in equilibrium. Therefore, in principle, firms must have knowledge of the cross-sectional distribution of the state variables in the state space in order to form a rational expectation of future permit prices. Because there are 138 firms in my sample, this approach would be infeasible due to the curse of dimensionality. This issue has been observed before in the literature by Krusell and Smith (1998) on a heterogeneous macro model, Lee and Wolpin (2006) for the structural estimation of a general equilibrium labor model, and Gillingham, Iskhakov, Munk-Nielsen, Rust, and Schjerning (2015, 2019) for a dynamic demand model for new and used car markets.

Given these complications, I choose to assume perfect foresight. Thanks to this assumption, the model can incorporate firm heterogeneity, which is an important consideration in the analysis of cap-and-trade, while still being tractable for estimation and simulation analyses. Such a simplifying assumption, however, is not completely innocuous. I discuss how this assumption affects the model prediction and thus policy simulations.

The assumption would affect the model prediction of scrubber investment and permit banking. As Dixit, Dixit, and Pindyck (1994) note, the degree of uncertainty, along with the irreversible nature of investment goods, reduces the level of investment. Thus, my model predicts over-investment in a scrubber. Secondly, with uncertainty, regulated firms bank more permits due to the precautionary motive, in addition to the inter-temporal smoothing of abatement costs, as was highlighted by my model.

In policy simulations, the improvement of cost-effectiveness due to the permit banking system would be underestimated under perfect foresight. Faced with uncertainty, though, firms have additional incentives to bank permits because they need to hedge future uncertainty. Thus, the role of permit banking in improving the cost-effectiveness would be more significant under uncertainty. The cost-saving due to the permit banking in my analysis can be considered as the lower bound.

Extending my framework to incorporate uncertainty is a challenging task yet it provides a fruitful direction to analyze the implications of cap-and-trade. Recent cap-and-trade programs struggle with the uncertainty that stems from various factors (see, e.g., Borenstein, Bushnell, Wolak, and Zaragoza-Watkins, 2019). I leave this matter open to future research. [84]

[84] There are two potential approaches to deal with this problem. The first approach, taken by Gillingham, Iskhakov, Munk-Nielsen, Rust, and Schjerning (2015), is to use the concept of temporary equilibrium. This approach assumes that firms have stationary expectations of future prices of emissions permits; in other words, firms expect permit prices tomorrow to be the same as they are today. Under this expectation, the equilibrium prices of permits are computed such that the markets are cleared in every period. The other

6.3 Rate-of-Return Regulation

The analysis in this paper does not explicitly consider the presence of public utility regulation (i.e., rate-of-return regulation). The literature has examined how a public utility regulation affects the outcomes of market-based environmental regulations. Such analysis includes the Averch–Johnson effects on compliance decisions (Fowlie, 2010b in the NO_x regulation and Cicala, 2015 in the SO_2 regulation), the regulatory uncertainty associated with the cost-recovery rule of permits (see, e.g., Montero, 1998 and Arimura, 2002), and inefficiency in production decision (Abito, 2020). To fully account for these points, I would need to incorporate the rate-of-return regulation, which adds further complications to the model.[85]

I discuss the potential impacts of the rate-of-return regulation on my analysis in terms of the Averch–Johnson effects and inefficiency in production decisions.[86] First, the Averch-Johnson effects imply that the regulated utilities tend to invest in capital-intensive equipment to expand their rate base. In the current context, the utilities prefer to install a scrubber instead of trading permits for their compliance purpose. Thus, the model that assumes away the Averch-Johnson effect would under-predict installment of a scrubber. Secondly, Abito (2020) shows that the rate-of-return regulation would distort incentives to operate in an efficient manner. Such sub-optimal behavior would lead to higher abatement costs, which implies that the model that ignores this effect would lead to lower abatement costs.

6.4 Policy Design of Permit Allocation Rule

Investigating the implications of the permit allocation rule is worthwhile according to my framework. The Coase theorem states that, in the absence of transaction costs and other conditions being satisfied, the way the regulator allocates emissions permits at the beginning does not affect the equilibrium outcome of emissions.[87] This does not hold true in my setting, however, due to the presence of transactions costs. Thus, the initial allocation of permits affects equilibrium outcomes and the welfare implications of cap-and-trade policy. Analyzing the optimal allocation policy in the current framework is an interesting extension.

approach, followed by Krusell and Smith (1998), parameterizes the expectation of future permit prices as a function of a small set of "sufficient statistics," such as the aggregate demand shock, cost shock, and permit prices in the current period. It then determines the parameters of the expectation process that yield the smallest excess demand of emissions permits across periods. The drawback of this approach is that, although consumers have expectations about future permit prices that are consistent with realized prices, the market clearing may not be satisfied in a given period.

[85] See Lim and Yurukoglu (2018), who structurally estimate a model of rate-of-return regulation in a dynamic setting.

[86] The regulatory uncertainty in the cost-recovery rules has a nuanced impact on permit trading. On the one hand, if the regulated utilities can include the cost of purchasing permits in the rate base and their allowed rate of return is higher than their cost of capital, utilities are encouraged to purchase permits for their compliance. On the other hand, many public utility commissions were unclear regarding their policies toward permit trades in their cost recovery rule, which discourages utilities from relying on permit trading as a compliance strategy. See, e.g., Montero (1998) and Arimura (2002) for further discussions.

[87] For example, Footnote 1 in Fowlie and Perloff, 2013 provides the detailed conditions for the Coase theorem to be satisfied. See e.g., Hahn and Stavins (2011) for a survey of the implication of allowance allocations in cap-and-trade regulation.

However, this is not a straightforward exercise. Indeed, numerically analyzing the optimal initial allocation would be infeasible because the number of choice variables (i.e., the initial permit allocation for each firm) is too high.[88] Therefore, a theoretical approach is required to study the optimal design of initial permit allocation under the presence of transactions costs in a dynamic framework. I leave this for future work.[89]

Another policy design that may improve the cost efficiency of cap-and-trade is "depreciating permits," which was initially proposed as the depreciating license mechanism by Weyl and Zhang (2018). Under the depreciating permits system, the regulator collects a certain fraction of banked permits at the end of the period and then reallocates them through auctions. As a result, firms have a greater incentive to trade in the market, owing to the depreciation, but can still smooth their abatement costs over time using permit banking. Although this is an interesting extension, implementing this policy model requires additional modeling and consideration. The key to depreciating license mechanism, according to Weyl and Zhang (2018), is to reallocate depreciated license through an auction mechanism, which requires additional modeling in my framework.

These two policy counterfactual simulations are a fruitful extension of the paper, though they require additional considerations. I therefore leave this topic open for further discussion in future work.

7 Conclusion

This study examines the dynamic incentives of firms regulated by a cap-and-trade program in the context of SO_2 emissions regulations in the US electricity industry. I construct a dynamic equilibrium model of a cap-and-trade program, in which firms make decisions regarding abatement investment, permit trading, and banking. I apply the model to data from the US Acid Rain Program and estimate the model primitives. My estimates suggest that the variable transaction costs associated with permit trading are substantial. Through policy simulations, I find that the total costs could be reduced by $814 million in total without any transaction costs. This additional cost saving is achieved by more active trading of permits and a more cost-effective pattern of abatement. The emissions tax policy can achieve the outcome closer to the cost-effective one. Lastly, I examined the role of a permit banking system, finding that it helps firms to mitigate the negative effect of transactions costs. Indeed, permit banking induces earlier abatement of emissions, which contributes to lower health and environmental damages.

[88] In principle, I could search for an initial allocation that minimizes some objective function (i.e., the total abatement costs, the health and environmental damages, or the sum of these two). Such computation, however, would be infeasible for the following two reasons. First, the number of choice variables is too high, as I need to consider firm-level initial allocation in each phase. Second, to evaluate the outcome under a particular candidate of initial allocation, I must solve a dynamic competitive equilibrium.

[89] Fowlie, Reguant, and Ryan (2016) empirically study the implications of an initial allocation scheme under the presence of market power and emissions leakage.

The proposed framework can be applied beyond a cap-and-trade program on air pollutants. Governments now use a market-based policy in various settings, including credit trading in the CAFE regulation and Renewable Energy Certificates in the Renewable Portfolio Standard (RPS).[90] Under these policies, firms face a similar problem to that examined here: they can either trade these credits, or invest in technology (i.e., improve fuel efficiency in the CAFE credit trading, or build renewable generators in the RPS program). The proposed empirical framework can be used to study the effectiveness of these market-based policies and the implications of alternative regulatory designs. Future research can further develop these topics.

References

Abito, Jose Miguel (2020). "Measuring the welfare gains from optimal incentive regulation". In: *The Review of Economic Studies* 87.5, pp. 2019–2048.

Aguirregabiria, Victor and Pedro Mira (2002). "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models". In: *Econometrica* 70.4, pp. 1519–1543.

— (2007). "Sequential estimation of dynamic discrete games". In: *Econometrica* 75.1, pp. 1–53.

— (2010). "Dynamic discrete choice structural models: A survey". In: *Journal of Econometrics* 156.1, pp. 38–67.

Arimura, Toshi H (2001). "How Much Can We Save From Emissions Permit Banking? The Case Of The U.s. So2 Allowance Market". In: *Working Paper*.

— (2002). "An Empirical Study of the So2 Allowance Market: Effects of PUC Regulations". In: *Journal of Environmental Economics and Management* 44.2, pp. 271–289.

Asker, John, Allan Collard-Wexler, and Jan De Loecker (2019). "(Mis) allocation, market power, and global oil extraction". In: *American Economic Review* 109.4, pp. 1568–1615.

Bajari, Patrick, C Lanier Benkard, and Jonathan Levin (2007). "Estimating Dynamic Models of Imperfect Competition". In: *Econometrica* 75.5, pp. 1331–1370.

BGC Environmental Brokerage Services (2011). *SO2 ALLOWANCE PRICE INDICATIONS Historical Monthly Bulletins*. URL: `http://www.bgcebs.com/Emissions/?page=USAComp_MarketData-BulletinsHistoric`.

Borenstein, Severin, James B Bushnell, and Frank A Wolak (2002). "Measuring market inefficiencies in California's restructured wholesale electricity market". In: *The American Economic Review* 92.5, pp. 1376–1405.

Borenstein, Severin et al. (2019). "Expecting the unexpected: Emissions uncertainty and environmental market design". In: *American Economic Review* 109.11, pp. 3953–77.

[90] See, e.g., Lade, Lin Lawell, and Smith (2018) for a theoretical model for the Renewable Fuel Standard.

Breen, William J, Laurie Simon Hodrick, and Robert A Korajczyk (2002). “Predicting equity liquidity”. In: *Management Science* 48.4, pp. 470–483.

Buchanan, James M (1969). “External diseconomies, corrective taxes, and market structure”. In: *The American Economic Review* 59.1, pp. 174–177.

Burtraw, Dallas (2000). *Innovation under the tradable sulfur dioxide emission permits program in the US electricity sector*. Tech. rep.

Burtraw, Dallas and Erin Mansur (1999). “Environmental Effects of SO2 Trading and Banking”. In: *Environmental Science and Technology* 33, 3489r3494.

Bushnell, James B, Erin T Mansur, and Celeste Saravia (2008). “Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets”. In: *The American Economic Review* 98.1, pp. 237–266.

Cantillon, Estelle and Aurélie Slechten (2018). “Information aggregation in emissions markets with abatement”. In: *Annals of Economics and Statistics/Annales d’Économie et de Statistique* 132, pp. 53–79.

Carlson, Curtis et al. (2000). “Sulfur dioxide control by electric utilities: What are the gains from trade?” In: *Journal of political Economy* 108.6, pp. 1292–1326.

Chan, H Ron et al. (2015). *The Impact of Trading on the Costs and Benefits of the Acid Rain Program*. Tech. rep. National Bureau of Economic Research.

Chan, Ron (2015). “How Large are the Cost Savings from Emissions Trading? An Evaluation of the U.S. Acid Rain Program”. In: *Working Paper*.

Chen, Cuicui (2018). *Slow focus: Belief evolution in the US acid rain program*. Tech. rep. Working Paper 1–49.

Cicala, Steve (2015). “When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation”. In: *American Economic Review* 105.1, pp. 411–44.

Coase, Ronald H (1960). *The problem of social cost*. Springer.

Collard-Wexler, Allan (2013). “Demand Fluctuations in the Ready-Mix Concrete Industry”. In: *Econometrica* 81.3, pp. 1003–1037.

Conlon, Christopher T (2012). “A Dynamic Model of Prices and Margins in the LCD TV Industry”. In:

Cook, Jonathan and C-Y Cynthia Lin Lawell (2018). “Wind turbine shutdowns and upgrades in Denmark: Timing decisions and the impact of government policy”. In: *Available at SSRN 3336341*.

Dardati, Evangelina (2016). “Pollution permit systems and firm dynamics: how does the allocation scheme matter?” In: *International Economic Review* 57.1, pp. 305–328.

Dávila, Eduardo and Cecilia Parlatore (2017). “Trading Cost and Informational Efficiency”. In:

Dixit, Avinash K, Robert K Dixit, and Robert S Pindyck (1994). *Investment under uncertainty*. Princeton university press.

EIA (2012). *Electric Power Annual 2012*. Tech. rep.

Ellerman, A Denny, Paul L Joskow, and David Harrison Jr (2003). "EMISSIONS TRADING IN THE U.S.: EXPERIENCE, LESSONS, AND CONSIDERATIONS FOR GREENHOUSE GASES". In: *Prepared for the Pew Center on Global Climate Change*.

Ellerman, A Denny and Juan-Pablo Montero (2007). "The efficiency and robustness of allowance banking in the US Acid Rain Program". In: *The Energy Journal*, pp. 47–71.

Ellerman, A Denny et al. (2000). *Markets for clean air: The US Acid Rain Program*. Cambridge University Press.

Energy Information Administration (1990-2013). *Form EIA-860 detailed data*. URL: https://www.eia.gov/electricity/data/eia860/.

— (1990–2003). *Historic Form EIA-423 & FERC-423 Detailed Data*. (last accessed: April 29, 2023). URL: https://www.eia.gov/electricity/data/eia423/.

— (1985–2005). *EIA-767 Steam-Electric Plant Operation and Design Report*. (last accessed: April 29, 2023). URL: https://www.eia.gov/electricity/data/eia767/.

Fabra, Natalia and Mar Reguant (2014). "Pass-Through of Emissions Costs in Electricity Markets". In: *American Economic Review* 104.9, pp. 2872–99.

Federal Register (Jan. 2004). *Rule To Reduce Interstate Transport of Fine Particulate Matter and Ozone (Interstate Air Quality Rule)*. https://www.federalregister.gov/d/04-808.

Fowlie, Meredith (2010a). "Allocating emissions permits in cap-and-trade programs: Theory and evidence". In: *Working Paper*.

— (2010b). "Emissions trading, electricity restructing, and investment in pollution abatement". In: *The American Economic Review*, pp. 837–869.

Fowlie, Meredith L (2009). "Incomplete environmental regulation, imperfect competition, and emissions leakage". In: *American Economic Journal: Economic Policy* 1.2, pp. 72–112.

Fowlie, Meredith and Nicholas Muller (2013). *Market-based emissions regulation when damages vary across sources: What are the gains from differentiation?* Tech. rep. Working Paper.

Fowlie, Meredith and Jeffrey M Perloff (2013). "Distributing pollution rights in cap-and-trade programs: are outcomes independent of allocation?" In: *Review of Economics and Statistics* 95.5, pp. 1640–1652.

Fowlie, Meredith, Mar Reguant, and Stephen P Ryan (2014). "Market-based emissions regulation and industry dynamics". In: *forthcoming in Journal of Political Economy*.

— (2016). "Market-based emissions regulation and industry dynamics". In: *Journal of Political Economy* 124.1, pp. 249–302.

Gârleanu, Nicolae and Lasse Heje Pedersen (2013). "Dynamic trading with predictable returns and transaction costs". In: *The Journal of Finance* 68.6, pp. 2309–2340.

Gillingham, Kenneth et al. (2015). "A Dynamic Model of Vehicle Ownership, Type Choice, and Usage". In:

— (2019). *Equilibrium trade in Automobile Markets*. Tech. rep. National Bureau of Economic Research.

Gollop, Frank M and Mark J Roberts (1985). “Cost-minimizing regulation of sulfur emissions: Regional gains in electric power”. In: *The Review of Economics and Statistics*, pp. 81–90.

Hagerty, Nick (2019). “Liquid constrained in California: Estimating the potential gains from water markets”. In: *Working Paper*.

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

Helfand, Gloria E, Michael R Moore, and Yimin Liu (2006). “Testing for dynamic efficiency of the Sulfur Dioxide Allowance Market”. In:

Hotz, V Joseph and Robert A Miller (1993). “Conditional choice probabilities and the estimation of dynamic models”. In: *The Review of Economic Studies* 60.3, pp. 497–529.

Igami, Mitsuru (2017). “Estimating the innovator’s dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998”. In: *Journal of Political Economy* 125.3, pp. 798–847.

Jaraitė-Kažukauskė, Jūratė and Andrius Kažukauskas (2015). “Do transaction costs influence firm trading behaviour in the european emissions trading system?” In: *Environmental and Resource Economics* 62.3, pp. 583–613.

Jørgensen, Thomas H and Maxime Tô (2020). “Robust Estimation of Finite Horizon Dynamic Economic Models”. In: *Computational Economics* 55.2, pp. 499–509.

Joskow, Paul L and Richard Schmalensee (1998). “The Political Economy of Market-Based Environmental Policy: the US Acid Rain Program 1”. In: *The journal of law and economics* 41.1, pp. 37–84.

Joskow, Paul L, Richard Schmalensee, and Elizabeth M Bailey (1998). “The market for sulfur dioxide emissions”. In: *American Economic Review*, pp. 669–685.

Kalouptsidi, Myrto (2014). “Time to build and fluctuations in bulk shipping”. In: *The American Economic Review* 104.2, pp. 564–608.

Keane, Michael P and Kenneth I Wolpin (2001). “The effect of parental transfers and borrowing constraints on educational attainment”. In: *International Economic Review* 42.4, pp. 1051–1103.

Keohane, Nathaniel (2006). “Cost savings from allowance trading in the 1990 Clean Air Act: Estimates from a choice-based model”. In: *Moving to markets in environmental regulation: Lessons from twenty years of experience*, pp. 194–229.

Kolstad, Jonathan T and Frank A Wolak (2008). “Using Environmental Emissions Permit Prices to Raise Electricity Prices: Evidence from the California Electricity Market”. In: *Working Paper*.

Krusell, Per and Anthony A Smith Jr (1998). “Income and wealth heterogeneity in the macroeconomy”. In: *Journal of political Economy* 106.5, pp. 867–896.

Lade, Gabriel E, C-Y Cynthia Lin Lawell, and Aaron Smith (2018). “Policy shocks and market-based regulations: Evidence from the Renewable Fuel Standard”. In: *American Journal of Agricultural Economics* 100.3, pp. 707–731.

Lee, Donghoon (2005). "An estimable dynamic general equilibrium model of work, schooling, and occupational choice". In: *International Economic Review* 46.1, pp. 1–34.

Lee, Donghoon and Kenneth I Wolpin (2006). "Intersectoral labor mobility and the growth of the service sector". In: *Econometrica* 74.1, pp. 1–46.

Lillo, Fabrizio, J Doyne Farmer, and Rosario N Mantegna (2003). "Econophysics: Master curve for price-impact function". In: *Nature* 421.6919, pp. 129–130.

Lim, Claire SH and Ali Yurukoglu (2018). "Dynamic natural monopoly regulation: Time inconsistency, moral hazard, and political environments". In: *Journal of Political Economy* 126.1, pp. 263–312.

Liski, Matti and Juan-Pablo Montero (2011). "Market Power in an Exhaustible Resource Market: The Case of Storable Pollution Permits". In: *The Economic Journal* 121.551, pp. 116–144.

Mansur, Erin T (2007). "Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market". In: *The Journal of Industrial Economics* 55.4, pp. 661–689.

— (2008). "Measuring welfare in restructured electricity markets". In: *The Review of Economics and Statistics* 90.2, pp. 369–386.

Montero, Juan-Pablo (1998). "Marketable pollution permits with uncertainty and transaction costs". In: *Resource and Energy Economics* 20.1, pp. 27–50.

Muller, Nicholas Z and Robert Mendelsohn (2009a). "Additional Materials for "Efficient pollution regulation: getting the prices right"". In: *The American Economic Review* 99.5. (last accessed: May 1, 2023), pp. 1714–1739. URL: `https://www.aeaweb.org/aer/data/dec09/20071380_app.zip`.

— (2009b). "Efficient pollution regulation: getting the prices right". In: *The American Economic Review* 99.5, pp. 1714–1739.

Reguant, Mar (2014). "Complementary bidding mechanisms and startup costs in electricity markets". In: *The Review of Economic Studies* 81.4, pp. 1708–1742.

Reguant, Mar and A Denny Ellerman (2008). "Grandfathering and the endowment effect: An Assessment in the context of the Spanish National Allocation Plan". In: *Working Paper*.

Robert, Engle, Ferstenberg Robert, and Russell Jeffrey (2012). "Measuring and modeling execution cost and risk". In: *The Journal of Portfolio Management* 38.2, pp. 14–28.

Rubin, Jonathan D (1996). "A model of intertemporal emission trading, banking, and borrowing". In: *Journal of Environmental Economics and Management* 31.3, pp. 269–286.

Rust, John (1987). "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher". In: *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

Ryan, Stephen P (2012). "The costs of environmental regulation in a concentrated industry". In: *Econometrica* 80.3, pp. 1019–1061.

Schennach, Susanne M (2000). "The economics of pollution permit banking in the context of Title IV of the 1990 Clean Air Act Amendments". In: *Journal of Environmental Economics and Management* 40.3, pp. 189–210.

Schmalensee, Richard and Robert N Stavins (2013). "The SO2 Allowance Trading System: The Ironic History of a Grand Policy Experiment". In: *The Journal of Economic Perspectives* 27.1, pp. 103–121.

Smith, Anne E, Jeremy B Platt, and A Denny Ellerman (1998). "The cost of reducing utility SO2 emissions: not as low as you might think". In:

Stavins, Robert N (1995). "Transaction costs and tradeable permits". In: *Journal of environmental economics and management* 29.2, pp. 133–148.

U.S. Bureau of Labor Statistics (1984–2015). *Producer Price Indexes*. URL: `https://www.bls.gov/ppi/`.

U.S. Census Bureau (2011). *Census Bureau Region and Division Codes and Federal Information Processing System (FIPS) Codes for States*. URL: `https://www2.census.gov/programs-surveys/popest/geographies/2011/state-geocodes-v2011.xls`.

U.S. Environmental Protection Agency (1996-2010). *The Emissions & Generation Resource Integrated Database (eGRID)*. URL: `https://www.epa.gov/egrid`.

— (1993–2013a). *Acid Rain Program Allowance Transactions Data*. URL: `https://campd.epa.gov/`.

— (1993–2013b). *Clean Air Markets Program Data (CAMPD)*. URL: `https://campd.epa.gov/`.

— (1993a). *Technical Documentation for Phase II Allowance Allocations*. Tech. rep. EPA-430-R-93-002.

— (1993b). "The National Allowance Data Base Version 2.1 Technical Support Document". In:

— (Dec. 2003). *Air Quality Proposal to Deeply Cut Power Plant Emissions is Signed*. `https://archive.epa.gov/epapages/newsroom_archive/newsreleases/b5264a20d8436a0685256dff005d9ce8 html`.

— (Sept. 2004). *Acid Rain Program 2003 Progress Report*. Tech. rep.

— (2009). *Continuous Emissions Monitoring Fact Sheet*. `https://web.archive.org/web/20090211082920/http://epa.gov/airmarkets/emissions/continuous-factsheet.html`. Retrieved from Internet Archive. Last accessed on October 14, 2021.

Weitzman, Martin L (1974). "Prices vs. quantities". In: *The review of economic studies* 41.4, pp. 477–491.

Weyl, E Glen and Anthony Lee Zhang (2018). "Depreciating Licenses". In: *Available at SSRN 2744810*.

Appendix For Online Publication

Dynamic Incentives and Permit Market Equilibrium in Cap-and-Trade Regulation

Yuta Toyama

A Online Appendix: Supplemental Analysis on Electricity Production Decision

In this Appendix, I provide a descriptive and simulation analysis on electricity production decisions. The purpose of these analyses is to examine the validity of exogenous production assumption in the structural model. Section 6.1 in the main body of the paper presents a summary of these analyses. Below, however, I provide their details. First, I conduct the difference-in-differences analysis to determine how the introduction of the Acid Rain Program affects the electricity generation in . Then, I run the regression analysis to estimate how the output prices respond to the regulation in A.2. Lastly, I examine the effect of the Acid Rain Program on the reallocation of production across generating units by constructing a merit-order curve in A.3.

A.1 Difference-in-differences Analysis on Electricity Generation

This section conducts a difference-in-difference (hereafter DID) regression to examine how the introduction of the Acid Rain Program affects the electricity production. To do this, I exploit the variation in the timing of the regulation across generating units. As was mentioned in Section 2.1, there are two groups of generating units: those regulated since 1995 (Group I units) and those regulated since 2000 (Group II units). Figure A1 shows the average capacity factor for each group each year.

While the data cover the period until 2003, all the units are treated (i.e., regulated) from 2000, implying that there are no control units after 2000. Following the practice in the recent DID literature (i.e., Goodman-Bacon 2021; Callaway and SantAnna 2021; Sun and Abraham 2021), I restrict the sample to the period between 1990 and 1999.[91]

I specify the regression equation as follows:

$$Y_{jm} = \alpha \text{GroupI}_j \cdot 1\{\text{after1995}\}_m + \beta' X_{jm} + u_j + u_m + u_{jm},$$

where Y_{jm} is the outcome variable of unit j in year-month m. I use as the main outcome variable capacity factor defined by $cf_{jm} = q_{jm}/k_j$, where q_{jm} is the net generation and k_j is

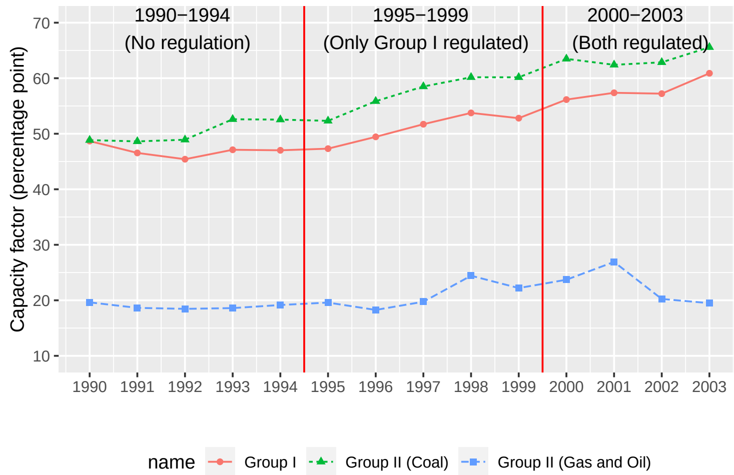
[91] Using the terminology of the recent DID literature, there are no "never-treated" units in the sample. The difference-in-difference regression using the sample after 2000 compares the Group II units with the Group I units, which are "already treated" units and thus are not adequate control units for the Group II units. See Goodman-Bacon 2021; Callaway and SantAnna 2021; Sun and Abraham 2021 for the details.

the nameplate capacity. GroupI_j is the dummy variable for the Group I units. $1\{\text{after1995}\}_m$ is the dummy variable that indicates the periods after 1995 (the beginning of Phase I). X_{jm} includes control variables such as the state-level electricity demand. Unit and time fixed effects are captured by u_j and u_m, respectively.

It is worth explaining the interpretation of the DID estimate of α. To estimate it, I compare the change in the utilization rate of the Group I units before and after 1995 and the change of the Group II units. Once the regulation is introduced in 1995, the Group I units have a higher marginal cost of generation due to the opportunity cost of SO_2 emissions under cap-and-trade. Such a cost increase might induce lower utilization of the Group I units. However, the Group II units could also change their electricity production via production reallocation. Specifically, the firm may reallocate the electricity production from Group I to Group II. Therefore, the DID estimate of α should be interpreted as an upper bound of the regulation effect on the Group I units.

The regression results are shown in Table A1. The estimates suggest that introducing the ARP decreases the capacity factor of group I units by 0.6–3.8 percentage points, depending on the choice of units in the control group. Although the effects are statistically significant, as shown in column (1), the economic significance is limited. Because the mean of the capacity factor falls within the range of 50–60 percentage points in the sample, electricity generation fell by at most 7.6% after the introduction of the cap-and-trade program. As shown in Section 2.3.1, this magnitude cannot account for the significant decrease in SO_2 emissions in the sample period. Instead, the adjustment of emissions rate of fuel, as discussed in Section 2.3.2, is the key channel of SO_2 abatement. The difference-in-differences analysis confirms that the abatement of SO_2 emissions was achieved primarily through adjusting emissions rates of fuel. This finding supports my modeling assumption of exogenous electricity production.

Figure A1: Trend of Capacity Factor of Group I and Group II units



Notes: The figure shows the trend of the capacity factor, defined by the ratio of net generation (output) to generation capacity, over time. I calculate the mean of the monthly-level capacity factor in each year for three groups: units regulated since 1995 (denoted as Group I), coal units regulated since 2000 (denoted as Group II (Coal)), and gas and oil units regulated since 2000 (denoted as Group II (Gas and Oil)).

Table A1: Difference-in-differences Regression of Capacity Factor

	Dependent variable:	
	Capacity factor	
	(1)	(2)
Treatment	−3.823	−0.643
	(0.633)	(0.568)
log(Electricity Demand))	38.481	42.705
	(1.698)	(1.182)
Control group	Coal only	Coal, Gas, and Oil
Observations	127,200	201,960
Adjusted R^2	0.527	0.656

Notes: Unit-level dummies and year-and-month dummies are included. Standard errors are clustered at the unit level.

A.2 Effects of Acid Rain Program on Output Prices

In this subsection, I examine how output prices change due to the introduction of the Acid Rain Program. Specifically, I run the following panel regression:

$$\log P_{st} = \theta Z_{st} + u_s + u_t + u_{st},$$

where P_{st} is the output price in state s in month-year t and Z_{st} is a measure of exposure to the Acid Rain Program. The measure of output price is defined as the average price for industrial/residential retail customers obtained from the form Form EIA-826 "Monthly Electric Utility Sales and Revenue Report with State Distributions" Energy Information Administration (1990–2003). For Z_{st}, I use (1) the number of generating units affected by the Acid Rain Program in state s and period t, (2) the total generation capacity of units that are regulated under the Acid Rain Program, and (3) the dummy variable indicating whether there exists at least one regulated unit in state s and period t. I use the data from 1990 to 2003. Note that there is both cross-sectional and temporal variation in the exposure to the Acid Rain Program.

Table A2 presents the estimation results. All the columns show that the impact on output price is both statistically and economically insignificant, which suggests that the introduction of the Acid Rain Program seems to have a limited impact on output prices.

A.3 Reallocation of Production across Generating Units based on a Merit-order Curve

I examine how the electricity generation would change if I consider the increase in the marginal cost of generation due to cap-and-trade. While the model assumes the exogenous electricity production, the cap-and-trade regulation might change the marginal costs of generation by incorporating the shadow cost of SO_2 emissions and thus the allocation of production across generating units. To investigate this point, I construct a merit-order curve based on the marginal cost and the capacity of generating units as well as examining how the shadow cost of SO_2 emissions changes the production pattern (e.g., Borenstein, Bushnell, and Wolak (2002), Bushnell, Mansur, and Saravia (2008), and Asker, Collard-Wexler, and De Loecker (2019)). I first introduce a model of electricity generation based on a merit-order curve in A.3.1 and show the summary statistics of fuel and permit costs in A.3.2. Using the model, I analyze the reallocation of electricity generation due to the introduction of the cap-and-trade in A.3.3. Lastly, I examine the extent to which coal units are likely to be a marginal unit in A.3.4.

Table A2: Effects of Exposure to the Acid Rain Program on State-level Electricity Price

	Dependent variable:		
	log (output price)		
	(1)	(2)	(3)
log(number of regulated units + 1)	−0.005		
	(0.008)		
log(capacity of regulated units + 1)		0.0001	
		(0.002)	
1at least one regulated unit			0.023
			(0.020)
Observations	9,671	9,671	9,671
Adjusted R^2	0.872	0.872	0.872

Note: State FE, Year FE, and Month FE are included. I add one to the inside of the logarithm because the covariates take 0 if there are no regulated units in a state in a particular time period. Standard errors are clustered at the state level.

A.3.1 Model

A generating unit can produce electricity with constant marginal cost mc_{jt} up to the capacity constraint of k_j. The marginal cost of unit j in time t is given by

$$mc_{jt}(\lambda) = p_{jt}^{fuel} \cdot HR_j + \lambda(1 - \alpha_{jt})R_{jt} \cdot HR_j.$$

Notations follow those in the main body of the paper. The first component of marginal cost $p_{jt}^{fuel} \cdot HR_j$ is fuel cost. The second component $\lambda(1-\alpha_{jt})R_{jt} \cdot HR_j$ is the cost associated with SO_2 emissions, which I call permit cost. The permit cost depends on the shadow value of permit λ. In the absence of transaction costs, the shadow value of emissions permits is equal to the market price of permits. In the structural model of the paper, the shadow value λ is endogenously determined due to the presence of transaction costs and permit banking.

I consider two approaches that endogenize the decision on electricity generation. First, I consider the cost minimization problem for each firm given their firm-level output. In period t, the firm decides the production allocation across generating units to minimize their total cost:

$$
\begin{aligned}
\min_{\{q_{jt}\}_{j=1}^{J_{it}}} & \sum_{j=1}^{J_{it}} mc_{jt}(\lambda) \cdot q_{jt} \\
\text{s.t.} & \sum_{j=1}^{J_{it}} q_{jt} = Q, q_{jt} \in [0, k_j]
\end{aligned}
$$

Note that the total output at the firm level Q is exogenously given.

Without loss of generality, I order the generation units $i = 1, \cdots, N$ according to increasing marginal costs, i.e., $mc_{1t}(\lambda) \leq mc_{2t}(\lambda) \leq \cdots \leq mc_{Jt}(\lambda)$ in each period t. Then, the optimal choice of production allocation is given by

$$
q_j = \begin{cases} k_j & \text{if } j = 1, \dots, J^* - 1 \\ Q - \sum_{j=1}^{J^*-1} k_j & \text{if } j = J^* \\ 0 & \text{if } j = J^* + 1, \dots \end{cases}
$$

where J^* is the minimum number of generating units whose total generation capacity exceeds the given amount of total generation Q, i.e.,

$$
J^* = \arg\min \left\{ J \,\middle|\, \sum_{j=1}^{J} k_j \geq Q \right\}.
$$

Intuitively speaking, the firm operates the generation units with cheaper costs until it satisfies the total demand of Q.

I also construct a merit-order curve at the state level. Let $\mathcal{J}_{st}$ be the set of generating units located in state s. Then, given the state-level generation Q_{st}, the cost minimization problem is given by

$$
\begin{aligned}
\min_{\{q_{jt}\}_{j \in \mathcal{J}_{st}}} & \sum_{j \in \mathcal{J}_{st}} mc_{jt}(\lambda) \cdot q_{jt} \\
\text{s.t.} & \sum_{j \in \mathcal{J}_{st}} q_{jt} = Q_{st}, q_{jt} \in [0, k_j].
\end{aligned}
$$

The solution to this problem is similarly given as that for the firm-level problem.

A.3.2 Descriptive Statistics of Fuel and Permit Cost

Before I analyze the endogenous response of electricity generation to cap-and-trade, I first report the descriptive statistics of the marginal cost in Table A3. To do this, I decompose the marginal cost into the fuel cost and the permit cost. To calculate the permit cost, I use the market price of emissions permits as shadow cost λ. I report the decomposition for three groups of generating units. First, the permit cost accounts for 7.9% and 6.2% of the marginal cost for Group I and Group II (coal). These numbers suggest that the introduction of cap-and-trade does not significantly increase the marginal cost of generation. Secondly, even if I consider the permit cost associated with SO_2 emissions, the marginal cost of coal units is substantially higher than that of gas and oil units.

Table A3: Summary Statistics of Fuel and Permit Cost

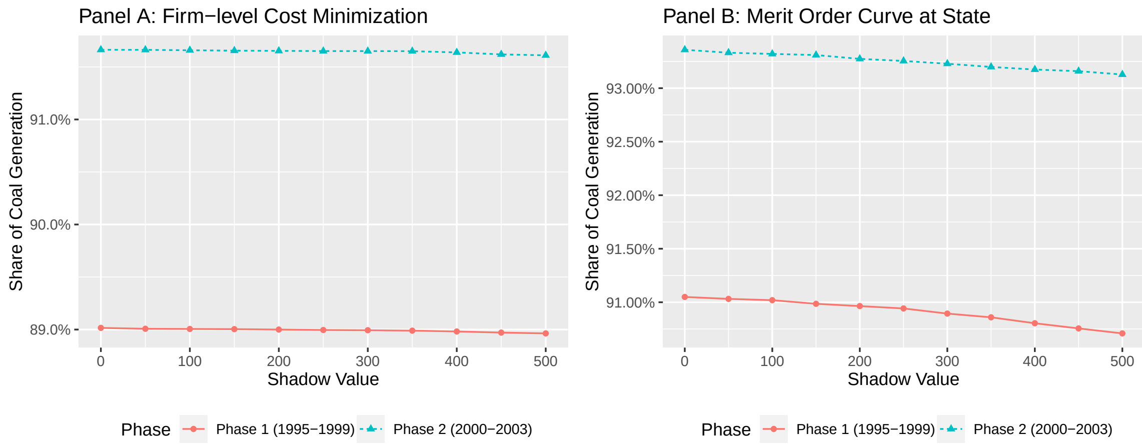
Variable	Mean	Std. Dev.	25 Percentile	Median	75 Percentile
Group: Group I					
Fuel cost	16.29	15.55	11.41	13.16	15.97
Permit cost	1.39	1.19	0.52	1.08	1.93
Group: Group II (Coal)					
Fuel cost	13.99	5.82	11.31	13.42	15.93
Permit cost	0.93	0.81	0.45	0.73	1.21
Group: Group II (Gas and Oil)					
Fuel cost	54.58	48.37	35.59	46.65	59.71
Permit cost	0.17	0.41	0	0	0.04

Note: The unit is USD per MWh. The permit cost is the shadow cost of SO_2 emissions evaluated by the observed permit price.

A.3.3 Reallocation across generating units

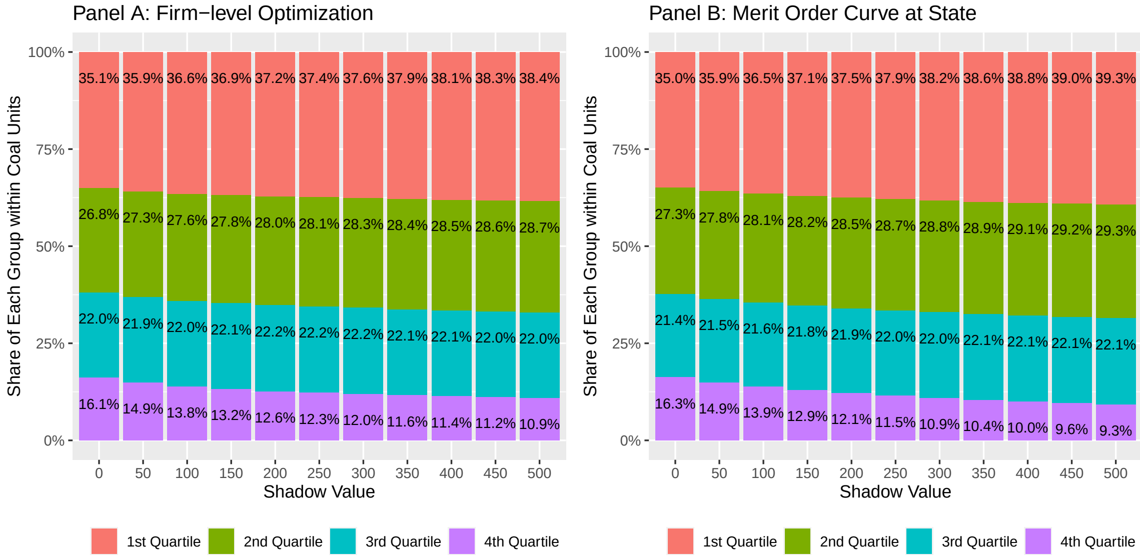
I now use the merit-order curve to investigate the endogenous change in the electricity generation. Figure A2 plots the share of coal generation as a function of the shadow value λ. In theory, the higher λ implies the higher costs of permits, leading to lower utilization of coal units. Although the share of coal generation is indeed decreasing as the shadow value λ increases, its magnitude is quite small. Even if the shadow value is set at \$500, which is quite high given that the permit price in my sample period is within the range of \$100–\$200, the share of coal generation only decreases by 0.2 percentage point in Panel B. The analysis implies that even though we consider the cost of emissions permits, reallocation of production across coal and other units (namely gas and oil units) is very limited. This finding is also consistent with the cost difference between coal and other units, as I discuss in Table A3.

Figure A2: Share of Coal Generation



I then investigate the reallocation of generations within coal units. Figure A3 illustrates how the generation share of each group of coal units changes according to the shadow value of emissions permits. Each group is defined by the quartiles of emissions rate of fuel (lbs/MMBtu). The lower quartile corresponds to cleaner units (i.e., lower emissions rate of fuel). The figure shows that reallocation from dirtier (i.e., 4th Quartile) to cleaner (i.e. 1st and 2nd Quartile) occurs when the shadow value of permits increases. Specifically, when the shadow value increases from 0 (i.e., no cap-and-trade) to 150 (which is in the ballpark of the observed permit prices in 2000–2003), the share of the dirtiest group falls from 16.1% to 13.2%, while that of the cleanest group rises from 35.1% to 36.9% in Panel A. This result indicates that introducing cap-and-trade may lead to the reallocation of generation within coal units to some extent. However, its magnitude is not sufficient to achieve the required level of emissions abatement under the Acid Rain Program.

Figure A3: Possible Reallocation within Coal Units in Phase II



Note: Quartile is defined by emissions rate of fuel (lbs/MMBtu) observed in the data. 1st quartile: lower than 0.56 lbs/MMBtu, 2nd quartile: higher than 0.56 lbs/MMBtu and lower than 1.00 lbs/MMBtu, 3rd quartile: higher than 1.00 lbs/MMbtu and lower than 1.70 lbs/MMbtu, 4th Quartile: higher than 1.70 lbs/MMBtu

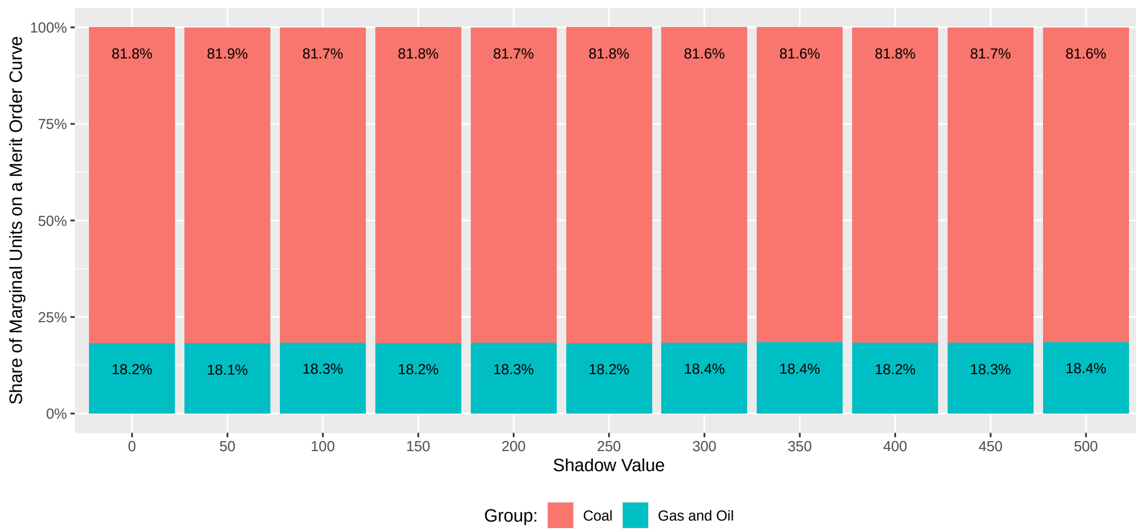
A.3.4 Marginal Units on a Merit-Order Curve at the State Level

Using a state-level merit-order curve for each month and year, I examine the share of marginal units. Figure A4 shows the share of marginal units as a function of the shadow value. Overall, the share of marginal coal units is around 82%, and this figure is almost constant across different values of the shadow values. Note that the share of coal units that are marginal is likely overestimated because I construct a merit-order curve using the monthly-level data, rather than high-frequency data (i.e., hourly-level data available in CEMS).[92]

This result raises the concern that the introduction of cap-and-trade could affect the output price determined by the marginal unit on a merit-order curve. Cap-and-trade increases the marginal cost of coal units due to the additional permit costs, which might be passed through to the output price. However, I believe this magnitude is likely to be relatively small. As I have shown in the descriptive statistics presented in Table A3, the permit cost is quite small compared to the fuel cost. Therefore, the potential impact of cap-and-trade on output price is likely to be quite small too.

[92]If I consider the demand fluctuation during the day, the model based on a merit-order curve produces a period of time when gas and oil units are marginal units.

Figure A4: Share of Marginal Units



B Online Appendix: Estimation of Fringe Demand

This subsection explains the estimation of the fringe demand function. I consider the following linear specification:

$$\bar{B}_t^{fringe} = \kappa_0 + \kappa_1 P_t + \kappa_2 Phase2_t, \tag{B.1}$$

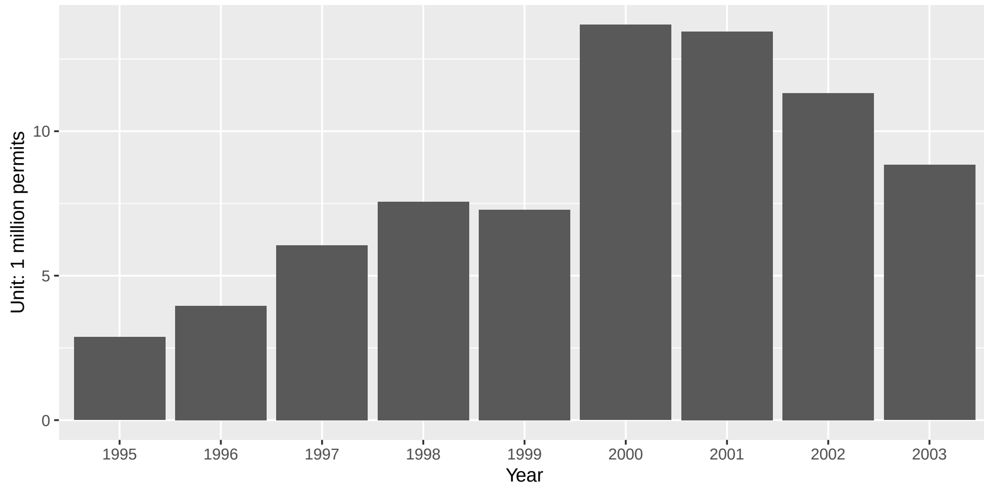
where $Phase2_t$ is the dummy for Phase II. The permit price P_t is subject to the endogeneity problem because the equilibrium permit price depends on the aggregate demand from the fringe firms. Thus, I use the sum of the initial allocation of permits owned by the firms in my sample as an instrument for P_t. The initial allocation of firms in the estimation sample is excluded from the fringe demand equation. Moreover, it is the part of the total amount of permits available in the market and thus affects permit prices. Table A4 reports IV estimates. Given the few data points available in my data (9 yearly observations from 1995 to 2003), the price coefficient is rather imprecisely estimated. For a comparison, Table A5 in the Online Appendix reports the first-stage result and the OLS result.

Table A4: Parameter Estimates of Fringe Demand

	Parameter	Description	Estimate	Standard Errors
Fringe Demand $\bar{B}_t(\cdot)$	κ_0	Constant	630110.45	577998.93
	κ_1	Permit Price	-3424.25	4133.94
	κ_2	Phase II dummy	-270591.92	191833.72

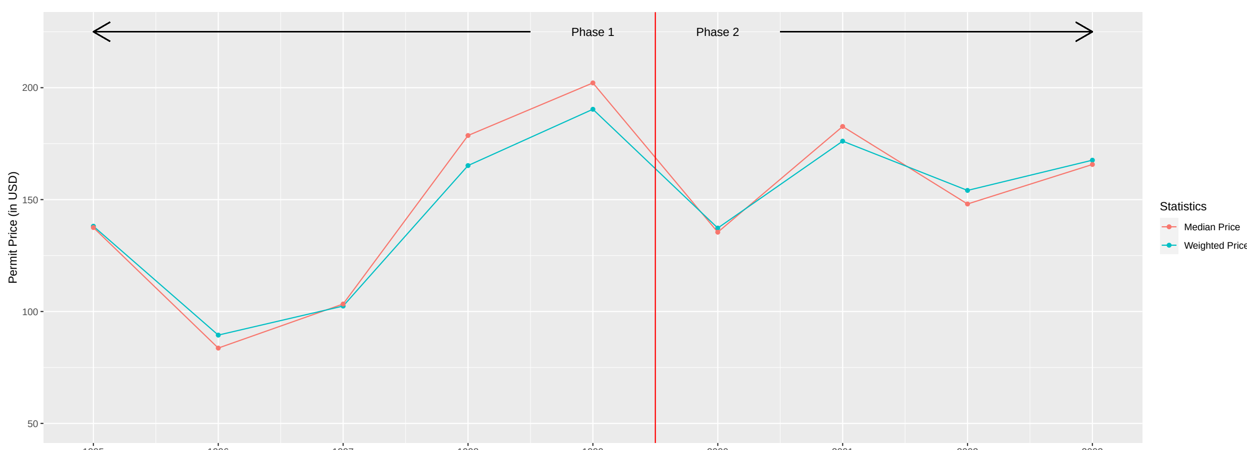
C Online Appendix: Additional Tables and Figures

Figure A5: Trading Volume over Time



Notes: The figure shows the aggregate volume of permit trading in each year. The unit is 1 million permits.

Figure A6: Comparison of volume-weighted mean and median prices of permits.



Notes: Prices are normalized to January 2000 prices using the producer price index. The prices are the weighted mean across months in each year. The weight is the aggregate trading volume of permits.

Figure A7: Frequency of Trading Normalized by Firm Size

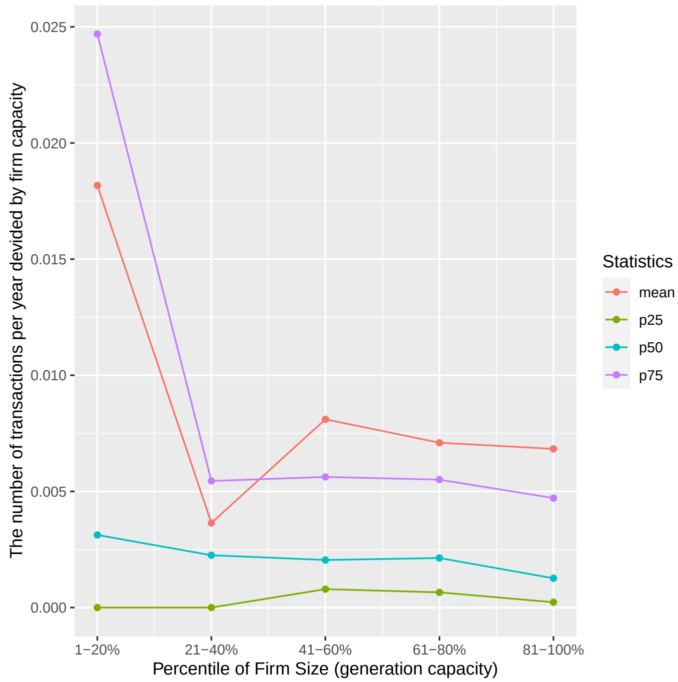
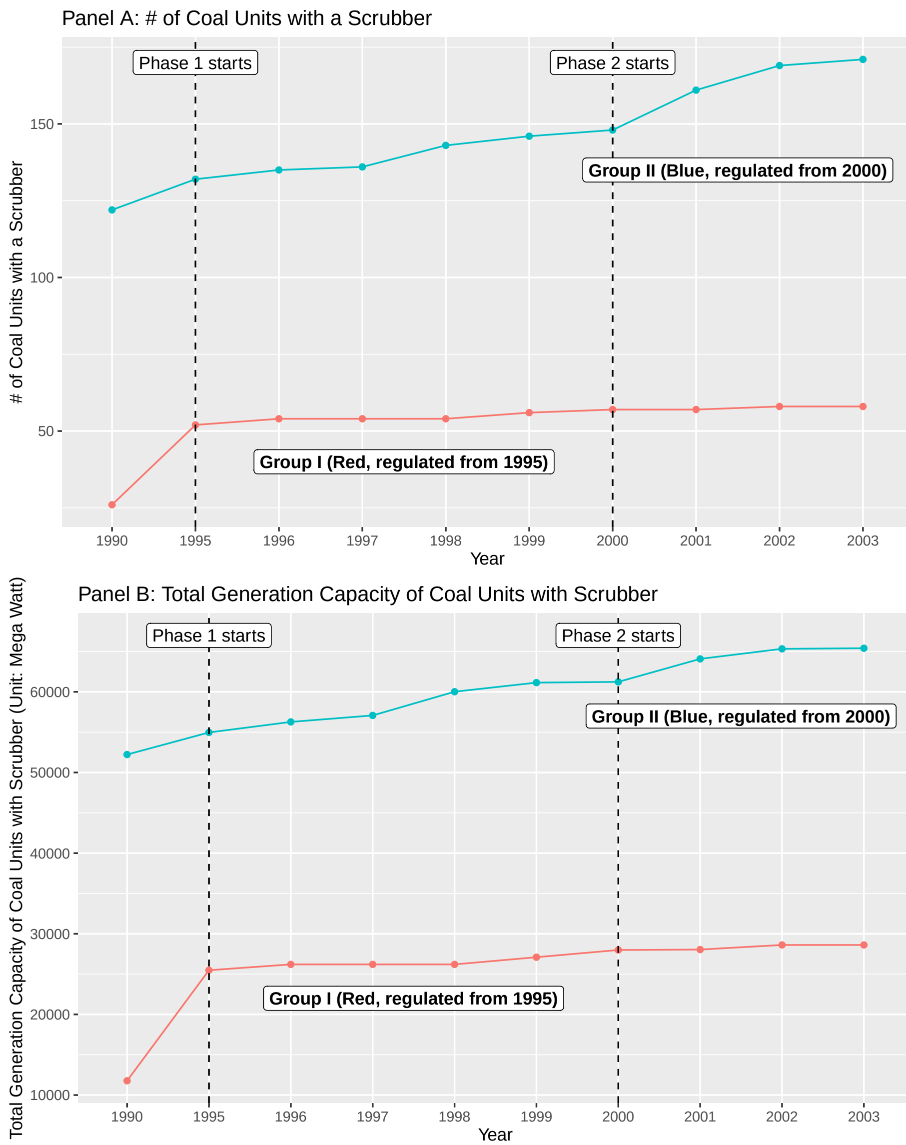


Figure A8: Scrubber Adoption



Notes: Panel A reports the number of coal units with a scrubber, while Panel B reports the total generation capacity of coal units with a scrubber. The Acid Rain Program was announced in 1990.

Table A5: Estimation Results of Fringe Demand

	OLS	1st Stage	IV
(Intercept)	974674.854	447.841	630110.447
	(377303.896)	(135.208)	(577998.931)
Permit Price	−5937.203		−3424.250
	(2642.120)		(4133.937)
Phase2 dummy	−216148.484	77.717	−270591.917
	(168181.241)	(30.112)	(191833.720)
Initial Allocation		−0.054	
		(0.024)	
R^2	0.620	0.531	0.563
Adj. R^2	0.494	0.375	0.417
Num. obs.	9	9	9

D Online Appendix: Details of Model Derivation

D.1 Derivation of the Optimality Conditions

In this appendix, I provide a detailed derivation of the optimality conditions for the constrained optimization problem introduced in Section 3.4. Recall that the constrained optimization problem is given by

$$\begin{aligned}\max_{\{R_{jt}\}_{j=1}^{J_{it}}, b_{it}, h_{i,t+1}} \quad & \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) - (P_t b_{it} + TC(|b_{it}|)) + \beta EV_{i,t+1}(h_{i,t+1}, 1, R_{i,t+1}) \\ \text{s.t.} \quad & e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha_{it}\right) + h_{i,t+1} = a_{it} + h_{it} + b_{it}, \\ & h_{i,t+1} \geq 0.\end{aligned}$$

The Lagrangian for this problem is

$$\begin{aligned}\mathcal{L} =& \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) - (P_t b_{it} + TC(|b_{it}|)) + \beta EV_{i,t+1}(h_{i,t+1}, 1, R_{i,t+1}) \\ &+ \lambda_{it}\left(a_{it} + h_{it} + b_{it} - e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha_{it}\right) - h_{i,t+1}\right) + \mu_{it} h_{i,t+1},\end{aligned}$$

where λ_{it} denotes the Lagrange multiplier on the transition of permit holding, $e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha_{it}\right) + h_{i,t+1} = a_{it} + h_{it} + b_{it}$, and μ_{it} denotes the Lagrange multiplier on the nonborrowing constraint, $h_{i,t+1} \geq 0$. Taking the first-order conditions, I have

$$\frac{\partial \mathcal{L}}{\partial R_{jt}} = \frac{\partial \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right)}{\partial R_{jt}} - \lambda_{it} \frac{\partial e_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right)}{\partial R_{jt}} = 0 \tag{D.1}$$

$$\frac{\partial \mathcal{L}}{\partial b_{it}} = -P_t - \frac{dTC(|b_{it}|)}{db_{it}} + \lambda_{it} = 0 \tag{D.2}$$

$$\frac{\partial \mathcal{L}}{\partial b_{it}} = \beta \frac{dEV_{i,t+1}(h_{i,t+1}, I_{i,t+1}, R_{i,t+1})}{dh_{i,t+1}} + \mu_{it} - \lambda_{it} = 0 \tag{D.3}$$

Equation (3.1) implies that $\frac{\partial \pi_{it}\left(\{q_{jt},R_{jt}\}_{j=1}^{J_{it}}\right)}{\partial R_{jt}} = -\frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}} HR_j q_{jt}$ and Equation (3.2) implies that $\frac{\partial e_{it}\left(\{q_{jt},R_{jt}\}_{j=1}^{J_{it}}\right)}{\partial R_{jt}} = (1-\alpha_{jt})HR_j q_{jt}$. Thus, Equation (D.1) can be written as $\lambda_{it}\left((1-\alpha_{jt})HR_j q_{jt}\right) = -\frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}} HR_j q_{jt}$, as shown in Equation (3.6). It is clear to see that Equations (D.2) and (D.3) can be written as Equations (3.7) and (3.8), respectively. Lastly, the complementary slackness condition with respect to the banking constraint $h_{it+1} \geq 0$ is given by $\mu_{it}h_{it+1} = 0, \mu_{it} \geq 0, h_{i,t+1} \geq 0$, as is shown in Equation (3.9).

D.2 Derivation of the Participation Probability $\mathbb{P}_{it}(h_{it}, \alpha_{it})$ and Ex-ante Value Function $EV_{it}(h_{it}, I_{it}, \alpha_{it})$

Participation probability of $\mathbb{P}_{it}(h_{it}, \alpha_{it})$ Recall that the participation probability is given by

$$\mathbb{P}_{it}(h_{it}, \alpha_{it}) = \int \mathbf{1}\left\{V_{it}^1(h_{it}, \alpha_{it}) - (F + \sigma_F \epsilon_{it}) > V_{it}^0(h_{it}, \alpha_{it})\right\} dG(\epsilon_{it}).$$

Since I assume the type-I extreme value distribution of ϵ_{it}, the participation probability is given by the well-known logit formula:

$$\mathbb{P}_{it}(h_{it}, \alpha_{it}) = \frac{\exp\left(\frac{V_{it}^1(h_{it},\alpha_{it})-F}{\sigma_F}\right)}{\exp\left(\frac{V_{it}^0(h_{it},\alpha_{it})}{\sigma_F}\right) + \exp\left(\frac{V_{it}^1(h_{it},\alpha_{it})-F}{\sigma_F}\right)}$$

Ex-ante value function of $EV_{it}(h_{it}, I_{it}, \alpha_{it})$ Recall that the ex-ante value functions are given by

$$EV_{it}(h_{it}, I_{it}, \alpha_{it}) = \begin{cases} \int \max\left\{V_{it}^0(h_{it}, \alpha_{it}), V_{it}^1(h_{it}, \alpha_{it}) - (F + \sigma_F \epsilon)\right\} dG(\epsilon) & if\ I_t = 0 \\ V_{it}^1(h_{it}, \alpha_{it}) & if\ I_t = 1. \end{cases}$$

Under the assumption that ϵ follows an i.i.d. type-I extreme value distribution, the expected value function when $I_{it} = 0$ can be written as

$$EV_{it}(h_{it}, I_{it} = 0, \alpha_{it}) = \sigma_F \log\left[\exp\left(\frac{V_{it}^0(h_{it}, \alpha_{it})}{\sigma_F}\right) + \exp\left(\frac{V_{it}^1(h_{it}, \alpha_{it}) - F}{\sigma_F}\right)\right].$$

By applying the Williams–Daly–Zachary theorem and the envelope theorem, the derivative of the expected value function with respect to the state variable h_{it} can be expressed as follows:

$$\frac{dEV_t(h_{it}, 0, \alpha_{it})}{dh_{it}} = \mathbb{P}_{it}(h_{it}, \alpha_{it})\lambda_{it}^1 + (1 - \mathbb{P}_t(h_{it}, \alpha_{it}))\,\lambda_{it}^0. \tag{D.4}$$

$$\frac{dEV_t(h_{it}, 1, \alpha_{it})}{dh_{it}} = \lambda_{it}^1, \tag{D.5}$$

where λ_{it}^1 and λ_{it}^0 are the Lagrange multipliers on the constraints for permit transitions in the optimization problems for traders and nontraders, respectively. I now provide a detailed derivation of the above equations.

Derivation of $\partial EV_t(h_t, I_t, \alpha_t)/\partial h_t$ I omit the index i for a firm for ease of exposition. I focus on the derivation of $\frac{\partial EV_t(h_t,0,\alpha_t)}{\partial h_t}$. Recall that

$$EV_t(h_t, 0, R_t) \quad = \quad \int \max\left\{V_t^0(h_t, R_t), V_t^1(h_t, R_t) - F - \sigma_F \epsilon\right\} dG(\epsilon).$$

By the chain rule, I have

$$\frac{dEV_t(h_t, 0, \alpha_t)}{dh_t} = \frac{\partial EV_t}{\partial V_t^0}\frac{dV_t^0}{dh_t} + \frac{\partial EV_t}{\partial V_t^1}\frac{dV_t^1}{dh_t}.$$

First, I derive $\frac{\partial EV_t}{\partial V_t^k}$ for $k = 0, 1$. This is an application of the Williams–Daly–Zachary theorem (see Theorem 3.1 in Rust, 1994). Using the interchange of integration and differentiation, I arrive at the following (I omit h_t for ease of exposition in the following derivation):

$$\begin{aligned}
\frac{\partial EV_t}{\partial V_t^1} &= \frac{\partial}{\partial V_t^1}\int \max\left\{V_t^1 - F - \sigma_F\epsilon, V_t^0\right\} dG(\epsilon) \\
&= \frac{\partial}{\partial V_t^1}\int_{\Upsilon^1}(V_t^1 - F - \sigma_F\epsilon)dG(\epsilon) + \frac{\partial}{\partial V_t^1}\int_{\Upsilon^0} V_t^0 dG(\epsilon) \\
&= \int_{\Upsilon^1}\frac{\partial}{\partial V_t^1}(V_t^{trade} - F - \sigma_F\epsilon)dG(\epsilon) + \int_{\Upsilon^0}\frac{\partial}{\partial V_t^1}V_t^0 dG(\epsilon) \\
&= \int_{\Upsilon^1} dG(\epsilon) \\
&= \mathbb{P}_t(\cdot),
\end{aligned}$$

where Υ^1 is the set of ϵ such that a firm chooses to participate (i.e., $\Upsilon^1 \equiv \{\epsilon : V_t^1 - F - \sigma_F\epsilon > V_t^0\}$), and Υ^0 is defined similarly. Note that I can apply a similar derivation to obtain $\frac{\partial EV_t}{\partial V_t^0} = 1 - \mathbb{P}(h_t)$.

Next, I calculate $\frac{\partial V_t^k}{\partial h_t}$, for $k = 0, 1$. The derivation is a direct application of the envelope theorem (or the Benveniste–Scheinkman formula):

$$\frac{\partial V_t^k}{\partial h_t} = \lambda_t^k,$$

where λ_{it}^k denotes the Lagrange multipliers in the corresponding optimization problems. Thus, I obtain

$$\frac{dEV_t(h_t, 0)}{dh_t} = \mathbb{P}_t(h_t)\lambda_t^1 + (1 - \mathbb{P}_t(h_t))\,\lambda_t^0.$$

D.3 Incentives in Abatement Investment

Here, I discuss how the incentive to invest in abatement is determined in the model. Using the envelope theorem, the marginal return from increasing the removal rate of a scrubber is given as follows:

$$\frac{\partial EV_{1995}}{\partial \alpha^1} = \sum_{t=1995}^{1999} \beta^{t-1995} \left(\lambda_{it} \cdot \sum_{j=1}^{J_{it}} R_{jt} \cdot HR_{jt} \cdot q_{jt} \right) - \beta^{2000-1995} \frac{\partial}{\partial \alpha^1} \Gamma(\alpha^2 - \alpha^1).$$

The first term is the returns from emissions abatement evaluated at the shadow value λ_{it}. The second term is the saving of investment costs in Phase II due to the earlier investment in Phase I.

The primary component in the return on investment is the first term. By increasing the removal rate of a scrubber, a firm can marginally reduce its emissions by $\sum_{j=1}^{J_{it}} R_{jt} \cdot HR_{jt} \cdot q_{jt}$. This marginal abatement is evaluated at the shadow value of λ_{it}. Thus, the return on investment is given by the discounted sum of the returns on the marginal abatement. The path of shadow values λ_{it} is key for the investment incentives. As discussed in Section 3.7, the shadow value λ_{it} and equilibrium permit price P_t are affected by both permit banking and transaction costs.

D.4 Details of Decomposition of Change in Health and Environmental Damages

The aggregate health and environmental damage is given by

$$D = \sum_{t=1995}^{2003} \beta^{t-1995} \left(\sum_{i=1}^{N} \sum_{j=1}^{J_{it}} d_j e_{jt} \right),$$

where e_{jt} is emissions from unit j in year t and d_j is the health and environmental damage from emissions produced by unit j. Note that d_j is the county-level estimates of SO_2 damage constructed by Muller and Mendelsohn (2009).

I rewrite the aggregate damage as

$$\begin{aligned} D &= \sum_{t=1995}^{2003} \beta^{t-1995} E_t \left(\frac{\sum_{i=1}^{N} \sum_{j=1}^{J_{it}} d_j e_{jt}}{E_t} \right) \\ &= \sum_{t=1995}^{2003} \tilde{d}_t E_t, \end{aligned}$$

where $E_t = \sum_{i=1}^{N} \sum_{j=1}^{J_{it}} e_{jt}$ is the aggregate emissions in year t and $\tilde{d}_t = \beta^{t-1995} \left(\frac{\sum_{i=1}^{N} \sum_{j=1}^{J_{it}} d_j e_{jt}}{E_t} \right)$ is the (discounted) average SO_2 damages in year t. Note that $\tilde{d}_t$ interpreted as the weighted average of health and environmental damages, where the weight is given by the amount of SO_2 emissions.

E Online Appendix: Computational Details of Solving the Model

Appendix E explains the computational procedure used to solve the structural model.

E.1 Decomposition of the Per-Period Problem

One of the choice variables in the individual dynamic decision problem is the unit-level coal quality R_{jt}, which appears in the profit function π_{it}, given by Equation (3.1), and the firm-level emissions, $e_{it} = \sum_{j=1}^{J_{it}}(1-\alpha_{jt})R_{jt} \cdot HR_j \cdot q_{jt}$. Because each firm has multiple generation units, solving unit-level production in a dynamic framework seems computationally demanding. Therefore, to reduce the computational burden, I decompose the per-period problem into the following two problems. First, I consider the following optimization problem with respect to the unit-level coal quality $\{R_{jt}\}_{j \in J_{it}}$, *holding firm-level emissions e_{it} fixed:*

$$\begin{aligned} \Pi_{it}(e_{it}, \alpha_{it}) \equiv \quad & \max_{\{R_{jt}\}_{j \in J_{it}}} \quad \pi_{it}\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) \\ & \text{s.t.} \qquad \sum_{j=1}^{J_{it}}(1-\alpha_{it})R_{jt} \cdot HR_j \cdot q_{jt} = e_{it}. \end{aligned}$$

$\Pi_{it}(e_{it}, \alpha_{it})$ is the optimal profit *as a function of the firm-level emissions e_{it}.* Note that the FOCs for this subproblem are

$$\begin{aligned} \lambda_{it}^{sub}\left((1-\alpha_{jt})HR_j q_{jt}\right) &= -\frac{\partial p_{jt}^{fuel}(R_{jt})}{\partial R_{jt}} HR_j q_{jt} \\ \sum_{j=1}^{J_{it}}(1-\alpha_{it})R_{jt} \cdot HR_j \cdot q_{jt} &= e_{it}, \end{aligned}$$

where λ_{it}^{sub} is the Lagrange multiplier of the constraint on firm-level emissions in the above problem.

I now use $\Pi_{it}(e_{it}, \alpha_{it})$ to consider the dynamic decision problem:

$$\begin{aligned} \max_{e_{it}, b_{it}, h_{i,t+1}} \quad & \Pi_{it}(e_{it}, \alpha_{it}) - (P_t b_{it} + TC(b_{it})) + \beta EV_{i,t+1}(h_{i,t+1}, 1, R_{i,t+1}) \\ \text{s.t.} \quad & e_{it} + h_{i,t+1} = a_{it} + h_{it} + b_{it}, \\ & h_{i,t+1} \geq 0. \end{aligned}$$

Note that the choice variables are now reduced to e_{it}, b_{it}, and $h_{i,t+1}$.

When I numerically solve the individual dynamic decision problem, I follow two steps. First, I construct $\Pi_{it}(e_{it}, \alpha_{it})$ using the unit-level FOC for production. I then use the precomputed $\Pi_{it}(e_{it}, \alpha_{it})$ to solve the individual dynamic decision problems.

E.2 Individual Optimization

I explain the computational procedure for solving an individual problem. For notational simplicity, I omit the script i for a particular firm. Because the model has a finite period, it can be solved using backward induction.

1. Phase II (2003 to 2000): I solve the optimization problem from 2003 to 2000. Note that I use $CV_{T+1}(h_{T+1}, \alpha^2)$ as a continuation value in the terminal period 2003. By solving with backward induction, I obtain the policy function $\hat{x}_t(h_t, I_t, \alpha^2)$ for emissions e_t, net purchase b_t, and banking h_{t+1}, and the expected value function in 2000 $EV_{2000}(h_{2000}, I_{2000}, \alpha^2)$.

2. Investment decision for Phase II: I define the continuation value at the timing of making the investment decision for Phase II by $W_{2000}(h_{2000}, I_{2000}, \alpha^1)$. The decision problem is given by

$$\begin{aligned} W_{2000}(h_{2000}, I_{2000}, \alpha^1) \equiv \quad & \max_{\alpha^2} \quad EV_{2000}(h_{2000}, I_{2000}, \alpha^2) - \Gamma(\alpha^2, \alpha^1). \\ & \text{s.t.} \quad \alpha^2 \leq \alpha^1 \end{aligned}$$

   By solving this problem, I obtain the investment policy function $\alpha^{2*}(h_{2000}, I_{2000}, \alpha^1)$.

3. Phase I (1999 to 1995): I repeat the same procedure as that in step 1. Note that the continuation value in the problem at $t = 1999$ is given by $W_{2000}(h_{2000}, I_{2000}, \alpha^1)$.

4. Investment for Phase I: The problem is given by

$$\begin{aligned} & \max_{\alpha^1} EV_{1995}(0, 0, \alpha^1) - \Gamma(\alpha^1, \alpha^0). \\ & \text{s.t.} \alpha^1 \leq \alpha^0 \end{aligned}$$

   Note that $h_{1995} = 0$ and $I_{1995} = 0$ in 1995.

E.3 Computation of a Dynamic Competitive Equilibrium

The computational procedure for finding an equilibrium is parallel to the estimation procedure introduced in Section 4.

1. Fix a candidate of permit prices: $\mathbf{P} = \{P_t\}_{t=1995}^{2003}$.

2. Solve the individual problem using backward induction and obtain the policy function $\hat{x}_{it}(h_{it}, I_{it}, \alpha_{it})$ for emissions e_t, net purchase b_t, and banking h_{t+1}, participation probability $P_{it}(h_{it}, \alpha_{it})$, and the investment decisions $\alpha_i^1(h_{i1995}, I_{i1995})$ and , $\alpha_i^2(h_{i,2000}, I_{i,2000}, \alpha_i^1)$.

3. Consider the timing of market participation. Denote the year of participation by $s \in \{\emptyset, 1995, \cdots, 2003\}$. Here, $s = \emptyset$ means that a firm does not trade in a period.

4. For each path of participation timing, I simulate the optimal decisions using the policy functions.

5. Calculate the probability that each path of participation timing is realized.

6. The simulated optimal decisions are given as

$$\hat{x}_{it} = \sum_{s \in \{\emptyset, 1995, \cdots, 2003\}} Prob_i^{enter}(s) \hat{x}_{it}(s),$$

where x denotes the choice variables.

7. Check the market-clearing condition as

$$\sum_{i=1}^{N} \hat{b}_{it}(\mathbf{P}) + \bar{B}_t^{fringe}(P_t) = 0 \ \forall t = 1995, \cdots, 2003.$$

8. Stop the iteration when the following condition is satisfied:

$$\max_{t=1995, \cdots, 2003} \left| \sum_i \hat{b}_{it}(\mathbf{P}) + \bar{B}_t^{fringe}(P_t) \right| < 1000.$$

Note that this criterion is sufficiently tight to ensure that the absolute value of the price change is in the order of magnitude of 1e-1.

9. If the above is not satisfied, repeat steps 1–7 with the updated price vector (explained below), until the market-clearing conditions are satisfied.

Price Update Rule To update the price in each iteration, I construct the following heuristic rule that exploits the market-clearing conditions and the optimality conditions. Denote the current candidate of an equilibrium price vector by $\mathbf{P}^l = \{P_t^l\}_{t=1995}^{2003}$. The next candidate of price in year t, P_t^{l+1}, is given by solving the following equation:

$$\sum_{i=1}^{N} \sum_{s \in \{\emptyset, 1995, \cdots, 2003\}} P_{i,enter}(s) \cdot TC'^{(-1)} \left(\hat{\lambda}_{it}(\mathbf{P}^l, s) - P_t^{l+1} \right) + \bar{B}_t^{fringe}(P_t^{l+1}) = 0,$$

where $\hat{\lambda}_{it}(\mathbf{P}^l, s)$ is the prediction of the shadow value when the current price candidate is $\mathbf{P}^l$ and the year of participation is s. Note that at the fixed point of this equation, where $\mathbf{P}^l = \mathbf{P}^{l+1}$,

$$TC'^{(-1)} \left(\hat{\lambda}_{it}(\mathbf{P}^l, s) - P_t^l \right) = b_{it}(\mathbf{P}^l, s),$$

such that the market-clearing conditions are satisfied in all periods.

The computation procedure with this price update rule works relatively well in numerical simulations. The algorithm finds an equilibrium price vector in fewer than 10 iterations in most cases, though I do not have a formal proof of this property of the algorithm.

F Online Appendix: Details of Counterfactual Simulations

F.1 Eliminating Transaction Costs

I now consider the case with permit banking. In the absence of transaction costs, Rubin (1996) has shown that the equilibrium path of permit prices grows at the rate of β^{-1}, as long as the aggregate banking is positive, which implies that

$$\begin{aligned} P_{t+1} &= \beta^{-1} P_t \\ \Longleftrightarrow P_t =& \beta^{-(t-1)} P_{1995} \text{ for } t \in \{1995, \cdots, 2003\}. \end{aligned}$$

The optimal decision on emissions, given the emissions rate of fuel, is determined by $\partial \pi_{it} / \partial R_{jt} = P_t \ \forall j$. As discussed in Section 3.7.2, individual decisions on net purchases and banking are not determined from the model because the current shadow value $\lambda_t = P_t$ is equal to the discounted marginal value of banking $\beta \lambda_{t+1} = \beta P_{t+1} = P_t$. In other words, banking and trading decisions are arbitrary as long as a firm can produce the level of emissions determined by the optimality condition.

Now, I consider the investment decisions. The continuation value at the beginning of Phase II is given by

$$\begin{aligned} V_{i,2000}(h_{i,2000}, \alpha_i^2) &= \sum_{t=2000}^{2003} \beta^{t-2000} \left[\pi_{it} \left(\{q_{jt}\}_{j=1}^{J_{it}}, \alpha_i^2 \right) - P_t b_{it} \right] + \beta^{2003-2000} CV(h_{i,T+1}) \\ &= \sum_{t=2000}^{2003} \beta^{t-2000} \left[\pi_{it} \left(\{q_{jt}\}_{j=1}^{J_{it}}, \alpha_i^2 \right) - P_t \cdot (e_{it} - a_{it}) \right] \\ &\quad + \beta^{2003-2000} \left\{ CV(h_{i,T+1}) - P_T h_{i,T+1} \right\} \\ &\quad + \sum_{t=2000}^{2003} \beta^{t-2000} P_t h_{it} + \sum_{t=2000}^{2002} \beta^{t-2000} P_t h_{it+1} \\ &= \sum_{t=2000}^{2003} \beta^{t-2000} \left[\pi_{it} \left(\{q_{jt}\}_{j=1}^{J_{it}}, \alpha_i^2 \right) - P_t \cdot (e_{it} - a_{it}) \right] \\ &\quad + \beta^{2003-2000} \left\{ CV(h_{i,T+1}) - P_T h_{i,T+1} \right\} + P_{2000} h_{i,2000}, \end{aligned}$$

where the last equality uses the equilibrium relationship $\beta P_{t+1} = P_t$. The investment problem is

$$\begin{aligned} W_{i,2000}(h_{i,2000}, \alpha_i^1) = \quad & \max_{\alpha_i^2} \quad V_{2000}(h_{i,2000}, \alpha_i^2) - \Gamma(\alpha_i^2, \alpha_i^1). \\ & s.t. \quad \alpha_i^2 \leq \alpha_i^1. \end{aligned}$$

Note that $h_{i,2000}$ does not affect the optimal investment level of α_i^2.

The continuation value at the beginning of Phase I is given as

$$\begin{aligned} V_{1995}(h_{i,1995}, \alpha_i^1) \quad = \quad & \sum_{t=1995}^{1999} \beta^{t-1995} \left[\pi_{it} \left(\{q_{jt}\}_{j=1}^{J_{it}}, \alpha_i^1 \right) - P_t(e_{it} - a_{it}) \right] \\ & + \beta^{1999-1995} \left(\beta W_{2000}(h_{i,2000}, \alpha_i^1) - P_{1999} h_{i,2000} \right). \end{aligned}$$

The investment problem is similar to that in Phase II.

Finally, I consider the market-clearing condition. By aggregating the transition equation of permit holding (3.3) over individual firms and time, I have

$$\sum_{t=1995}^{2003} E_t(P_t) + H_{T+1} = \sum_{t=1995}^{2003} A_t + \sum_{t=1995}^{2003} B_t, \tag{F.1}$$

where $E_t = \sum_{i=1}^{N} e_{it}(P_t)$, and other uppercase variables are defined similarly. The market-clearing condition in each period is

$$B_t + \bar{B}_t^{fringe}(P_t) = 0.$$

By substituting this condition into Equation (F.1), I have

$$\sum_{t=1995}^{2003} E_t \left(\beta^{-(t-1)} P_{1995} \right) + H_{T+1}(\beta^{-(T-1)} P_{1995}) = \sum_{t=1995}^{2003} A_t + \sum_{t=1995}^{2003} -\bar{B}_t^{fringe} \left(\beta^{-(t-1)} P_{1995} \right).$$

The equilibrium price P_{1995} is determined by this equation and, thus, so is the whole path of the equilibrium price.

F.2 Model without Permit Banking between Phase I and II

I explain the case in which firms are not allowed to bank emissions permits between Phases I and II. The decision problem is the same as that introduced in Section 3, except for 1999, the last year of Phase I.

I first consider the problem for a trader in 1999 (i.e.,$t = 1999$). I omit the subscript i for simplicity. The problem is given by

$$
\begin{aligned}
V^1_{1999}(h_{1999}, I_{1999} = 1, \alpha^1) = \quad & \max_{\{R_{jt}\}_{j=1}^{J_{it}}, b_t} \quad \pi_t\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) - (P_t b_t + TC(|b_t|)) + \beta W_{2000}(0, I_{2000}, \alpha^1) \\
& \text{s.t.} \qquad e_t\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha^1\right) = a_t + h_t + b_t.
\end{aligned}
$$

Note that permit banking h_{2000} is not among the choice variables, while the continuation value $W_{2000}(0, I_{2000}, \alpha^1)$ is evaluated at $h_{2000} = 0$. The optimality conditions of the problem are given by Equations (3.6) and (3.7).

Next, consider the case in which a firm is a non-trader:

$$
\begin{aligned}
V^0_{1999}(h_{1999}, I_t = 0, \alpha^1) = \quad & \max_{\{R_{jt}\}_{j=1}^{J_{it}}, b_t} \quad \pi_t\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}\right) + \beta W_{2000}(0, I_{2000}, \alpha^1) \\
& \text{s.t.} \qquad e_t\left(\{q_{jt}, R_{jt}\}_{j=1}^{J_{it}}, \alpha^1\right) \leq a_t.
\end{aligned}
$$

In this case, a firm may not consume all its permits owing to the capacity constraints of production. The emissions level is given by

$$
e_t^* = \min\left\{a_t, e_t^{max}\right\},
$$

where e_t^{max} is the emissions level when a firm faces zero shadow costs of permits $\lambda_t = 0$.

Other components, including the participation and the investment decisions, are the same as in the baseline case (i.e., the case that includes both permit banking and transaction costs).

Reference for Online Appendix

Asker, John, Allan Collard-Wexler, and Jan De Loecker (2019). "(Mis) allocation, market power, and global oil extraction". In: *American Economic Review* 109.4, pp. 1568–1615.

Borenstein, Severin, James B Bushnell, and Frank A Wolak (2002). "Measuring market inefficiencies in California's restructured wholesale electricity market". In: *The American Economic Review* 92.5, pp. 1376–1405.

Bushnell, James B, Erin T Mansur, and Celeste Saravia (2008). "Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets". In: *The American Economic Review* 98.1, pp. 237–266.

Callaway, Brantly and Pedro HC SantAnna (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Energy Information Administration (1990–2003). *Form EIA-826 detailed data*. (last accessed: September 17, 2014). URL: `http://www.eia.gov/electricity/data/eia826/`.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Muller, Nicholas Z and Robert Mendelsohn (2009). "Efficient pollution regulation: getting the prices right". In: *The American Economic Review* 99.5, pp. 1714–1739.

Rubin, Jonathan D (1996). "A model of intertemporal emission trading, banking, and borrowing". In: *Journal of Environmental Economics and Management* 31.3, pp. 269–286.

Rust, John (1994). "Structural estimation of Markov decision processes". In: *Handbook of econometrics* 4, pp. 3081–3143.

Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.