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

Yuta Toyama†
Waseda University

April 22, 2019

Abstract

While the cap-and-trade program was originally proposed as a static regulation, its implementation introduces dynamic incentives such as saving (banking) of emissions permits. I examine the performance of the program by accounting for dynamic regulatory design and firms' incentives in the context of the US Acid Rain Program. I develop and estimate a dynamic equilibrium model of abatement investment and permit trading and banking, subject to transaction costs. Simulations reveal that although permit banking improves the cost-efficiency, the aggregate level of banking is excess due to transaction costs. Distribution of emissions would be more dispersed in the first best.

Key words: *cap-and-trade regulation, dynamic equilibrium model, gains from trade, permit banking, transaction costs, electricity industry.*
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 Vivek Bhattacharya, Igal Hendel, Gaston Illanes, Koichiro Ito, Kei Kawai, Mark Roberts, Lola Segura, Max Tabord-Meehan, Yuta Takahashi, Nicholas Vreugdenhil, Yasutora Watanabe, Anthony Lee Zhang, and seminar participants at many institutions and conferences for their helpful comments and discussions. I acknowledge financial support from a Waseda University Grant for Special Research Projects (2018S-223). 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.

†School of Political Science and Economics, Research Institute for Environmental Economics and Management, Waseda University, 1-6-1 Nishi-Waseda, Tokyo, Japan. E-mail: yuta.toyama@gmail.com

1 Introduction

Achieving environmental sustainability without compromising economic efficiency is a central question in policy debates and the economics literature. The traditional approach adopted by policymakers is command-and-control regulation. However, economists advocate market-based approaches that provide incentives to reduce emissions through a market mechanism. The cap-and-trade program, or emissions trading program, is one such example. Regulated firms are allowed to trade emissions permits in order to achieve a target level of *aggregate* emissions in a flexible way. The idea of cap-and-trade was originally proposed in the seminal work of Coase (1960), then formalized by Montgomery (1972). Such schemes are now widely adopted, including in the air pollution regulations of the United States and in greenhouse gas regulations in the European Union.

While these seminal works consider cap-and-trade regulation as a static regulatory framework, its implementation has dynamic aspects. A cap-and-trade program spans multiple periods, and the regulatory standard is typically rising over time (i.e., emissions cap is decreasing). Moreover, the regulator often allows for the intertemporal reallocation of emissions permits, namely saving of emissions permits. Although these dynamic features prevail in many cap-and-trade programs, empirical studies in the literature focus mainly on static decisions in the steady state (see, e.g., Carlson et al. 2000, Fowlie 2010b, and Chan 2015).

The dynamic nature of the regulation affects firms' behavior and thus the consequence of the regulation. In particular, permit banking exhibits a tradeoff in efficiency. On the one hand, regulated firms can smooth abatement costs through permit banking, improving intertemporal efficiency. Theoretical studies of the cap-and-trade (e.g., Rubin, 1996) emphasize this effect in a frictionless environment. However, permit banking can lower firms' incentive to trade (especially sell) permits, hindering efficient reallocation of permits through the market mechanism. This channel could be important in a setting where the Coase theorem is unlikely to hold.

The goal of this study is to propose an empirical framework that evaluates the performance of the cap-and-trade program by accounting for the dynamic nature of the regulatory design and the forward-looking incentives of regulated firms. To do so, I develop and estimate a new model of abatement investment 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 power plants. Using counterfactual simulations based on the estimated model, I examine the consequences of the Acid Rain Program, emphasizing the dynamic aspects of firms' behavior and regulatory design.

In a cap-and-trade program, regulated firms must surrender emissions permits to offset their actual emissions. To meet this regulatory requirement, firms face a sort of "make-or-buy" decision problem: reduce emissions or trade (buy) emissions permits. Dynamic incentives matter 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 on abatement investment 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 affect the performance of a cap-and-trade program.

First, the model allows for rich observed firm heterogeneity in terms of abatement costs and the initial allocation of emissions permits. Firm heterogeneity is why the trading of emissions permits occurs. Trading leads to a reallocation of emissions permits and achieves a more efficient distribution of emissions. However, incorporating heterogeneity in a dynamic equilibrium framework can be computationally intensive. To circumvent this issue, I introduce a simple econometric strategy that avoids computation of dynamic competitive equilibrium in estimation. This strategy is based on the literature on the estimation of dynamic structural models.

In addition, 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 those firms' investment 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, every firm should have the same shadow value of permits, given by the market price. Furthermore, the equilibrium permit price should increase at the inverse of the interest rate (i.e., Hotelling rule). This places a strong restriction on an empirical analysis. By including transaction costs, a firm's shadow value of emissions permits is determined endogenously.

Stavins (1995) was the first theoretical study to investigate how transaction costs discourage permit trading and lead to inefficient outcomes in a static setting. Concerns related to transaction costs have been pointed out in practice. Previous studies document that many firms tend not to trade emissions permits, and instead comply with the regulation using their allocated permits (see, e.g., Jaraitė-Kažukauskė and Kažukauskas, 2015 for the EU Emissions Trading Scheme). 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. I structurally estimate the costs using the proposed model 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 sulfur dioxide (SO_2) emissions that targets the US electricity industry.[1] The aim of the Acid Rain Program is to reduce the aggregate SO_2

[1] I terminate my analysis in 2003 because of the proposal of the Clean Air Interstate Rule in December

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.[2] Regulated sources could choose how to comply with the regulation. For example, they could switch to a 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. Before the regulation was implemented in 1995, in 1990, the US EPA (regulator) announced the permit allocation schedule. The schedule is generous in the first five years of the regulation (1995–1999, called Phase I) and then decreases by almost half in the period after 2000 (Phase II). Casual observation suggests that firms took this schedule into account, banking a significant number of permits in the first five years, and then using them once the cap became tighter after 2000. While this observation suggests the importance of banking in firms' compliance strategies, some are concerned about the excessive banking of permits (e.g., Smith et al., 1998). My framework analyzes how such dynamic incentives affect firms' compliance decisions and the performance of the program.

An econometric analysis poses a challenge in terms of computation. My model belongs to a class of models that feature a dynamic competitive equilibrium with multiple heterogeneous firms. A full solution approach (i.e., solving a dynamic competitive equilibrium for each evaluation of the model parameters) is computationally prohibitive. To circumvent the computation costs, I use the observed permit prices as a sequence of equilibrium prices, and avoid solving equilibrium permit prices in the estimation. This trick is similar in spirit to the two-step estimators in dynamic Markov games, such as that in Aguirregabiria and Mira (2007). Because each firm faces a different optimization problem, owing to the firm heterogeneity, this approach reduces the computation costs in the estimation significantly.[3] I match the model predictions with the data to construct the least squares criterion function.

The estimates of my model parameters imply that the variable transaction costs from permit trading are substantial. The median of the marginal transaction cost is estimated to be USD 36, whereas the permit prices range between USD 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.

2003, which had a major impact on the regulatory environment for SO_2 emissions. See Section 2.2 for details.

[2]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.

[3]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 many covariates affect optimal decisions, estimating policy functions from the data in a flexible way is quite difficult.

Using the estimated model, I conduct two counterfactual exercises to investigate the performance of the regulation. In my first counterfactual experiment, I examine the effect of introducing a centralized trading platform that eliminates all transaction costs. This simulation quantifies the first-best outcome under the cap-and-trade. While previous works, including Carlson et al. (2000) and Gollop and Roberts (1985), conduct such an exercise in a static framework, my analysis focuses on dynamic decisions such as abatement investment and permit banking.[4] I find that the centralized platform would lead to a more dispersed distribution of emissions and investment, and less banking of emissions permits. These effects reflect more active trading of emissions permits. The average abatement costs would decrease by 18% under the centralized platform, implying that "unrealized" gains from trade are significant in my sample period.

In the second counterfactual experiment, I examine the equilibrium outcome when permit banking between Phase I and II is not allowed. I find more active trading of emissions permits in the absence of permit banking. Although active trading could lead to a more efficient allocation of emissions in the cross-sectional sense, permit banking allows for the smoothing of abatement costs across periods. In my simulation result, the latter effect dominates the former, with permit banking reducing the average abatement costs by 2.8%.

The paper proceeds as follows. I first briefly review the related literature. Section 2 provides some background on the Acid Rain Program and a descriptive analysis of the data. Motivated by the descriptive findings, I introduce the model in Section 3. I then present the estimation methodologies and the results of the model in Sections 4 and 5, respectively. Section 6 presents the counterfactual experiments, through which I evaluate the consequences of the Acid Rain Program. Section 7 concludes the paper.

Related Literature My study is related to three strands of literature. First, my study contributes to the empirical literature on dynamic investment behavior (see, e.g., Ericson and Pakes, 1995, Bajari et al., 2007, Ryan, 2012, Collard-Wexler, 2013, and Kalouptsidi, 2014 in an oligopolistic setting, and Rust, 1987 and Aguirregabiria and Mira, 2002 in a competitive setting). My empirical setting 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. Though my model is tailored to cap-and-trade regulation, the model can be used to analyze firms' dynamic incentives when they are subject to frictions or imperfections in input markets. The model also can be applied to other market-based environmental regulations, such as the CAFE credit trading program and green certificate trading in the Renewable Portfolio Standard.

My study is also related to the empirical literature on cap-and-trade programs. Much of

[4] Carlson et al. (2000) and Gollop and Roberts (1985) quantify the abatement 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. Therefore, the marginal abatement costs would be equalized under a centralized trading platform.

this literature tests qualitative predictions of models of permit trading.[5] A few recent works take a structural approach to measure the welfare implications of cap-and-trade programs. In the context of NO_x regulation, Fowlie (2010b) estimates a model of abatement choice to study the implications of rate-of-return regulation on permit trading program. Fowlie et al. (2014) construct and estimate a dynamic game model of investment and entry/exit decisions to discuss the implications of hypothetical market-based environmental policies in the US cement industry.[6]

A distinctive feature of my paper is to model trading and banking decisions in a dynamic equilibrium framework.[7] Existing studies assume frictionless permit markets and stationary regulatory environment. In such a setting, cap-and-trade is equivalent to imposing a Pigouvian tax. My model describes how firms make decisions on investment, trading, and banking when transaction costs exist and regulatory standard is changing 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 on future permit prices, and examines the implications of these beliefs using a single-agent dynamic model of abatement decisions and permit trading. In contrast, this study provides a dynamic equilibrium framework in which permit prices are determined endogenously.

Finally, my study provides new insights for the evaluation of the Acid Rain Program by studying the intertemporal aspects of firms' behavior and the regulatory design. One approach adopted in the literature is to calculate the cost saving resulting from permit trading by estimating a cost function and a discrete choice model for abatement choices (see, e.g., Ellerman et al., 2000, Carlson et al., 2000, Keohane, 2006, and Chan, 2015). Researchers found that adopting a permit trading program led to significant cost savings, 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 et al. 1998, Ellerman and Montero 2007, and Helfand et al. 2006).[8] A recent paper by Chen (2018) examines the distortion in beliefs over future permit prices as a source of inefficiency. My study complements the previous works by empirically examining the dynamic aspects of compliance and abatement decisions under a cap-and-trade program.

[5] 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).

[6] Dardati (2014) studies how an allocation scheme for closing plants affects entry/exit decisions, using a calibrated model of industry dynamics in the context of the Acid Rain Program.

[7] An on-going work of Cantillon and Slechten (2015) studies participation decisions and the price formation of CO_2 permit prices using trading data from the EU-ETS scheme.

[8] Joskow et al. (1998) finds 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. Ellerman and Montero (2007) argues for an efficient market of permits by comparing the actual and theoretically predicted volume of aggregate banking. Helfand et al. (2006) uses monthly permit prices for the period 1994 to 2003 to test whether the price path follows the Hotelling r-percent rule for intertemporal arbitrage. They reject the Hotelling rule, which suggests there is inefficiency in the market.

In particular, my study is the first to quantify the role of permit banking system. Previous studies (e.g., Ellerman et al., 2000) note the importance of permit banking as a source of cost efficiency. For this purpose, I construct an equilibrium model of the cap-and-trade program that allows 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 with the Clean Air Act Amendments of 1970, such regulations have not been effective in reducing SO_2 emissions.[9] The failure of previous regulations led to the introduction of the Acid Rain Program (ARP), a cap-and-trade program, as part of Title IV of the 1990 Clean Air Act Amendments. The program began in 1995.

The target of the regulation is 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 especially dirty and old before the regulation, as well as an additional 182 EGUs from the Non-Table 1 group as substitution or compensating units. In PhaseⅡ (begun in 2000), all eligible EGUs are mandated to comply with the regulation.

The ARP aims to reduce SO_2 emissions by 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 incumbent units. The EPA adopts a rule that determines the unit-level allocation of emissions permits based on the characteristics of a unit.[10] The allocation is determined primarily by the average heat input during the period 1985–1987 and the target emissions rate (i.e., emissions per fuel input) for each phase. Specifically, the target emissions rate 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 also obtain an additional allocation of permits based on technical and political considerations (Joskow and Schmalensee, 1998). Note that the provisions in the 1990 legislation include detailed rules for permit allocations. Thus, generation facilities knew the schedule of permit allocation before the program started in 1995.

SO_2 permits are tradable goods. Firms can sell or buy permits with other firms, including financial companies or brokers that do not own any generating units and, thus, are not

[9] Ellerman et al. (2000) provide a brief history of the regulation on SO_2 emissions.

[10] 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 (1993a,b) for the details.

required to comply with the regulation. 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 way 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.[11] At the end of the 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.[12] The remaining permits are carried over to the next year, 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 emissions permits were allocated to existing units without cost, most still needed to decrease their emissions from their existing levels 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).[13] The latter option of reducing emissions rates was the primary channel of abatement, as explained in Section 2.3.2.

2.2 Data Sources

In this study, I focus on the period 1995 to 2003. Although the ARP continued after 2004, the proposal by the Clean Air Interstate Rule, announced in December 2003, had a large impact on regulated firms' expectations over the future regulatory environment. 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. Firms also started to invest in scrubbers in anticipation of a stricter regulation.[14] Thus, I do not include data after 2004, focusing instead on those periods when the regulatory environment for SO_2 emissions was stable.[15]

[11] 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 `https://web.archive.org/web/20090211082920/http://epa.gov/airmarkets/emissions/continuous-factsheet.html` for the details.

[12] 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.

[13] 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.

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

[15] One might be concerned that the regulator (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 so, new legislation would need to be passed by Congress.

My data are a combination of transaction data for emissions permits and various data on electricity production. The data on permit transactions are taken from the Allowance Tracking System (ATS), operated by the EPA. The latter data are compiled from various databases of the EPA and the US Energy Information Administration (EIA).

First, the EPA uses 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.[16] I constructed the transaction data at the firm and year levels from the database.[17] Specifically, I aggregated the account-level data into firm-level data using ownership information taken from various sources, including the EGrid database and EIA-860. The final data set includes the (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.[18]

The ATS does not collect information on transaction prices. Instead, I collected the market-price index of SO_2 permits provided by Cantor Fitzgerald, one of the biggest brokers in SO_2 permit markets. The frequency of the price data is monthly. I explain the price data further in Section 2.3.5.

The second part of my data set includes production data for the electricity companies. Here, I combined multiple databases to construct the data set, including EPA data and EIA survey data. 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. 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" 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" provides plant-level and monthly level information on fuel procurement, including fuel type, sulfur content, heat content, and purchase costs.

[16] 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 et al. (1998) for details.

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

[18] 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 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.

2.3 Descriptive Analysis

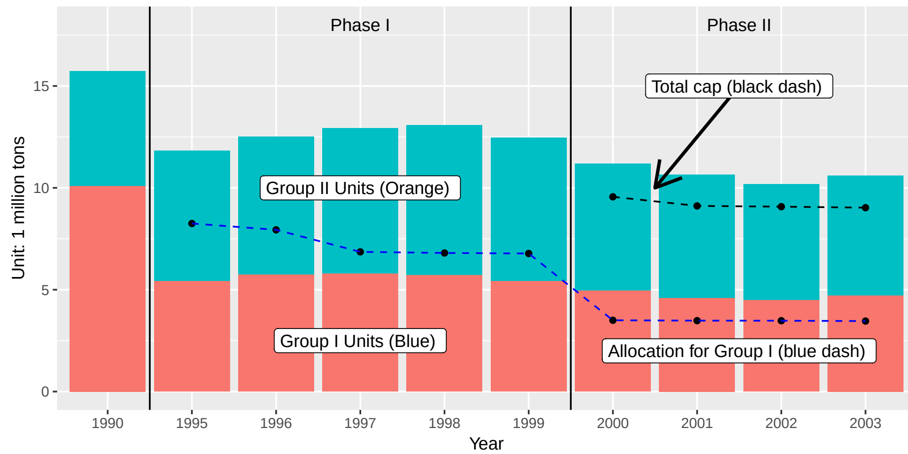
I now provide a descriptive analysis of my data set. I focus on various aspects of the ARP, including the banking of emissions permits, abatement decisions of regulated sources, and market for emissions permits. These descriptive findings motivate the modeling approach introduced in Section 3.

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 show the emission levels each year, and the dashed lines show the emission caps. 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 shows that Group I units reduced their emissions by almost half compared with their 1980 levels, once Phase I started in 1995. While both Group I and II units reduced their emissions further in 2000, the first year of Phase II, Group I units did not reduce their emissions by as much as they did in 1995. The emissions before 1999 were significantly lower than the emissions cap. However, after 2000, the aggregate emissions exceed the annual allocation of emissions permits, implying that Group I units saved their permits in Phase I, and then began using them after 2000 to ensure compliance. Note that the regulator had already announced the allocation schedule in 1990. Thus, regulated firms behaved in a forward-looking manner, taking this schedule into account.

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 shows the permit allocation for Group I units, and 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 (2) reducing the emissions rate (i.e., emissions per input).[19] However, the former strategy was not a major option for coal units regulated under the Acid Rain Program. In Appendix A, I run a difference-in-differences (DID) regression. Here, I exploit the variation of the timing of the regulation across units to estimate the effect of the regulation on the output (utilization rate) of EGUs. I found that the utilization rate decreased by only 2%-6% after the introduction of a cap-and-trade program.[20] In this subsection, I explain the abatement strategy of adjusting the emissions rate of EGUs whose primary fuel type is coal.[21]

Two common options are available to reduce the emissions rate of 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

[19]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.

[20]I focus on the intensive margin of operation, and treat entry/exit as given. Although the retirement of coal units is 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.

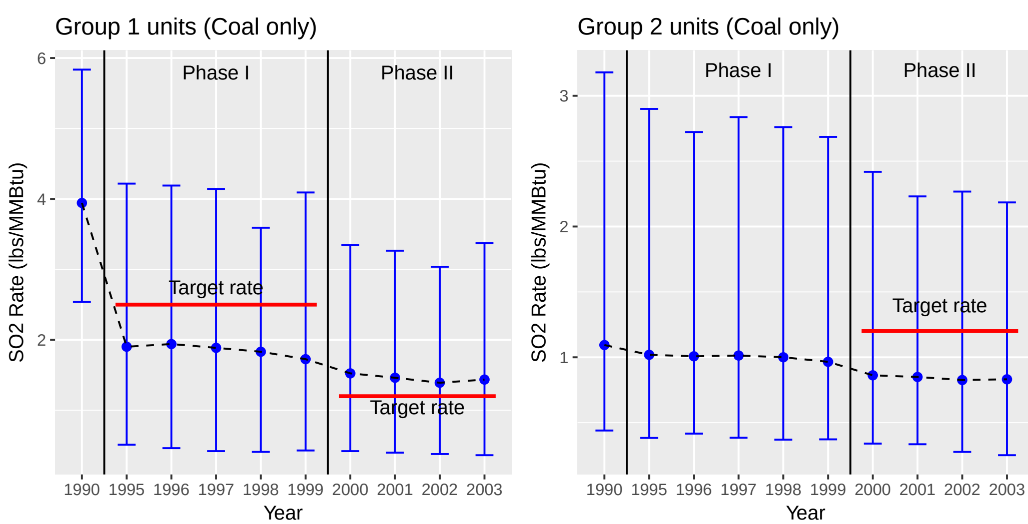
[21]Note that although the target of the ARP includes all types of fossil fuel units (coal, gas, and oil), SO_2 emissions from gas and oil units are relatively small, and no room remains to reduce the emissions rate of these units.

eliminates more than 80% of SO_2 emissions. However, this option incurs large a investment cost, as well as a long lead time (two to three years, on average).

Figure 2 shows the distribution of unit-level SO_2 emissions rates (measured in pounds per MMBtu) for each group in selected years. The left panel shows the distribution for Group I sources. The emissions rates of these sources decreased between 1990 and 1995, the beginning of Phase I. The rates then stayed almost constant during Phase I, and decreased further in 1999, anticipating the beginning of Phase II. The emissions rates of generating units in Group II did not change until 1999, and then decreased in 2000, the first year of the cap-and-trade program for these units. These observations imply that firms adjusted their emissions rates at the beginning of each phase, but that the emissions rates remained almost constant within the phase. This observation motivates the timing of abatement investment in the model I introduce in Section 3.

Another important finding from Figure 2 is the flexibility of the compliance patterns. The red vertical lines show the target emissions rates in each phase. Generating units would need 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 thus needed to obtain 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



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

2.3.3 Heterogeneity of Regulated Firms

The heterogeneity of firms is an important factor in the evaluation of a cap-and-trade program. Firm heterogeneity 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 efficient than that in an autarky, where no emissions permits are traded.

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 emissions rate before the regulation started, and initial allocation. These factors affect the firms' abatement and trading decisions. For example, firms that had a higher emissions rate before the regulation needed to make a greater 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.

Table 1: Firm Heterogeneity

	Mean	St. Dev.	25 Percentile	Median	75 percentile
Emissions Rate for Group 1 units in 1990	3.28	1.49	2.30	3.28	4.36
Emissions Rate for Group 2 units in 1990	1.11	0.77	0.69	0.94	1.31
# units in Group 1	5.66	4.87	2.25	4.00	6.00
# units in Group 1	4.39	5.78	1	2.5	5
Generation capacity for Group 1 units	19,953.60	21,897.87	6,425.00	14,025.00	24,182.50
Generation capacity for Group 2 units	13,190.18	17,004.14	1,615.5	6,530	18,145
Initial allocation	75,207.8	105,611.5	17,214.5	41,449	88,605

Notes: There are 114 firms in the sample. The unit for emissions rate (in rows 2 and 3) is lbs per MMBtu. The emissions rate is calculated at the firm-level. The unit for generation capacity (in rows 6 and 7) is megawatts.

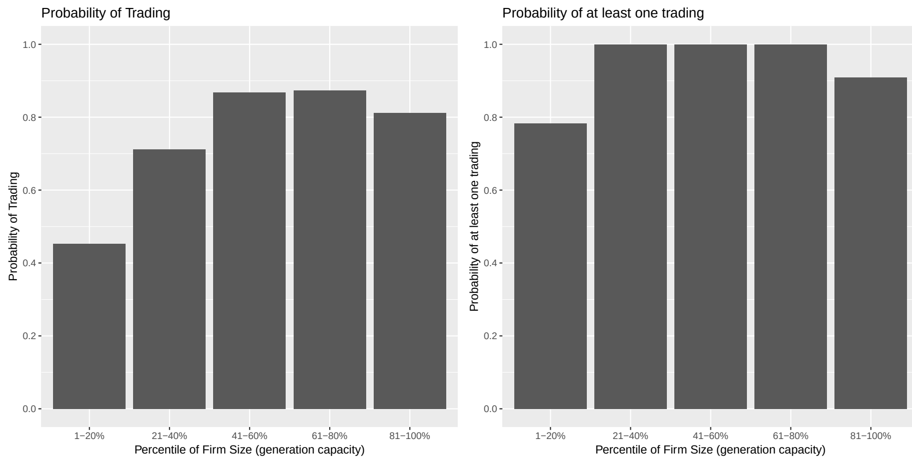
2.3.4 Firm-level Trading Information

I now explain firms' decisions related to the trading of emissions permits. U.S. Environmental Protection Agency (2004) reports that transactions of emissions permits between related entities (namely 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 shows the 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. 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.

The left panel shows that firms did not necessarily trade every year. The unconditional probability of conducting permit trading was 76%. 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 a suggestive evidence of fixed transaction costs, firms do not need to conduct a transaction in every period, owing to the storability of emissions permits. In the right panel, I show the firm-level experience of market trading during the sample period. As shown, 94% of firms had at least one experience of trading with another firm in the sample period, although some firms, most of which are small, did not trade at all.

Figure 3: Trading Pattern at Firm Level



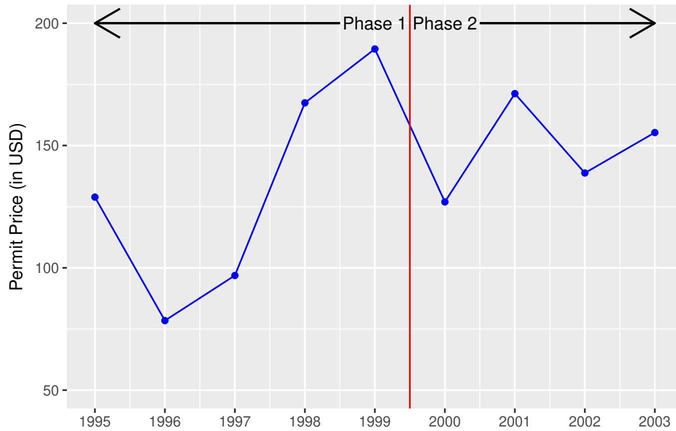
2.3.5 Price of Emissions Permits

As discussed in Section 2.1, there are no centralized trading exchanges for emissions permits under the Acid Rain Program. Although regulated firms need to 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 broker in this market. I use the monthly SO_2 price index as a price measure. Cantor Fitzgerald constructs this 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 median for each year. Note that the price is normalized to the 2000 level using the producer price index.

The price at the beginning was around USD 150, falling to below USD 100 in 1996 and 1997. Then, it increased to USD 200 in 1999, before fluctuating in the range USD 120–200 after 2000. The figure suggests that the market price reflects the availability of banking. In the absence of permit banking, I would expect to see a spike in the permit price between Phases I and II, because the target emissions rate in Phase II is much stricter than that in Phase I. Instead, the permit price increases gradually over time, though it is relatively volatile. [22]

[22] A key theoretical prediction on 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 et al. (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.

3 Model

3.1 Overview of the Model

This section introduces a model of regulated firms and permit market equilibrium under the cap-and-trade program. The model incorporates the descriptive findings in the previous section, including the nonstationary 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 model is set as a nonstationary and finite-horizon model, where time is indexed by $t = 1995, \ldots, 2003 (\equiv T)$. Each discrete decision period corresponds to one compliance year. Firms have a common discount factor of β. An overview of the firm-level decision problem is provided in Figure 5. The problem has two building blocks: (i) an investment in abatement options at the beginning of each phase (1995 and 2000), and (ii) decisions on production, trading, and banking in each year. At the beginning of each phase (i.e., 1995 and 2000), firms make a decision on abatement investment and determine the emissions rate R_i^k $(k = 1, 2)$. The emissions rates are assumed to be fixed within each phase. This assumption reflects the observation from Section 2.3 that the emissions rate changes at the beginning of each phase, and then stays constant within the phase.

Given the emissions rate, each firm i makes decisions on production, permit trading, and banking. The timeline of 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}.

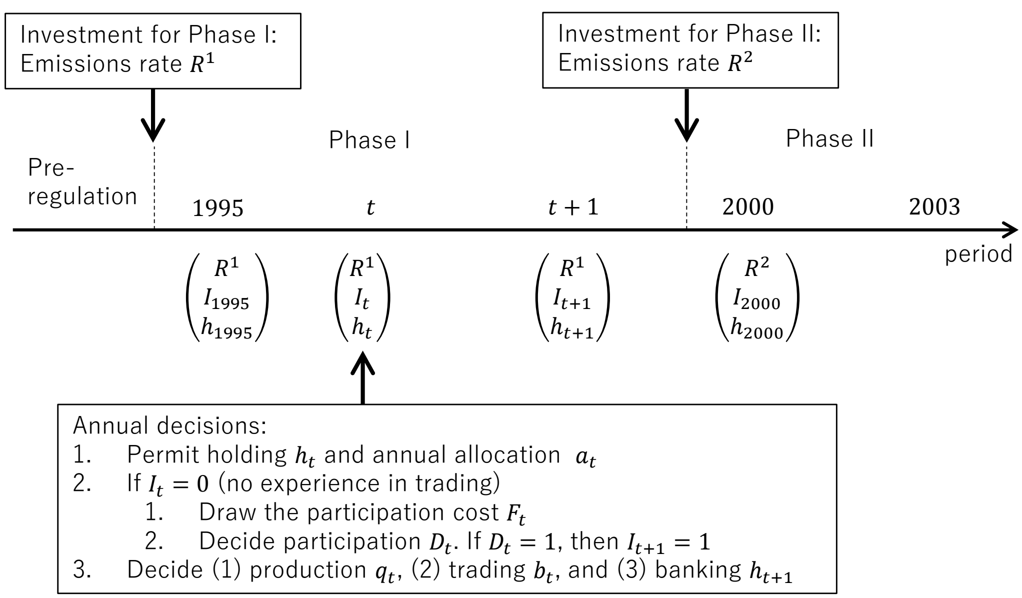
2. Participation decision: Denote firm i's experience of market trading by I_{it}; i.e., $I_{it} = 1$ if a firm has experience in market trading, and 0 otherwise. If $I_{it} = 0$, a firm can pay

the one-time sunk cost F_{it} to participate.

3. A firm chooses (i) the production quantity of each generating unit $\{q_{jt}\}_j$, (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 deciding on 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 the holding $h_{i,t+1}$.

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. I now explain each component of the model.

Figure 5: Overview of the Firm-level Decision Problem



3.2 Electricity Production and SO_2 Emissions

Firms earn profits from electricity production in a competitive electricity market. Firm i holds J_{it} units of the regulated sources, and chooses a production quantity q_{jt} for each generating unit j. The profit is given by

$$\pi_{it}\left(\{q_{jt}\}_j\right) = \sum_{j \in J_{it}} \left\{ (\tau_{st}^{elec} - c_{jt}^{fuel}(R_{jt})) \cdot q_{jt} - g(q_{jt}, k_j) \right\}, \tag{3.1}$$

where τ_{st}^{elec} is the electricity price in state s where unit j is located, and $c_{jt}^{fuel}(R_{jt})$ denotes the unit-specific fuel costs of production as a function of the emissions rate R_{jt}. I explain how the unit-specific fuel cost c_{jt}^{fuel} depends on the emissions rate R_{jt} shortly. Note that fuel costs account for around 75% of the total operating expenses (see EIA, 2012). Here, $g(q_{jt}, k_j)$ is the convex cost of production, and captures the increasing costs of operation near the capacity constraint (see, e.g, Ryan, 2012).

Note that the profit $\pi_{it}(\cdot)$ is the gross profit from electricity production, which excludes costs associated with permit trading. In their production decisions, though, firms should consider the cost of using emissions permits. I discuss the optimal decision on production quantity in Section 3.4.

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

$$e_{it}\left(\{q_{jt}, \rho_{jt}\}_j\right) = \sum_{j \in J_{it}} \rho_{jt} q_{jt}, \tag{3.2}$$

where ρ_{jt} is the unit-level SO_2 emissions rate per production.

I now explain how the profits π_{it} and emissions e_{it} depend on the emissions rate R_{jt}, defined as the emissions level per unit of fuel input (one MMBtu). First, ρ_{jt} in equation (3.2) is defined as

$$\rho_{jt} = HR_j \cdot R_{jt}.$$

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 fuel cost $c_{jt}^{fuel}(R_{jt})$ also depends on R_{jt} because cleaner coals tend to be more expensive. Specifically, the cost can be expressed as

$$c_{jt}^{fuel}(R_{jt}) = HR_j \cdot p_{jt}^{fuel}(R_{jt}),$$

where $p_{jt}^{fuel}(R_{j,t})$ is the fuel price per 1 MMBtu of fuel input. I estimate the hedonic function that predicts the fuel price as a function of the emissions rate R_{jt}.

As discussed in Section 2, firms can adjust the emissions rates of their coal units. In Section 3.5, I introduce the model of how firms determine their emissions rates by making an investment at the beginning of each phase. Note that I treat gas and oil units separately from coal units and R_{jt} for gas and oil units as fixed and exogenous in the model. This is because gas and oil units have quite low SO_2 emissions rates and are not able to reduce their emissions rates.

Aggregation of Unit-Level Emissions Rate at the Firm-Level When modeling investment decisions, a key issue is the dimensionality of the state space. If I model the investment decision at the generation unit level, the state variable should include the emissions rate for each generation unit, which would make the number of state variables high.

Because the median number of coal units is five in my sample, such a modeling approach would be subject to the curse of dimensionality and be intractable for the analysis.

To maintain the tractability of the model, I aggregate the unit-level emissions rates at the firm-level, and assume that each firm decides on a single emissions rate that is common across coal units within the same firm (i.e., $R_{it} \equiv R_{jt}$, for $j \in J_{it}$). Under this assumption, the number of state variables related to the emissions rate is just one. Table 2 examines the validity of this approach. In the third row, I report the firm-level emissions rate, which is equivalent to the weighted mean of the emissions rates across units. The fourth row shows the standard deviation of the unit-level emissions rates within a firm. Table 2 shows that the standard deviation is much smaller than the firm-level emissions rate. This implies that the variation in the emissions rate across coal units is small within a firm, thus supporting my modeling approach.

Table 2: Unit-level Variation of Emissions Rate of Coal Units at the Firm-level

	1995-1999		2000-2003	
	Mean	Std. Dev.	Mean	Std. Dev.
Emissions rate at the firm-level	1.53	0.96	1.05	0.83
Standard deviation of unit-level emissions rate within a firm	0.59	0.55	0.33	0.40

Notes: The unit is pounds (equivalently lbs) per 1 MMBtu.

3.3 Structure of Permit Trading and Transaction Costs

Each firm is allocated an annual allocation of permits a_{it} in each period. Because the allocation plan was announced before the regulation, the sequence of $\{a_{it}\}_t$ 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 decides on its emissions level e_{it}, which is determined by the production quantity $\{q_{jt}\}$ (equation (3.2)), net purchase volume b_{it}, and banking volume $h_{i,t+1}$. Here, 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

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

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

Note that equation (3.4) is the nonnegativity constraint of banking, and excludes the possibility of borrowing permits from a future allocation. I assume that firms achieve perfect compliance in my model, because the compliance rate under this regulation is nearly perfect.

I model the permit market as a competitive market with transaction costs. 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.[23] The presence of transaction costs reflects the fact that the majority of permit transactions were bilateral, because there were no centralized exchanges for emissions permits. Ideally, a model would incorporate bilateral trading of emissions permits across participants. However, such a model could be quite difficult to solve and analyze because emissions permits are divisible objects and my model also features dynamic investments in clean technology and permit banking. Thus, I capture the nature of the permit market by introducing transaction costs in a reduced form.[24]

Firms are price-takers in the permit market and face the market price P_t. In addition, they pay 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 participate in permit trading. 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 $F_{it} = F + \epsilon_{it}$, where F is the mean participation cost, and ϵ_{it} is an idiosyncratic cost shock that follows the type-I extreme value distribution $G(\cdot; \sigma_F)$.

Second, firms pay variable transaction costs associated with the net purchase of permits b_{it} (Stavins, 1995). This cost is given by $TC(|b_{it}|)$, where $TC(\cdot)$ is a differentiable and strictly convex function. Variable transaction costs include brokerage commissions and bid-ask spreads. The convex nature of the cost function also 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. Therefore, the firm needs to find a different trading partner, hence incurring a costly search process in a bilateral market. Convex transaction costs are employed in the theoretical finance literature (e.g., Gârleanu and Pedersen, 2013, and Dávila and Parlatore, 2017) and are motivated by empirical findings (see, e.g., Breen et al., 2002, Lillo et al., 2003, and Robert et al., 2012).

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 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 Section 4.3.

[23] 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.

[24] In the model, I consider that firm-level permit trading decisions incur transaction costs. This implicitly assumes that there are no costs associated with transactions across generating units within the same firm. The latter transactions simply reflect a reallocation of inputs within a firm, which should present decisions that are more flexible than those associated with permit trading with other firms. In fact, the trading of permits between related entities (reallocation) has been active since the implementation of the regulation, as mentioned in Section 2.3.4 (see, e.g., U.S. Environmental Protection Agency, 2004).

3.4 Optimal Choices of Production, Trading, and Banking

I now consider the optimization problem in year t. A firm makes both discrete (participation) and continuous decisions related to production, 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

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

Here, $EV_{it}(h_{it}, I_{it}, R_{it})$ 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 firm-level emissions rate is R_{it}. Recall that $I_{it} = 1$ if firm i has participated in the market previously, thus paying the participation cost. When a firm is a nontrader, 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, and 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}, R_{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}.

The optimality conditions for the traders are given by

$$\frac{\tau_{st}^{elec} - c_{jt}^{fuel} - g'(q_{jt})}{\rho_{jt}} = \lambda_{it} \tag{3.6}$$

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

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

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

where λ_{it} denotes the Lagrange multiplier on the transition of permit holding (3.3), and μ_{it} denotes the Lagrange multiplier on the nonborrowing constraint (3.4). I call λ_{it} the shadow value of emissions permits for firm i.

Equation (3.6) determines the optimal production decision given the shadow costs of emissions permits. The left-hand side is the marginal profit from additional emissions, which should be equal to the shadow costs of emissions permits λ_{it} at the optimum.

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 is 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 production $\{q_{jt}\}_j$, trading b_{it}, and banking $h_{i,t+1}$.

The optimality conditions for nontraders are the same as above, except $b_{it} = 0$ and I do not have equation (3.7). These conditions imply that, in this case, 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).

Next, 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 $F_{it}(= F + \epsilon_{it})$. A firm participates in the market if $V_{it}^1(h_{it}, R_{it}) - (F + \epsilon_{it}) > V_{it}^0(h_{it}, R_{it})$. This optimal decision leads to the participation probability given by

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

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 nontraders, I now provide the value function. Let $V_{it}(h_{it}, I_{it}, R_{it}, \epsilon_{it})$ be the value function after observing the random draw of the participation costs. The value function is given by

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

Finally, the ex ante value functions $EV_{it}(h_{it}, I_{it}, R_{it})$ (before observing ϵ_{it}) are given by the integral of V_{it} with respect to ϵ_{it}.

Continuation Value at the Terminal Period My model has a finite time period, and the terminal period T corresponds to the year 2003, which is the last period of my sample. However, the Acid Rain Program continued after 2003, and the level of banking at the end of 2003 was still substantial. To deal with this issue, I include the reduced-form continuation value function $CV_{T+1}(h_{i,T+1})$ in the model. This term captures the banking incentive at the terminal period $T(= 2003)$. In Section 4.2, I provide the functional form of $CV_{T+1}(h_{i,T+1})$, and estimate it along with the other parameters.

3.5 Investment Decisions on Emissions Rate

A firm decides on the phase-specific emissions rates R_i^k $(k = 1, 2)$ at the beginning of each phase. This implies that $R_{it} = R_i^1$ if $t \in \{1995, \ldots, 1999\}$ and $R_{it} = R_i^2$ if $t \in \{2000, \ldots, 2003\}$. The lower the emissions rate, the higher is the level of investment. I further assume that the emissions rate is a continuous choice variable. I denote the cost function of investment by $\Gamma(\bar{R} - R)$, where R is the emissions-rate level chosen by a firm, and $\bar{R}$ is the emissions rate before the investment.

The investment problem for Phase I is given by

$$\max_{R_i^1} \; EV_{i,1995}(0, 0, R_i^1) - \Gamma(R_i^0 - R_i^1) \;\; \text{s.t. } R_i^1 \leq R_i^0, \tag{3.10}$$

where R_i^0 is the emissions rate in 1990, that is, before the regulation. I incorporate the adjustment costs and the irreversibility of investment by allowing R_i^0 to affect both the investment cost and the choice set of the emissions rate R_i^1. Note that $h_{i,1995} = 0, I_{i,1995} = 0$, by definition. The problem for Phase II is defined similarly, except that the investment cost now depends on R_i^1, which is determined endogenously in Phase I.

3.6 Dynamic Competitive Equilibrium of Permit Trading

I now define an equilibrium for 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 of this assumption in Section 3.8.

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 $\{\{q_{jt}^*\}_{j \in J_{it}}, b_{it}^*, h_{i,t+1}^*\}_{t=1995}^{2003}$ and $\{R_i^{1*}, R_i^{2*}\}$, given a sequence of permit prices, and*
2. *[Market Clearing] $\sum_i b_{it}^* \left(\{P_t\}_{t=1995}^{2003}\right) + \bar{B}_t^{fringe}(P_t) = 0$ 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 by backward induction for all firms, and (ii) calculate the aggregate level of net purchases to check 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 finds an equilibrium vector of permit prices satisfying the market-clearing conditions. I explain how to numerically compute a vector of equilibrium prices in Appendix E.3.[25]

[25] Although I do not have a formal proof for the uniqueness of the equilibrium, I try different initial prices of emissions permits when I numerically solve a dynamic competitive equilibrium. Thus, I confirm that these initial values converge to the same equilibrium prices.

3.7 Model Implications

This subsection discusses the implications of the model. In particular, I argue the role of transaction costs, $TC(\cdot)$ and F_{it}, in firms' dynamic decisions and equilibrium outcomes.[26] To discuss this point, I first show the benchmark case in which no transaction costs exist; namely, $T(\cdot) = 0$ and $F_{it} + \epsilon_{it} = 0$. In this case, the optimality conditions (3.6)–(3.9) can be summarized as[27]

$$\begin{aligned} \frac{\tau_t^{elec} - c_{jt}^{fuel} - \frac{\partial g(q_{jt},k_j)}{\partial q_{jt}}}{\rho_{jt}} &= P_t. && (3.11) \\ P_t &= \beta P_{t+1} + \mu_{it},\ \mu_{it} \geq 0 \perp h_{i,t+1} \geq 0. && (3.12) \end{aligned}$$

The role of Transaction Costs in Avoiding the Indeterminacy of Trading/Banking Decision Equation (3.12) implies that the equilibrium permit prices P_t should increase over time at the rate β^{-1}, as long as the banking volume is positive and no transaction costs exist. This property is known as the Hotelling $r-$percent rule, which states that the price of an exhaustible resource should increase at a rate equal to the inverse of the interest rate (see, e.g., Rubin, 1996).

More importantly, the model suffers from indeterminacy of individual optimal decisions: 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, all choices are equivalent for individual firms, as long as a firm can produce the level of emissions given by the optimality condition on production quantity (3.11).

I now consider the case in which transaction costs are present. Combining optimality conditions (3.7) and (3.8) and using 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}. Intuitively, $TC(b_{it})$ prevents firms from engaging in complete intertemporal arbitrage. Without convex 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 infinitely many permits) in period t.

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

[26] 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.

[27] 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 Appendix C for the derivation.

from net purchases is increasing owing to the convex transaction costs $TC(b_{it})$. Intuitively, buying additional permits becomes more difficult. The discounted marginal value from banking, given by $\beta\left(P_{t+1}+\frac{dTC(|b_{it}|)}{db_{it}}\right)$, is decreasing 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 decreasing as firms sell additional permits because they have to pay transaction costs.

Role of Transaction Costs in Efficiency One of the virtues of a cap-and-trade regulation is that the equilibrium allocation of emissions, given the emissions cap, is efficient in the absence of transaction costs. This assertion is confirmed by equation (3.11), which implies that the marginal profit from producing one unit of emissions is equalized across firms at the level of permit price P_t. The key mechanism is that all firms are facing the same shadow value, given by the market price (i.e., $\lambda_{it}=P_t, \forall i$).

I now examine how the trading behavior affects the shadow costs of emissions permits and leads to an inefficient outcome of emissions in the presence of transaction costs. 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

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

The inequalities hold because $\frac{dTC(|b_{it}|)}{db_{it}}>0$ for $b>0$, and $\frac{dTC(|b_{it}|)}{db_{it}}<0$ for $b<0$. Intuitively, in the presence of variable transaction costs, buyers must pay 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 profit of emissions for the buyer is strictly higher than that for the seller. In other words, buyers produce less and sellers produce more than the efficient level at which the marginal profits of the two firms are equalized.

The heterogeneity of the shadow value has an implication for firms' dynamic decisions, namely, on their investment behavior and permit banking. The return on investment is determined by the marginal abatement of emissions, given by $\sum_j HR_{jt}q_{jt}$, evaluated at the the shadow value of λ_{it}.[28] 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 face a lower incentive to bank permits because the transaction costs lead to a higher shadow value today. On the other hand, sellers prefer to bank additional permits because of the lower shadow value. However, the aggregate level of permit banking can be higher or lower than the first-best case when no transaction costs exist.

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

3.8 Discussion of Modeling Assumptions

Output price τ_{st} I assume that output price τ_{st} is given as exogenous. Because the main target of the Acid Rain Program is coal units, which are typically infra-marginal units, the program does not affect the electricity price in the wholesale market.[29]

Public Utility Regulation An important factor when evaluating an environmental regulation in the electricity industry is the presence of a public utility regulation (rate-of-return regulation). The literature examines how a public utility regulation affects the performance of market-based environmental regulations, including the Averch–Johnson effects on compliance decisions (Fowlie, 2010b in the NO_x regulation and Cicala, 2015 in the SO_2 regulation) and the regulatory uncertainty associated with the cost-recovery rule of permits (see, e.g., Arimura, 2002).

Although my framework does not explicitly model the presence of a public utility regulation, it accounts for such an effect through the transaction costs of emissions permits. Given that the above two channels make firms less likely to trade emissions permits and more likely to reduce their emissions for compliance, the effect of a public utility regulation should be captured by the transaction costs of permit trading.

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 CAIR 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.[30]

No Aggregate Uncertainty The model assumes perfect foresight on the evolution of the profit π_t, namely, electricity demand and production costs.[31] This assumption implies that permit prices P_t are deterministic in equilibrium and therefore firms should have perfect foresight over permit prices. Under this assumption, the value function V_t, indexed by time script t, subsumes those deterministic state variables. The state variables that I need to track down to solve the individual optimization problem are permit holding h_{it}, trading experience I_{it}, and firm-level emissions rate R_{it}. Although perfect foresight is certainly a

[29] See Fowlie (2010b) for a similar discussion in the context of NO_x regulations.

[30] Alternatively, I could model the terminal period as a stationary and infinite-period dynamic programming problem by assuming that the same regulatory environment continues and the CAIR would not be introduced. Although this approach would allow me to avoid specifying a parametric form of the continuation value function, it is computationally more demanding to solve the firm's optimization problem.

[31] 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.

strong assumption, it reduces the computational burden and makes the model tractable for estimations and simulations. Incorporating aggregate uncertainty about demand and production makes the model intractable for two reasons.

First, incorporating a stochastic transition of demand, costs, and permit prices increases the dimension of the state space. In addition to (h_{it}, I_{it}, R_{it}), I would have to consider the transition of the profit function π_{it} and the permit price P_t. If I rewrite the profit function as $\pi_{it}(\{q_{jt}\}_j) \equiv \pi_i(\{q_{jt}\}_j; D_t, C_t)$, 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. Because my framework incorporates the rich heterogeneity of regulated firms, I need to solve the dynamic optimization problem separately for each firm. Therefore, expanding the state space would make the model more difficult to compute and estimate.

Another significant problem is how to model the transition of equilibrium permit prices. This problem was first identified by Krusell and Smith (1998), and is both conceptually and computationally difficult. With the aggregate uncertainty of demand and costs, the equilibrium permit prices become random variables. Thus, firms need to form an expectation over future equilibrium permit prices. In a rational expectation, firms need to track all information that forecasts the permit prices tomorrow P_{t+1}. Permit prices are determined by the market-clearing condition, which consists of the net purchases of emissions permits by every firm (i.e., $b_{i,t+1}$ for all i). Net purchase depends on the state variable (h_{it}, R_{it}, I_{it}), permit allocation a_{it}, and profit function π_{it}. Therefore, in principle, firms must know and incorporate the complete cross-sectional distribution of the state variable in the state space in order to form a rational expectation of future permit prices. Because there are 114 firms in my sample, this approach would be infeasible due to the curse of dimensionality. This issue has been known 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 et al. (2015) for a dynamic demand model for new and used car markets.

Given these complications, I choose to assume perfect foresight. As a result, the model can incorporate other important considerations, especially firm heterogeneity, while still being tractable for estimation and simulation analyses. I leave this extension for future work. [32]

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

4 Estimation Strategy

This section explains the estimation strategy for the model. An estimation follows three steps. First, I obtain the profit function $\pi_{it}(\{q_{jt}, \rho_{jt}\}_j)$ by estimating the convex cost of production $g(q, k)$ and the Hedonic function for coal price $p^{fuel}(R)$ without solving the dynamic decision problem. I will explain the details in estimation of $g(q, k)$ in section 4.1. Appendix B discusses estimation of the Hedonic function.

Using the estimated profit function $\pi_{it}(\{q_{jt}, \rho_{jt}\}_j)$, I then estimate the other model parameters in Section 4.2, including the variable transaction costs $TC(b)$, the distribution of the fixed transaction costs F_{it}, the continuation value at the terminal period $CV_{T+1}(h_{i,T+1}, R_i^2)$, and the costs of abatement investment $\Gamma(\bar{R} - R)$. 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. Lastly, I estimate the fringe demand $\bar{B}_t^{fringe}(P_t)$ in Section 4.3.

Identification Identification of the model relies on optimality conditions related to firms' decisions and detailed information on production and permit transactions. First, I use production data before and after the introduction of the cap-and-trade program to obtain the firm-level profit function from the electricity production. The profit function implies the marginal profit from emissions across firms. Without variable transaction costs, the marginal profits should be equalized across firms and equal to the permit price. Variable transaction costs are identified from how the marginal profits vary with the trading volume. Firm-level participation in permit trading is employed to identify the sunk costs of participation. Finally, I identify the costs of investment using the optimality condition for investment decisions.

4.1 Step 1: Estimation of the Profit Function

The purpose of step 1 is to estimate "offline" the profit function $\pi()$, which will be embedded in step 2. Here, I focus on estimating the convex cost function $g(q, k)$. I parameterize the function as

$$g(q_{jt}, k_j) = \frac{k_j}{2\gamma} \left(\frac{q_{jt}}{k_j} \right)^2, \tag{4.1}$$

which mimics the specification used in Ryan (2012).

I utilize the FOC for the unit-level production quantity q_{jt}, given by (3.6). Specifically, the FOC yields the following linear equation:

$$cf_{jt} = \gamma \left(\tau_t^{elec} - c_{jt}^{fuel} - \lambda_{it}\rho_{jt} \right),$$

where $cf_{jt} \equiv q_{jt}/k_j$ is the capacity factor. This model captures the relationship between the capacity factor and the markup of electricity production, $\tau_t^{elec} - c_{jt}^{fuel} - \lambda_{it}\rho_{jt}$.

In the empirical implementation, I use monthly level observations instead of yearly observations. In addition, the sample includes generation units that are not affected by the Acid Rain Program. For example, the sample includes observations before 1995 (i.e., before the ARP started), as well as observations for units that were not regulated at a particular time point (e.g., observations of Group II units before 2000). To accommodate these observations, I consider the following form of the regression equation, indexed by month m:

$$cf_{jm} = \gamma \left(\tau_m^{elec} - c_{jm}^{fuel} - \mathbf{1}\{SO_2reg\}_{jt} \cdot \lambda_{it}\rho_{jt} \right) + u_j + u_m + u_{jm}, \tag{4.2}$$

where u_j is a unit fixed effect, u_m denotes the time fixed effects, and u_{jm} is an error term. The dummy variable $\mathbf{1}\{SO_2reg\}_{jt}$ takes the value one if unit j is under the ARP in year t.

In equation (4.2), I observe output price τ_{jm}^{elec}, fuel costs c_{jt}, and emissions rate per production ρ_{jt} in the right-hand-side . However, I cannot directly observe the firm-level shadow costs λ_{it}, which are determined endogenously in the model. To solve this problem, I proxy λ_{it} using the optimality conditions implied from the model in the spirit of Olley and Pakes (1996). The model implies that λ_{it} can be written by some unknown function of the state variables, i.e., $\lambda_{it} = G(h_{it} + a_{it}, I_{it}, R_{it})$. I approximate this function using the second-order polynomial;

$$\begin{aligned} \lambda_{it} \quad = \quad & \theta_1 \log(h_{it} + a_{it}) + \theta_2 I_{it} + \theta_3 \log R_{it} + \theta_4 \{\log(h_{it} + a_{it})\}^2 + \theta_5 \{\log R_{it}\}^2 \\ & + \theta_6 I_{it} \cdot \log(h_{it} + a_{it}) + \theta_7 I_{it} \cdot \log R_{it} + \theta_8 \log R_{it} \cdot \log(h_{it} + a_{it}) \end{aligned}$$

I substitute this equation into equation (4.2) and estimate the resulting linear equation.

4.2 Step 2: Estimation of Remaining Parameters

The estimation in step 1 yields the profit function $\pi_{it}(\{q_{jt}\}_j)$. The next step is to estimate the remaining parameters, including the transaction costs, continuation value, and investment costs. I first provide specifications for these primitives.

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

$$TC(|b|) = \frac{1}{2} \exp \left(\eta_0^{buy} \mathbf{1}\{b > 0\} + \eta_0^{sell} \mathbf{1}\{b < 0\} + \eta_1 \log(size_i) \right) b^2, \tag{4.3}$$

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 level of transaction costs. Furthermore, I allow for the possibility that transaction costs can vary depending on the firm that is buying or selling permits, with η_0^{buy} and η_0^{sell}.

The participation cost F_{it} is specified as

$$F_{it} = F + \epsilon_{it}, \tag{4.4}$$

where ϵ_{it} follows an i.i.d. type-I extreme-value distribution, with standard deviation σ_F. Here, $size_i$ denotes the size of firm i, measured as the sum of the generation capacity of firm i.

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

$$CV(h_{i,T+1}, R_i^2) = \exp\left(\alpha_0 + \alpha_1 \log(size_i) + \alpha_2 R_i^2\right) h_{i,T+1}^{\alpha_3}. \tag{4.5}$$

The coefficient depends on the firm size, $size_i$, and on the emissions rate in Phase II, R_i^2. These variables capture the heterogeneity in the incentives to bank in the terminal period.

The specification for the investment cost $\Gamma(\cdot)$ is given by

$$\Gamma(\bar{R} - R) = \frac{exp(\zeta_0 + \zeta_1 \log(K_{i,\tau}))}{2} (\bar{R} - R)^2, \tag{4.6}$$

where $K_{i,\tau}$ is the generation capacity of the units regulated in Phase $\tau \in \{I, II\}$. The parameters estimated in this step are summarized as $\boldsymbol{\theta} = (\eta_0, \eta_1, F, \sigma_F, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \zeta_0, \zeta_1)$.

The estimation procedure builds on the literature of estimation of dynamic structural models in industrial organization and labor economics.[33] 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 choice variables and match the prediction with the data. Here, 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. The observed prices are used to solve the single-agent optimization problems, which are much easier to solve than the dynamic competitive equilibrium.[34]

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 by 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 $Prob_{it}^{enter}(s)$.

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

[34] 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. Then, they solve the optimal response of a player, given the estimated beliefs, to construct the pseudo-likelihood function.

5. Then, the prediction for firm i in year t is given by

$$\hat{x}_{it} = \sum_{s \in \{\emptyset, 1995, \cdots, 2003\}} Prob_i^{enter}(s) \hat{x}_{it}(s). \tag{4.7}$$

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 the emissions, net purchases, and permit banking in each year, 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.

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

4.3 Step 3: Estimation of Fringe Demand

I now estimate the fringe demand function. I consider the following specification, with constant elasticity:

$$\log \left(-\bar{B}_t^{fringe} \right) = \phi_0 + \phi_1 \log P_t + \phi_2 Phase2_t + e_t, \tag{4.8}$$

where $Phase2_t$ is the dummy for Phase II. I take the negative of $\bar{B}_t^{fringe}$, because $\bar{B}_t^{fringe} < 0$ for all years; that is, fringe firms are net sellers of permits. I estimate the model using the sum of the initial allocations for the firms in my sample as an instrument for P_t.

5 Estimation Results

Table 3 presents the parameter estimates of the structural model. The model incorporates two types of transaction costs: participation and variable costs. 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 $exp(\eta_0 + \eta_1 \log(size_i))|b_{it}|$. The mean of the costs is USD 52.6, and the median is USD 36.1. Considering that the permit prices range between USD 100 and USD 200, these numbers are substantial. This estimate indicates the large dispersion of the shadow value of emissions across firms, implying an inefficient outcome of the cap-and-trade in the baseline.

The mean and the standard deviation of the participation costs is around USD 0.45 million and USD 2.67 million, respectively. The estimated parameters in the continuation value function at the terminal period imply that bigger and cleaner firms obtain a higher value from banking, though the coefficients are small and imprecisely estimated. Estimates of the investment cost have reasonable signs. The bigger the capacity, the higher are the investment costs. Finally, the fringe elasticity is estimated to be 1.34.

Table 3: Parameter Estimates

	Parameter	Description	Estimate	Standard Errors
Convex Production Costs $g(\cdot)$	γ	Curvature	0.000647	0.000131
Variable Costs $TC(\cdot)$	η_0^{buy}	Constant for buying	-4.214	1.653
	η_0^{sell}	Constant for selling	-3.958	1.135
	η_1	Firm size	-0.068	0.063
Participation Costs F_{it}	F	Mean (USD 1 million)	0.455	0.200
	σ_F	Std. Dev. (USD 1 million)	2.673	2.266
Continuation Value $CV_{T+1}(\cdot)$	α_0	Constant	3.482	1.872
	α_1	Firm Size	0.097	0.186
	α_2	Emissions Rate	-0.046	0.270
	α_3	Curvature	0.051	0.095
Investment Costs $\Gamma(\cdot)$	ζ_0	Constant	9.504	3.794
	ζ_1	Capacity	0.838	0.352
Fringe Demand $\bar{B}_t(\cdot)$	ϕ_0	Constant	5.281	14.399
	ϕ_1	Elasticity	1.335	2.991
	ϕ_2	Phase II dummy	0.902	1.000

6 Counterfactual Experiments

This section provides a series of counterfactual exercises using the estimated model. I first quantify the first-best outcome of cap-and-trade by eliminating transaction costs (Section 6.1). Then, I evaluate the impact of the permit banking system in Section 6.2. Appendix F explains how to simulate the equilibrium outcome in each case.

6.1 The Potential Gains from Trade

I first simulate the outcome if the regulator introduced a centralized exchange for emissions permits that eliminates all transaction costs. Such a case is the first-best outcome in terms of abatement costs. By comparing the first-best outcome with the baseline outcome, I evaluate how well the cap-and-trade program works in the baseline.[35]

As discussed in Section 3.7, 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, and then solve the market equilibrium.

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 the excess permit banking in the baseline case, which is consistent with the concern argued by Smith et al. (1998). Under the presence of transaction costs, firms are less active in permit trading and, thus, prefer to save emissions permits. However, such an incentive would lower the allocative efficiency of emissions permits because the banked permits should be used by those who have a higher willingness to pay, namely buyers.

Figure 7 plots the distributions of emissions rates 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, making 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 4 shows the efficiency measures of the equilibrium outcome with and without transaction costs. With regard to the abatement costs for firms, the table shows that the costs would be lower by USD 2.7 billion, in total, or 25.4%. Although this partially reflects the higher level of aggregate emissions in the absence of transaction costs, the average cost of abatement is also lower, by USD 41 per SO_2 tons, or 18.9%. This finding indicates that the "potential" gains from trade, which could be achieved in the absence of transaction costs, are significant.

Implications on Net-benefit I also calculate the net-benefit of a cap-and-trade program. To do so, I calculate health and environmental damages in each case in Table 4 using the data from Muller and Mendelsohn (2009).[36] Health and environmental damages increase by USD 6.3 billion in the absence of transaction costs. This increase reflects both the increase in the aggregate level of SO_2 emissions, as well as the increase in the average health damages.

[35] The baseline is defined as the equilibrium outcome under the estimated model where transaction costs are present.

[36] Muller and Mendelsohn (2009) use the AP2 model, an integrated assessment model, to calculate marginal damages from SO_2 emissions at the county level. I use the marginal damages from point sources with effective height less than 250 meters (denoted as low point sources). Following Muller and Mendelsohn (2009), 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. See, e.g., Fowlie and Muller (2013) and Chan et al. (2015) for other attempts that calculates health and environmental damages of air pollutants using the AP2 model.

In particular, the average health damages increase by 1.5%, indicating that more active trading of emissions permits leads to greater emissions in regions where the health damage is higher.[37] In summary, the total costs (including firms' costs of abatement and health damages) increased by USD 3.6 billion (4.9%) in the absence of transaction costs.[38]

Figure 6: Permit Banking in the Absence of Transaction Costs

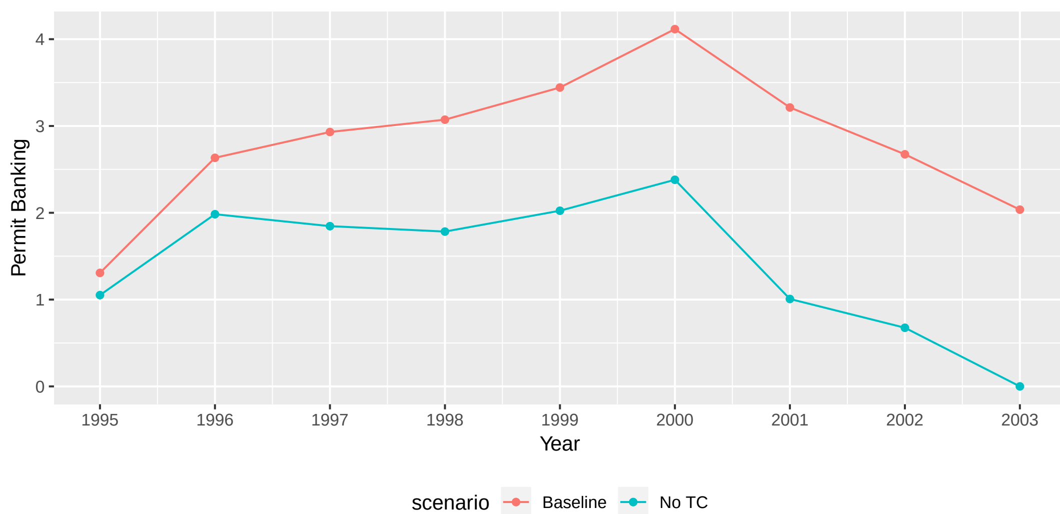
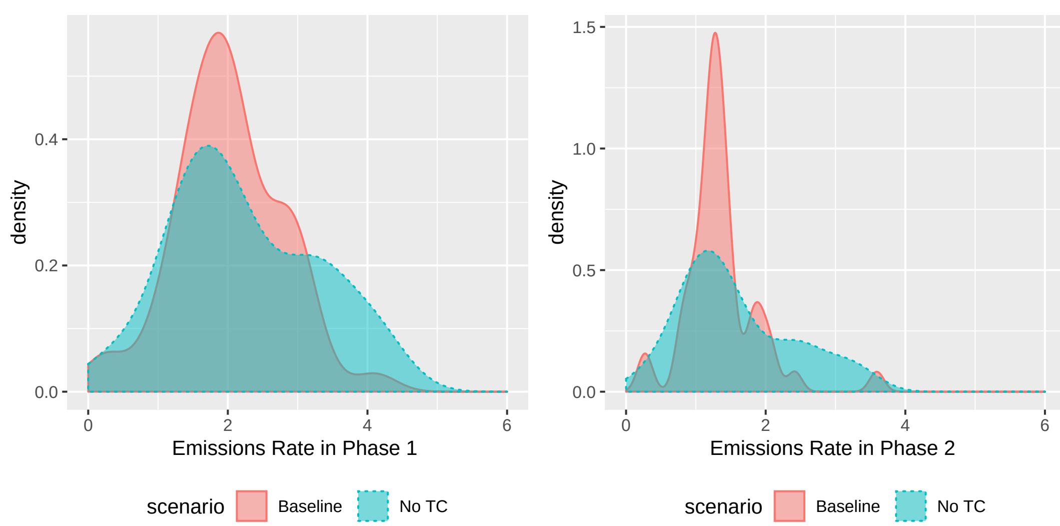


Figure 7: Distribution of Emissions Rate in the Absence of Transaction Costs



Notes: Emissions rate is measured by lbs per MMBtu.

---

[37] SO_2 emissions are known as nonuniformly mixed pollution; health and environmental damages depend on the location of the emissions' source.

[38] Note that the negative net effect of eliminating transaction costs is mostly driven by the fact that the health and environmental damage is much higher than the abatement cost under the regulatory intensity (i.e., emissions cap) in the Acid Rain Program. This fact motivates the further reduction of SO_2 emissions proposed in the subsequent regulations after 2003 (see, e.g., Schmalensee and Stavins, 2013).

Table 4: The Potential Gains from Trade

	Baseline	No Transaction Costs
Emissions (in 1 million tons)	53.20	57.68
Banking at the terminal period (in 1 million tons)	2.04	0.00
Abatement costs (in USD 1 million)	10,663	7,951
Change from baseline (in USD 1 million)		-2,711
Average abatement cost (in USD)	217	178
Health and environmental damages in USD 1 million)	63,297	69,656
Change from baseline (in USD 1 million)		6,359
Average damage (in USD)	1,190	1,208

Notes: The numbers are the total from 1995 to 2003. The unit of emissions and banking at the terminal period is 1 million SO_2 tons. Abatement cost is defined as the sum of investment costs and the profit loss due to the regulation. The latter is the difference between the realized profit and the business as usual (BAU) profit. The BAU is defined as the case in which the shadow value of permits λ_{it} is 0 and firms do not change emissions rate at all. The average abatement cost is the ratio of total abatement costs to the total abatement, which is defined as the difference between the actual and the BAU emissions.

6.2 The Role of the Permit Banking System

A key feature of the Acid Rain Program is the permit banking system and the pre-announced schedule of permit allocation which decreased from Phase II (2000). While the regulated firms utilized permit banking in their compliance strategies, the simulation analysis in the previous section shows an excess banking of permits, which contributes to less trading of permits and lower allocative efficiency. Here, I examine the effects of the permit banking system, given the presence of transaction costs. I simulate an equilibrium outcome in which permit banking between Phases I and II is not allowed. In other words, 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.[39]

Table 5 shows the firm- and year-level simulation outcomes in the baseline and no-banking cases. The trading volume, defined as the absolute value of net purchases, $|b_{it}|$, is higher by 26.8% in the absence of permit banking than in the baseline case, implying that permit trading is more active without banking. When banking is not allowed, firms have a greater incentive to trade. Because firms cannot rely on their banked permits for compliance, they have to buy permits from other firms. Furthermore, sellers have to discard their emissions permits, unless they participate in permit trading to sell.

Though more active trading of emissions permits could lead to an efficient reallocation of emissions permits, permit banking allows firms to smooth the abatement between phases.

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

Given that the initial allocation of permits changes significantly between Phase I and Phase II, the smoothing role is, presumably, important. To see the net effect of permit banking, Table 6 shows the efficiency measures in both cases.[40] The total abatement costs are estimated to be USD 11.3 billion without the permit banking system. I also calculate the average abatement cost, which is 2.8% higher than that in the baseline case. This result indicates that a permit banking system improves the cost efficiency of the cap-and-trade program, on aggregate.

The simulations in Sections 6.1 and 6.2 highlight the interaction of transaction costs and dynamic incentives. In particular, in the presence of transaction costs, a permit banking system might have subtle and counteracting implications for efficiency. The permit banking, on one hand, allows a firm to smooth costs over time, improving its inter-temporal efficiency. However, it discourages the trading of permits, reducing the firm's intra-temporal allocative efficiency.

Given this trade-off, a policy design that can improve the outcome is "depreciating permits," which is akin to the "depreciating license" mechanism of Weyl and Zhang (2018). Under the depreciating permits system, a certain fraction of banked permits is collected by the regulator at the end of the period, and then reallocated 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.

Table 5: Simulation Outcomes at the Firm-and-year Level

	Baseline			No Banking		
	Mean	Std. Dev	Median	Mean	Std. Dev	Median
Emissions e_{it}	76,992.2	100,173.9	44,791.3	74,954.6	98,889.5	44,232.7
Net purchase b_{it}	2,609.0	9,794.4	-158.4	2,846.6	12,454.1	-589.7
Trading volume $\|b_{it}\|$	6,454.5	7,811.9	4,147.2	8,183.4	9,805.9	5,238.5
Left-over permits				3,230.2	24,815.5	0.0

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 on 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 a simulation

[40]The aggregate emissions are different from the baseline case owing to the presence of leftover permits. Around 4% of permits would expire if a permit banking system were not available. Note that emissions permits might expire for two reasons. First, if a firm does not participate in permit trading, it cannot sell permits and, thus, the remaining permits must expire. Even though the firm participates, the marginal revenue from selling permits could be less than zero, owing to transaction costs. In such a case, firms do not sell all of their remaining permits.

Table 6: Effects of Permit Banking

	Baseline	No Banking
Emissions (in 1 million tons)	53.20	51.79
Left-over permits (in 1 million tons)	0.00	2.23
Banking at the terminal period (in 1 million tons)	2.04	1.52
Abatement costs (in USD 1 million)	10,663	11,287
Change from baseline (in USD 1 million)		624
Average abatement cost (in USD)	217	223
Health and environmental damage (in USD 1 million)	63,297	62,057
Change from baseline (in USD 1 million)		-1,240
Average damage (in USD)	1,190	1,198

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. See the notes in Table 4 for the definition of the abatement costs.

analysis, I find that the average costs of abatement could be reduced by 18% in the absence of transaction costs. This additional cost saving is achieved by more active trading of permits and a more efficient allocation of investment. I also examined the role of a permit banking system, finding that it helps firms to smooth their abatement costs across periods, although it could discourage the trading of permits.

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). 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. I leave these topics for future work.

References

Aguirregabiria, Victor and Pedro Mira, "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 2002, *70* (4), 1519–1543.

_ **and** _ , "Sequential estimation of dynamic discrete games," *Econometrica*, 2007, *75* (1), 1–53.

— **and** — , "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 2010, *156* (1), 38–67.

Arimura, Toshi H, "An Empirical Study of the So2 Allowance Market: Effects of PUC Regulations," *Journal of Environmental Economics and Management*, 2002, *44* (2), 271–289.

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

Breen, William J, Laurie Simon Hodrick, and Robert A Korajczyk, "Predicting equity liquidity," *Management Science*, 2002, *48* (4), 470–483.

Cantillon, Estelle and Aurelie Slechten, "Price formation in the European carbon market: the role of firm participation and market structure," *Working Paper*, 2015.

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

Chan, H Ron, B Andrew Chupp, Maureen L Cropper, and Nicholas Z Muller, "The Impact of Trading on the Costs and Benefits of the Acid Rain Program," Technical Report, National Bureau of Economic Research 2015.

Chan, Ron, "How Large are the Cost Savings from Emissions Trading? An Evaluation of the U.S. Acid Rain Program," *Working Paper*, 2015.

Chen, Cuicui, "Slow focus: Belief evolution in the US acid rain program," Technical Report, Working Paper 1–49 2018.

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

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

Collard-Wexler, Allan, "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 2013, *81* (3), 1003–1037.

Dardati, Evangelina, "Pollution Permit Systems and Firm Dynamics: How does the Allocation Scheme Matter?," *Working Paper*, 2014.

Dávila, Eduardo and Cecilia Parlatore, "Trading Cost and Informational Efficiency," 2017.

EIA, "Electric Power Annual 2012," Technical Report 2012.

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

_ , Paul L Joskow, Richard Schmalensee, Juan-Pablo Montero, and Elizabeth M Bailey, *Markets for clean air: The US Acid Rain Program*, Cambridge University Press, 2000.

Ericson, Richard and Ariel Pakes, "Markov-perfect industry dynamics: A framework for empirical work," *The Review of Economic Studies*, 1995, *62* (1), 53–82.

Fabra, Natalia and Mar Reguant, "Pass-Through of Emissions Costs in Electricity Markets," *American Economic Review*, 2014, *104* (9), 2872–99.

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

_ , "Emissions trading, electricity restructing, and investment in pollution abatement," *The American Economic Review*, 2010, pp. 837–869.

_ and Jeffrey M Perloff, "Distributing pollution rights in cap-and-trade programs: are outcomes independent of allocation?," *Review of Economics and Statistics*, 2013, *95* (5), 1640–1652.

_ and Nicholas Muller, "Market-based emissions regulation when damages vary across sources: What are the gains from differentiation?," Technical Report, Working Paper 2013.

_ , Mar Reguant, and Stephen P Ryan, "Market-based emissions regulation and industry dynamics," *forthcoming in Journal of Political Economy*, 2014.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Dynamic trading with predictable returns and transaction costs," *The Journal of Finance*, 2013, *68* (6), 2309–2340.

Gillingham, Kenneth, Fedor Iskhakov, Anders Munk-Nielsen, John Rust, and Bertel Schjerning, "A Dynamic Model of Vehicle Ownership, Type Choice, and Usage," 2015.

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

Helfand, Gloria E, Michael R Moore, and Yimin Liu, "Testing for dynamic efficiency of the Sulfur Dioxide Allowance Market," 2006.

Hotz, V Joseph and Robert A Miller, "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Jaraitė-Kažukauskė, Jūratė and Andrius Kažukauskas, "Do transaction costs influence firm trading behaviour in the european emissions trading system?," *Environmental and Resource Economics*, 2015, *62* (3), 583–613.

Joskow, Paul L and Richard Schmalensee, "The Political Economy of Market-Based Environmental Policy: the US Acid Rain Program 1," *The journal of law and economics*, 1998, *41* (1), 37–84.

_ , _ , and Elizabeth M Bailey, "The market for sulfur dioxide emissions," *American Economic Review*, 1998, pp. 669–685.

Kalouptsidi, Myrto, "Time to build and fluctuations in bulk shipping," *The American Economic Review*, 2014, *104* (2), 564–608.

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

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

Krusell, Per and Anthony A Smith Jr, "Income and wealth heterogeneity in the macroeconomy," *Journal of political Economy*, 1998, *106* (5), 867–896.

Lee, Donghoon and Kenneth I Wolpin, "Intersectoral labor mobility and the growth of the service sector," *Econometrica*, 2006, *74* (1), 1–46.

Lillo, Fabrizio, J Doyne Farmer, and Rosario N Mantegna, "Econophysics: Master curve for price-impact function," *Nature*, 2003, *421* (6919), 129–130.

Liski, Matti and Juan-Pablo Montero, "Market Power in an Exhaustible Resource Market: The Case of Storable Pollution Permits," *The Economic Journal*, 2011, *121* (551), 116–144.

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

Muller, Nicholas Z and Robert Mendelsohn, "Efficient pollution regulation: getting the prices right," *The American Economic Review*, 2009, *99* (5), 1714–1739.

Olley, G.S. and A. Pakes, "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 1996, *64* (6), 1263–1297.

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

Robert, Engle, Ferstenberg Robert, and Russell Jeffrey, "Measuring and modeling execution cost and risk," *The Journal of Portfolio Management*, 2012, *38* (2), 14–28.

Rubin, Jonathan D, "A model of intertemporal emission trading, banking, and borrowing," *Journal of Environmental Economics and Management*, 1996, *31* (3), 269–286.

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

_ , "Structural estimation of Markov decision processes," *Handbook of econometrics*, 1994, *4*, 3081–3143.

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

Schmalensee, Richard and Robert N Stavins, "The SO2 Allowance Trading System: The Ironic History of a Grand Policy Experiment," *The Journal of Economic Perspectives*, 2013, *27* (1), 103–121.

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

Stavins, Robert N, "Transaction costs and tradeable permits," *Journal of environmental economics and management*, 1995, *29* (2), 133–148.

U.S. Environmental Protection Agency, "The National Allowance Data Base Version 2.1 Technical Support Document," 1993.

_ , "Technical Documentation for Phase II Allowance Allocations," Technical Report, EPA-430-R-93-002 1993.

_ , "Acid Rain Program 2003 Progress Report," Technical Report September 2004.

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

Appendix For Online Publication

A Difference-in-differences Analysis on Utilization Rate

To estimate the effects of the ARP on production output, I exploit the variation of the timing of the regulation across units in a Difference-in-differenes (DID) framework. There are two groups of units: those regulated since 1995 (Group I units), and those regulated since 2000 (Group II units). I plot the trend of the capacity factor for each group's units in Figure A.1, which supports the parallel-trend assumption in the DID framework.

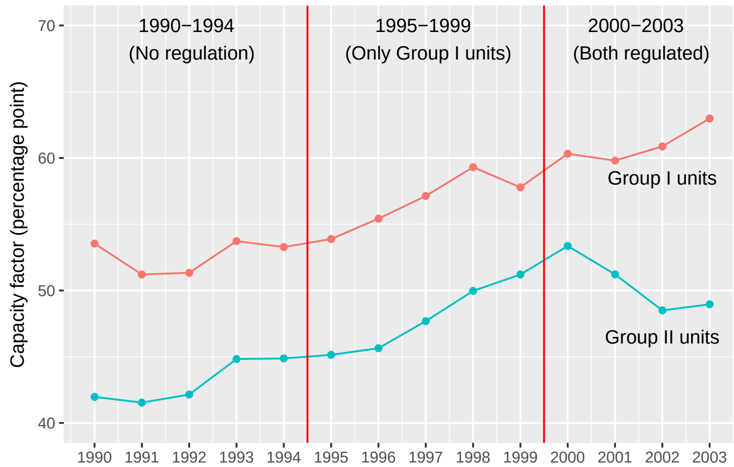
The regression equation is given by

$$cf_{jm} = \alpha_1 \text{GroupI}_j \cdot 1\{\text{after1995}\}_m + \alpha_2 \text{GroupII}_j \cdot 1\{\text{after2000}\}_m + \gamma X_{jm} + u_j + u_m + u_{jm},$$

where cf_{jm} is the capacity factor of unit j in month-year m. The capacity factor is defined by $cf_{jm} = q_{jm}/k_j$, where q_j is the net generation and k_j is the nameplate capacity. GroupI_j and GroupII_j are dummy variables for each group. $1\{\text{after1995}\}_m$ and $1\{\text{after2000}\}_m$ are dummy variables indicating the periods after 1995 (the beginning of Phase I) and 2000 (the beginning of Phase II), respectively. X_{jm} includes control variables such as fuel costs. Unit and time fixed effects are captured by u_j and u_m, respectively.

The regression results are shown in Table A.1. I find that introducing the ARP decreases the capacity factor by 1 to 2.5 percentage points, which is statistically significant. This finding is consistent with the idea that introducing a cap-and-trade program increases the marginal costs of production, because firms are facing opportunity costs of emissions under such a program. Thus, the increase in marginal costs decreases the output of generating units under a cap-and-trade regulation. Although the effects are statistically significant, the economic significance of the effects seems to be limited. Because the mean of the capacity factor is within the range of 40–60 percentage points in the sample, electricity generation decreased by around 2%–6% after the introduction of the cap-and-trade program. This magnitude is not as great as the decrease in emissions over time, as shown in Section 2.3.1. Combined with the findings from Section 2.3.2, this regression analysis indicates that the abatement of SO_2 emissions was achieved primarily through the adjustment of emissions rates.

Figure A.1: 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 two groups: those regulated since 1995 (Group I units), and those regulated since 2000 (Group II units). The figure shows that these two groups have a similar trend in their capacity factor from 1990 to 1994, supporting the parallel-trend assumption in the DID framework.

Table A.1: Difference-in-differences Regression of Capacity Factor

	Dependent variable:			
	Capacity factor in pct-point (0 to 100)			
	(1)	(2)	(3)	(4)
Treatment (Group I)	−0.506	−1.920***	−2.000***	−2.900***
	(0.571)	(0.657)	(0.569)	(0.691)
Treatment (Group II)	−3.802***	−2.456***	−2.636***	−1.865***
	(0.542)	(0.580)	(0.565)	(0.613)
log(fuel costs)			−10.004***	−9.959***
			(0.458)	(0.459)
log(electricity demand)	38.703***	38.742***	41.937***	41.937***
	(1.052)	(1.053)	(1.162)	(1.164)
Group-trend	No	YES	No	YES
Observations	373,934	373,934	278,956	278,956
Adjusted R^2	0.674	0.674	0.612	0.612

Notes: Unit-level dummies, year dummies, and month-of-year dummies are included. Standard errors are clustered at the unit level. $*p < 0.1; **p < 0.05; ***p < 0.01$

B Hedonic Regressions of Coal Price

This appendix explains the hedonic regressions of coal prices. I use the estimation results to obtain the fuel costs c_{jt} in the profit function as a function of the SO_2 emissions rate R_{jt}. I use the data from the Form FERC No. 423 (EIA-423) "Monthly Report of Cost and Quality of Fuels for Electric Plants" for the estimation. This data reports plant- and month-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 $p^{fuel}_{k,l,m}$ in fuel delivery k for plant l in month-year m:

$$\begin{aligned} p^{fuel}_{k,l,m} &= \exp(\phi_m + \phi_t + \phi_r + u_{l,k,m})\,(R_{k,l,m})^{\phi} \\ \iff \log(p^{fuel}_{k,l,m}) &= \phi \log(R_{k,l,m}) + \phi_m + \phi_t + \phi_r + u_{l,k,m}, \end{aligned}$$

where $p^{fuel}_{k,l,m}$ is the coal price, measured in cents per MMBtu, and $R_{k,l,m}$ is the SO_2 emissions rate, measured in lbs per MMBtu. ϕ_m, ϕ_t, and ϕ_r are the fixed effects for a month, year, and region, respectively.[41]

Note that I only include the SO_2 emissions rate $R_{k,l,m}$, along with time and region dummies, as covariates. Other product characteristics, such as ash content and the distance between plants and coal mines, are available in the data set and are used by other studies in regressions (See, e.g., Chan, 2015). However, the primary purpose of this regression is to predict how the choice of an SO_2 emissions rate affects the coal price. Because the other product characteristics could change with the choice of the SO_2 emissions rate, it would be erroneous to use a regression equation that includes other characteristics to predict how the coal price would change with respect to the SO_2 rate, *holding other characteristics fixed.* Therefore, I use the function with respect to the emissions rate $R_{l,k,m}$ to predict the coal price.

Table B.2 reports the results of coal price regression.

[41] I use the definition of US regions provided by the US Census Bureau. 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).

Table B.2: Hednic Regression

	Dependent variable:
	log(fuel price)
log(emissions rate)	-0.049^{***}
	(0.004)
Observations	304,470
Adjusted R^2	0.205

Notes: Estimates of month, year, and region dummies are omitted. Robust standard errors are used. *p<0.1; **p<0.05; ***p<0.01.

C Details on Model Derivations

This appendix explains the derivation of the ex ante value function and its derivative, $\partial EV_t(h_t, I_t)/\partial h_t$. Remember that the ex ante value functions are given by

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

Note that under the assumption that ϵ follows an i.i.d. type-I extreme value distribution with standard deviation σ_F, which I impose in the estimation, the expected value function when $I_{it} = 0$ can be written as

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

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:

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

$$\frac{dEV_t(h_{it}, 1, R_{it})}{dh_{it}} = \lambda_{it}^1, \tag{C.2}$$

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.

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

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

By the chain rule, I have

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

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 have the following (I omit h_t for ease of exposition in the following derivation):

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

where Υ^1 is the set of ϵ such that a firm chooses to participate (i.e., $\Upsilon^1 \equiv \{\epsilon : V_t^1 - F_t - \epsilon > V_t^0\}$), 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 Incentives in Abatement Investment

Here, I discuss how the incentive to invest in abatement is determined in the proposed model. Using the envelope theorem, the marginal return from reducing the emissions rate R^1 is given as follows:

$$
\begin{aligned}
-\frac{\partial EV_{1995}}{\partial R^1} = & \sum_{t=1995}^{1999} \beta^{t-1995} \left(\lambda_{it} \cdot \sum_j HR_{jt} q_{jt}^* \right) \\
& + \sum_{t=1995}^{1999} \beta^{t-1995} \left(\sum_j \frac{\partial c_{jt}}{\partial R^1} q_{jt}^* \right) \\
& + \beta^{2000-1995} \frac{\partial}{\partial R^1} \Gamma(R^1 - R^2).
\end{aligned}
$$

The first component is the returns from reducing emissions evaluated at the shadow value λ_{it}. The second component is the additional costs of using a cleaner fuel. Note that $\frac{\partial c_{jt}}{\partial R^1} < 0$, because fuel costs are higher for low-sulfur coals. The last component is the saving of investment costs in Phase II by investment in Phase I.

The primary component in the return on investment is the first term. By reducing the emissions rate, the firm can marginally reduce emissions by $\sum_j HR_{jt} q_{jt}^*$. 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.

E Computational Details on 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 generation q_{jt}, which appears in the profit function π_{it}, given by equation (3.1), and the firm-level emissions, $e_{it} = \sum_j \rho_{jt} q_{jt}$. Because each firm has multiple generation units, solving unit-level production in a dynamic framework seems computationally demanding. 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 generation $\{q_{jt}\}_{j \in J_{it}}$, *holding firm-level emissions* e_{it} *fixed:*

$$
\begin{aligned}
\Pi_{it}(e_{it}, \{\rho_{jt}\}_{j \in J_{it}}) \equiv \quad & \max_{\{q_{jt}\}_{j \in J_{it}}} \quad \pi_{it}\left(\{q_{jt}\}_j\right) \\
& \text{s.t.} \qquad \sum_{j \in J_{it}} \rho_{jt} q_{jt} = e_{it}.
\end{aligned}
$$

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

$$\frac{\tau_{st}^{elec} - c_{jt}^{fuel} - g'(q_{jt})}{\rho_{jt}} = \lambda_{it}^{sub} \forall j$$
$$\sum_{j\in J_{it}} \rho_{jt} q_{jt} = e_{it},$$

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

I now use $\Pi_{it}(e_{it}, \{\rho_{jt}\}_{j\in J_{it}})$ to consider the dynamic decision problem:

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

Note that the choice variables are e_{it}, b_{it}, and $h_{i,t+1}$, which is fewer than in the original problem.

When I numerically solve the individual dynamic decision problem, I use two steps. First, I construct $\Pi_{it}(e_{it}, \{\rho_{jt}\}_{j\in J_{it}})$ using the unit-level FOC for production. I then use the pre-computed $\Pi_{it}(e_{it}, \{\rho_{jt}\}_{j\in J_{it}})$ 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 by backward induction.

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

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

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

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

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 $W_{2000}(h_{2000}, I_{2000}, R^1)$.

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

$$\max_{R^1} EV_{1995}(0, 0, R_{P1}) - \Gamma(R^1, R^0).$$
$$\text{s.t.} R^1 \leq R^0$$

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 5.

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 $\hat{x}_{it}(h_{it}, I_{it}, R_{it})$ for emissions e_t, net purchase b_t, and banking h_{t+1}, participation probability $P_{it}(h_{it}, R_{it})$, and the investment decisions $R_i^1(h_{i1995}, I_{i1995})$ and , $R_i^2(h_{i,2000}, I_{i,2000}, R_i^1)$.

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

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

7. Check the market-clearing condition as

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

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 so 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 equation

$$\sum_i \sum_s P_{i,enter}(s) \cdot TC'^{(-1)} \left(\hat{\lambda}_{it}(\mathbf{P}^l, s) - P_t^{l+1}\right) + \bar{B}_t^{fringe}(P_t^{l+1}) = 0,$$

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.

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

F Details in Counterfactual Simulations

F.1 Shutting Down 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, is determined by $\partial \pi_{it} / \partial q_{jt} = P_t \ \forall j$. As discussed in Section 3.7, 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

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

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

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

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

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

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

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, we 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 $E_t = \sum_i e_{it}(P_t)$, 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), we 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 omit the subscript i for simplicity. The problem is given by

$$\begin{aligned} V_{1999}^1(h_{1999}, I_t = 1, R_t) = \quad & \max_{\{q_{jt}\}_j, b_t} \quad \pi_t\left(\{q_{jt}\}_j\right) - \left(P_t b_t + TC(b_t)\right) + \beta W_{2000}(0, I_{2000}, R^1) \\ & \text{s.t.} \quad e_t\left(\{q_{jt}, \rho_{jt}\}_j\right) = a_t + h_t + b_t. \end{aligned}$$

Note that permit banking h_{2000} is not part of the choice variables, and the continuation value $W_{2000}(0, I_{2000}, R^1)$ 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 nontrader:

$$\begin{aligned} V_{1999}^0(h_{1999}, I_t = 0, R_t) = \quad & \max_{\{q_{jt}\}_j, b_t} \quad \pi_t\left(\{q_{jt}\}_j\right) + \beta W_{2000}(0, I_{2000}, R^1) \\ & \text{s.t.} \quad e_t\left(\{q_{jt}, \rho_{jt}\}_j\right) \le a_t. \end{aligned}$$

In this case, a firm may not consume all its permits owing to 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).