Estimation of Production Technologies with Output and Environmental Constraints

Scott E. Atkinson[*] Rong Luo[†]

University of Georgia *Renmin University of China*

January 25, 2022

Abstract

Plants in many industries minimize costs subject to output and emission constraints, which restrict input choices. We develop a novel cost function approach to estimating their production function parameters, allowing heterogeneity in generation productivity and pollution abatement efficiency. We apply this technique to a panel of 76 US coal-fired power plants, which choose coal characteristics to minimize the total costs of coal and pollution control. We find substantial heterogeneity in both efficiency terms and a dramatic growth in abatement efficiency over time, consistent with a major goal of the Acid Rain Program's cap-and-trade system. Counterfactual analysis compares the cost efficiencies for different permit allocation methods, with highly restricted permit trading. With no trading, allocation based on generation can reduce total variable costs by \$.9 billion and abatement costs by up to 82%, compared with allocations based on the other methods. Thus, generation-based allocation is the safest hedge against a non-functioning permit system.

KEYWORDS: Output and Emission Constraints, Cost Minimization, Transaction Costs, Emission Permit Allocation, Generation Productivity, Abatement Efficiency.

JEL CODES: D24, L11, L94

[*]Email: `atknsn@uga.edu`.

[†]Email: `luorong@ruc.edu.cn`.

The authors thank participants at the 2018 International Industrial Organization Conference, the 2018 North American Chinese Economics Society, the 2018 Asian Econometric Society Meetings, and the 2018 Chinese Econometric Society Meetings for helpful comments. We especially thank Chad Syverson, Mark Roberts, Paul Grieco, David Rivers, John Turner, Tom Quan, Jon Williams, and Mike Abito. Rong Luo also thanks the Bonbright Center, Department of Economics, University of Georgia for funding support.

1 Introduction

Plants in many industries, especially electricity generation, face constraints on the quantities of outputs (goods) they must produce and the quantities of pollutants (bads) they can produce. The power plants cannot flexibly choose output to maximize profits. Regulatory agencies typically require that electric utilities, through orders given to their power plants, satisfy all demand at regulated prices. Even in restructured jurisdictions, plants take output orders from parent utilities.

Regulatory agencies also restrict plant emissions of co-generated bads. Under the US sulfur dioxide (SO_2) cap-and-trade system, also called the Acid Rain Program (ARP), coal-fired power plants must hold enough tradable permits to cover their emissions, where the total allowable SO_2 from all plants is scaled down from pre-ARP levels. Plants can reduce emissions by switching to cleaner fuels or operating flue-gas-desulfurization (FGD) devices (also called scrubbers). Thus, restrictions on the generation of bads have a significant and complex impact on input choices, due to the trade-off between the added expenses of cleaner inputs and pollution abatement.

Given the output and emission constraints, coal-fired power plants operate as cost minimizers. To model these plants, we develop a structural cost function approach to estimating their production technologies. Plants choose the levels of sulfur and the British Thermal Unit (Btu) or heat content of coal to minimize the sum of coal and pollution control costs, subject to output targets and emission constraints. This choice implies trade-offs between production and abatement costs. A higher sulfur content reduces the unit cost of coal but increases the pollution control cost, while a higher Btu content increases the unit cost of coal but reduces the coal quantity and pollution control costs.

For plants that generate more SO_2 than allowed by the initial permit allocation, we consider three options. First, they can abate emissions if they have FGD devices. Second, they can switch to lower-sulfur coal, which has higher unit costs. Lastly, they can purchase permits from other plants to cover emissions. For FGD and non-FGD plants, we solve for their optimal input choices and derive their total cost functions for the generation of the good and control of the bad.

However, input choices are endogenous, since they are correlated with generation productivity and abatement efficiency, which are observed by the plants but not in the data. Plants with higher generation productivities produce more electricity for the same amount of inputs, and plants with higher abatement efficiencies incur lower abatement costs to scrub a given amount of SO_2. The major innovation of this paper is that we mitigate this endogeneity by including generation productivity and pollutant abatement efficiency terms in a cost minimization framework.

We apply this model to a balanced panel of the 76 largest US coal-fired power plants during the 1995-2005 period, where substantial variation exists in production productivity and abatement efficiency. Data show that the more productive plants generate up to twice as much output as others with the same Btu inputs, especially among the smaller plants. For FGD plants using scrubbers, the operating and maintenance (O&M) cost to abate one ton of sulfur ranges from \$29 to \$533 across plants, with a standard deviation of \$145. These extreme variations heighten our concerns about the endogeneity of input choices and justify the inclusion of the productivity and efficiency terms. The sample period includes the years when the ARP operated with negligible transaction costs, allowing us to determine the growth in output productivity and abatement efficiency with a well-functioning cap-and-trade system.

To control for this endogeneity, the literature estimating production functions for other industries (e.g., Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), Wooldridge (2009), Ackerberg, Caves, and Frazer (2015) (ACF), and Doraszelski and Jaumandreu (2013)) incorporates an output productivity term and uses a control function approach to obtain consistent production function parameter estimates in a profit-maximization framework. Gandhi, Navarro, and Rivers (2020) shows that an under-identification issue exists in the previous approach, but a known flexible input elasticity can solve this issue. Our model extends this literature by including both output productivity and abatement efficiency in production and abatement equations, which are reformulated in the framework of constrained cost minimization. We follow this literature by assuming that both processes are Markov. The total cost functions depend on the parameters of the Leontief production function and the abatement cost function. We invert the production function to control for the unobserved productivity and substitute it into the cost functions. Thus, our cost function approach to estimating the production technologies avoids the need to assume monotonicity between productivity and investment or material inputs.[1]

Our estimation strategy consists of three steps. First, we estimate the hedonic price of coal as a function of its sulfur and Btu content using coal shipment data. We combine the estimated price of sulfur in coal and the optimality condition of sulfur content choice to calculate the implied permit prices by plant and year. Second, we estimate the FGD plants' abatement cost function for SO_2 removal using annual data on O&M costs and the abated amount of sulfur. Lastly, we estimate the derived total variable cost function, which depends on the production function parameters, excluding separable abatement costs. This step uses data on output, inputs of labor and capital, estimated permit prices, and sulfur and Btu content. In the second and third steps, we use the generalized methods of moments (GMM) with instrumental variables (IV) to deal with the endogeneity in sulfur and Btu content.

Two of our empirical results are consistent with known plant behavior, while two others shed light on the less-well-understood trends in plant productivity and efficiency. First, FGD plants have increasing marginal abatement costs, which implies that a plant would abate SO_2 until the marginal cost equals the permit price. This is consistent with

[1]In OP, LP, and ACF, if the firms' investments or intermediate inputs are monotonic functions of productivity, one can invert these functions and replace the unobserved productivity in the production function to deal with endogeneity in inputs.

plants abating some emissions and covering others with permits. Second, plants exhibit minor increasing returns to scale, which is consistent with the the literature, as summarized in Atkinson (2019). Third, the estimated unobserved generation productivities and abatement efficiencies vary substantially across plants. Lastly, while generation efficiency remained stable over time, abatement efficiency improved substantially during the sample period. In particular, the plants' FGD O&M costs to remove a unit of SO_2 decreased by 37% from 1995 to 2005 on average. This result is consistent with the stated goal of the ARP, which is to provide sources with the flexibility to select the most cost-effective approach to reducing emissions. Our methodology, which estimates separate output productivity and efficiency abatement terms, can be directly applied in future analysis of the growth in the efficiency of abating SO_2 and carbon dioxide (CO_2) emissions by the hundreds of newly-constructed and planned coal-fired power plants worldwide under cap-and-trade systems.

Our counterfactual relates to the literature that studies the existence of trade impediments for cap-and-trade systems. The Coase (1960) theorem states that, without trade impediments, the market equilibrium in a cap-and-trade system will be independent of the initial allocation of permits. An extensive literature calls this the independence property, as summarized in Fowlie and Perloff (2013). However, they find mixed evidence regarding the existence of this property.

Developing rapidly from a series of regulations which began in 2004, by 2009 the ARP spot market failed to operate effectively due to extremely high transaction costs. A logical extension of the Coase (1960) theorem is that, under high transaction costs, the initial permit allocation will significantly affect plants' costs. To model this, our first counterfactual assumes that the spot trading system is non-functioning and evaluates the cost implications of initial permit allocation schemes based on the initial emissions, total generation, and generation efficiency. To perform the counterfactual, we use our cost function estimates, which are based on data from a period of negligible transaction costs.

We obtain four results which are consistent with the Coase Theorem, but of magnitudes greater than expected. First, shutting down the permit trading system would increase the plants' variable costs, by as much as \$1 billion in 2005 dollars during the sample period. Second, each permit allocation method produces significant differences in the demand for sulfur, Btu content, and the abatement costs of FGD plants. Third, if allocations are based on generation, the total costs of coal and abatement for FGD plants are lowest, while if permits are based on emissions, total costs for non-FGD plants are lowest. Lastly, allocation based on generation reduces total costs by \$0.9 billion and abatement costs by 32%-82% compared with other allocation methods. We are unaware of any other study that estimates the impact of different initial allocation systems on plants' costs under substantial trade impediments.

A second counterfactual considers intra-state trading under the current ARP system. Allowing the same range of sulfur content as with the first counterfactual, the permit markets do not clear in some states.

The rest of the paper contains eight parts. Section 2 provides additional background on the regulatory enviroment and related cost-function literature. Section 3 examines our data sources, while Section 4 formulates the cost-minimization problems. We develop the estimation methods and identification strategies in Section 5, with results in Section 6. Finally, counterfactuals follow in Section 7. Section 8 concludes the paper.

2 Regulatory Environment and Related Literature

The major bad output of the generation process is the pollutant SO_2, which is produced by the combustion of sulfur in coal. Under the ARP cap-and-trade program for SO_2, plants with FGD devices must determine the optimal trade-off between abating emissions and holding permits to cover any unabated emissions. Plants are allocated enough permits to cover a grandfathered level of emissions. If plants generate SO_2 emissions above these levels, they must cover them by purchasing permits or remove them from the stack gas using FGD devices. Plants can sell any unused permits in a well-functioning permit market. The FGD devices employ either wet scrubbing or dry injection methods. A wet-scrubbing FGD device, by far the most common, typically consists of a cylindrical container into which a slurry of sorbents is sprayed over the generated SO_2, which has also been forced into the container. The sorbents combine with the sulfur from the SO_2, and the resulting by-product is solidified and removed from the plant.

Established under Title IV of the 1990 Clean Air Act (CAA) Amendments, the ARP required major emission reductions of SO_2 from coal-fired power plants by setting a permanent cap on total emissions but allowing the trade of emission permits. This program included plants in Phase I (the largest, dirtiest plants from 1995-1999) and Phase II (smaller, cleaner plants from 2000 and beyond). The initial allocation of permits was principally based on emission rates (emissions per unit of output). Among the plants in our sample, 91% of parent firms face rate-of-return regulation where regulatory commissions set output prices, and consumers determine the output quantity. For the other 9%, the auctions of Regional Transmission Organizations (RTOs) determine output prices.[2] Regardless, all plants face production decisions made by parent firms, so that they cost minimize subject to constraints on output and emissions.

Previous to 2004, the ARP operated with very low transaction costs. However, a series of restrictions by the Environmental Protection Agency (EPA) and the courts imposed high transaction costs on the ARP from 2004 onward. These restrictions included the Clean Air Interstate Rule (CAIR) in 2004, state- and source-level constraints on emissions, and the Cross-State Air Pollution Rule (CSAPR) in 2010.[3] By 2009, the

[2]RTOs and closely related Independent System Operators are very similar in nature. The former typically cover larger geographic areas. Both operate under the approval of the Federal Energy Regulatory Commission.

[3]The CAIR in 2004 cut by two-thirds the number of tradable permits for each source. From 2005

spot permit trading market failed to operate effectively, with permit trades and prices dropping to nearly zero.

Since the efficiency of a cap-and-trade system relies on minimal impediments to trade, a substantial literature has investigated the extent to which such impediments under these systems have violated the independence property. Theory papers examine the effect on permit trading of impediments such as firms' market power (Hahn (1984)), transaction costs (Stavins (1995)), transaction costs and uncertainty (Montero (1998)), and trade restrictions due to government regulations (Hahn and Stavins (2011)).

Several empirical studies assume that if the initial allocation of permits is close to the final one, impediments to trade must have caused a violation of the independence property.[4] However, Reguant and Ellerman (2008), Fowlie and Perloff (2013), and Hahn and Stavins (2011) fail to find significant evidence against the independence property.[5] As an alternative, our counterfactual examines a market where transaction costs are extremely high, namely the ARP after 2009, and estimates the effects on coal composition and costs of abatement and generation for different initial permit allocations.

A number of papers have estimated cost functions for coal-fired power plants, examining different aspects of SO_2 control. Some papers estimate a translog cost function including SO_2 control as an output before computing productivity growth. Examples

onward, the spot permit price dropped dramatically as the prices of coal and natural gas fell and EPA announced that it would reexamine CAIR. Athough the courts vacated CAIR in 2007, EPA mandated state- and source-level constraints on emissions, which made emission permits less useful. Further, EPA announced restrictions on interstate trades with the CSAPR in 2010, which substituted for CAIR. This substitute rule, which is still in effect today, established state-specific emissions caps for power plants, allowing only intrastate trading and limited interstate trading. See Schmalensee and Stavins (2013) for more details.

[4]Ellerman, Joskow, Schmalensee, Bailey, and Montero (2000) find trading within but not between firms under the SO_2 Acid Rain program. Gangadharan (2000) determines that transaction costs substantially reduce trades in the RECLAIM market. Montero, Sanchez, and Katz (2002) find limited trades under the total suspended particulates trading program in Santiago, Chile. Hanemann (2009) documents limited trades of SO_2 permits under the Acid Rain Program.

[5]Counter to these results is the study by Fowlie (2010) which finds that the rate-of-return regulation of electric utilities in the NO_x cap-and-trade system causes over-capitalization which, as an impediment to trade, may not be consistent with the independence property.

are Baltagi and Griffin (1988) and Atkinson and Dorfman (2005). However, neither of these papers models the sulfur/Btu trade-off or the operating costs of scrubbers. Further, only the latter deals with endogeneity of inputs. Neither recognizes that coal prices (which are arguments of their cost functions) are endogenous, since they depend upon Btu and sulfur content, which are choice variables for the firm.

Other papers explicitly modeled the sulfur/Btu trade-off using cost functions for coal-fired power plants, but did not model the endogeneity of hedonic coal prices or incorporate generation productivity and abatement efficiency to mitigate endogeneity. Examples are Gollop and Roberts (1983, 1985) and Carlson, Burtraw, Cropper, and Palmer (2000), where the latter study assumes that the only endogenous explanatory variable is emissions.[6]

Kolstad and Turnovsky (1998) is the only cost function study that modeled the Btu/sufur trade-off and recognized the endogeneity of coal prices. They estimated a cost system for Eastern coal-fired power plants in their first year of operation from the pre-ARP period, 1976-85. While they utilized utility and state characteristics as instruments, they did not include productivity/efficiency terms to mitigate endogeneity and they did not estimate abatement costs. They found no significant technical progress, measured as the reduction in total cost over time.

As in their paper, we model Btu and sulfur content as choice variables and treat coal prices as endogenous. However, we update and extend their approach to model abatement costs and control more extensively for endogeneity. We derive and estimate separate cost functions for generation and abatement at the plant level using panel data during the period when the ARP permit market was fully functional. Since the sulfur/Btu trade-off is correlated with productivity/efficiency components of the error terms, we mitigate this endogeneity by including terms to measure these effects.

Recently, several hundred coal-fired power plants have been constructed world-wide,

[6]This is estimated as a function of output, input prices, and the emission standard. The negative of the derivative of total costs with respect to emissions is the marginal abatement cost function.

many with SO_2 and CO_2 cap-and-trade systems. Our new methodology can be directly employed to examine the growth in output productivity and abatement efficiency, as well as the impact of initial permit allocations if permit market fail.

3 Data

We employ a balanced panel of the 76 largest coal-fired power plants (without entry or exit to this set) in the US from 1995 to 2005.[7] The sample ends in 2005 because an increasing number of utilities redacted capital and labor data after this date. While the majority of our sample plants are located in the Southern, Mid-Atlantic, or Midwestern states, a few reside in the Rocky Mountain and Far-Western regions.

Coal-fired power plants use capital, labor, and Btu to produce electricity. On average, 99% of the Btu for our plants come from consuming coal.[8] The combusted fuel super-heats water in a boiler until it produces steam, which is then forced under high pressure through turbines, causing magnets to spin inside coils of wire, generating electricity. Each plant produces electricity with one or more of these boiler-turbine-generator (BTG) complexes.

We measure plant-level capital in terms of plant megawatt (MW) capacity, which is the potential for generating electricity in all BTGs, if operated as designed. The MW capacity is baked into each BTG. We adjust capital as the plant augments or reduces existing capacity. Labor is the number of full-time employees plus one-half the number of part-time employees.

We obtained most of our data from several government sources. The Federal Energy Regulatory Commission (FERC) Form 1 provides labor and capital data for private electric power plants, and the Energy Information Administration (EIA) EIA-412 survey is the source of this data for public power plants. While the US Department of

[7]See Appendix A for our plants and parent utilities.

[8]These plants only use very small amounts of oil or natural gas for generation.

Energy (DOE) halted the EIA-412 survey after 2003, the Tennessee Valley Authority voluntarily posted 2004-06 data for its electric power plants online. The DOE Form EIA-767 is the source of information about fuel consumption and net MW hour (MWh) generation by plant. We use data on SO_2 emissions at the plant level collected by the EPA as part of its Continuous Emissions Monitoring System. The Btu and sulfur content data comes from EIA-423, which also supplies data on the transaction-level price of coal delivered to each plant.[9]

Although the prices of capital and labor inputs are only available at the utility level, we make the reasonable assumption that plant-level prices are identical to firm-level prices for these inputs. We compute the user cost of capital at the firm level using the corporate tax rate, the corporate property tax rate, the depreciation rate, the firm's yield on capital, and the Handy-Whitman Index as in Atkinson, Primont, and Tsionas (2018).[10] From FERC Form 1, we construct the wage as salaries plus wages for electric operating and maintenance workers divided by the quantity of labor.

We also collected variables that measure pollution control costs from EIA Forms 767 and 860. These include the SO_2 removal rate of scrubbers and the O&M costs of FGD devices. These costs include the cost of FGD labor, feed materials and chemicals, waste disposal, and other related costs.

Table 1 shows the summary statistics of the plant-level annual data. Among the 76 plants, only 16 plants employed FGD units throughout the 1995-2005 period. The other 60 plants have either never installed FGD units or installed them for only part of this period.[11] Panel A shows the outputs and inputs of the two types of plants. The median annual generation of an FGD plant is 8.20 million MWh, while that of a non-FGD plant is 4.21 million MWh. The median annual SO_2 emission of an FGD plant is

[9]We wish to thank Carl Pasurka for supplying us with data on input and output quantities.

[10]The yield on the firm's latest issue of long-term debt comes from Moody's Public Utility Manual before 2001 and from Mergent's Public Utility Manual after that time.

[11]If a plant did not have FGD devices in all years in the data, we do not include it when estimating the abatement cost function.

about 22,271 tons, while that of a non-FGD plant is 27,074 tons. In terms of inputs, the FGD plants have greater generation capacities, more employees, and higher coal consumption. They use coal with a significantly higher sulfur content and a slightly lower Btu content than non-FGD plants. The median sulfur content is 1.63% for FGD plants, but only 0.87% for non-FGD plants, while the median Btu content is 22.64 million/ton for FGD plants and 24.27 million/ton for non-FGD plants.

Table 1: Data Summary Statistics for Plants

	FGD			Non-FGD		
	median	min	max	median	min	max
A. Outputs and inputs						
Generation (10^6 MWh)	8.20	2.09	20.32	4.21	0.12	22.33
SO_2 emission (tons)	22,271	3,242	113,073	27,074	631	186,470
Generation capacity (MW)	1,620	411	2,600	740	110	3,499
Labor (# employees)	219	68.90	454.64	125	23.74	578
Coal (10^6 tons)	3.98	0.86	9.17	1.73	0.35	11.29
Sulfur content (%)	1.63	0.35	4.12	0.87	0.19	2.94
Btu content (10^6/ton)	22.64	15.58	24.91	24.27	16.56	26.25
B. Input prices						
Yield (%)	7.58	5.38	8.32	7.55	5.38	8.95
Wage (10^4\$)	4.44	2.67	8.34	4.36	2.49	9.47
Coal price (\$/ton)	24.28	7.32	74.10	32.98	11.09	65.83
C. Abatement						
Coal abatement (10^6 tons)	2.79	0.42	8.59			
SO_2 removal rate (%)	87.66	57.00	95.00			
FGD O&M costs (10^3\$)	4,864	300	95,656			
O&M costs/(O&M + coal costs) (%)	4.77	0.25	31.07			
Observations	176	176	176	660	660	660

Panel B of Table 1 shows the input prices at the plant-year level. The two types of plants have similar yield of capital and wage rates. The coal prices are lower for FGD plants because of higher sulfur and lower Btu content. Median coal prices are \$24.28/ton for FGD plants and \$32.98/ton for non-FGD plants.

Panel C shows the data on SO_2 abatement of the FGD plants. The median amount of coal abatement per year is 2.79 million tons, which is less than the coal input, implies that the FGD plants do not abate all coal. The sulfur removal rate is the percentage of

SO_2 that is abated, measured as "percent removal of SO_2 at 100% generation load", as stated in the variable definitions on the Directory of Form 767 files of the DOE EIA.[12] Thus, the removal rate is baked into the FGD device when it is manufactured, and the plants cannot change it once installed. Among the 16 FGD plants, the removal rates did not change during the 11 years for 11 FGD plants. For the other five plants, the standard deviations are extremely small.[13] This removal rate varies substantially across plants, with a median of 87.66%. The median annual O&M cost is $4.86 million per plant, while the average share of O&M cost in the plants' total costs of coal and FGD O&M is 4.77%, with a maximum of 31.07%. Thus, the O&M costs are an important part of the plants' variable costs.

We calculate trends of inputs and generation over time for our plants. From 1995 to 2005, the plant-level heat input from coal has increased slightly, with a mean annual growth rate of 3.88%. The average sulfur and Btu content of all 76 coal-fired power plants has decreased. Generation capital has been slowly increasing for all plants, with an average annual growth rate of 3.93%. The quantity of labor used in generation has been decreasing for all plants, with an annual growth rate of −0.94%. The plant-level electricity generation growth rate is close to the heat growth rate. The average electricity generation among the plants in million MWh increased from 5.71 in 1995 to 6.65 in 2005.

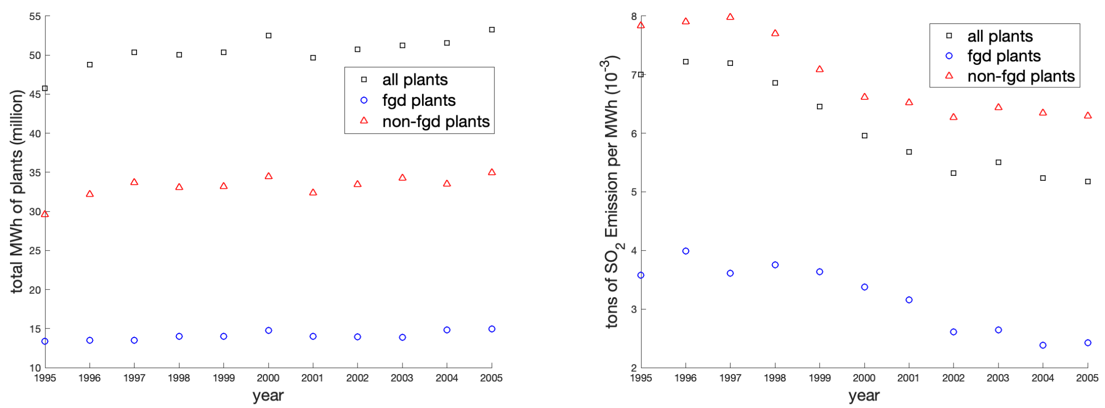
Figure 1a plots the aggregate electricity generation of the plants by year. We represent the aggregate generation for all plants with squares, for non-FGD plants with triangles, and for FGD plants with circles. Due to a larger number of non-FGD plants, their total electricity generation is about twice that of their FGD counterparts. Figure 1b plots the average SO_2 emission per MWh generation of the plants. We represent the

[12]See the F767_DATA_USERS_GUIDE.xls variable definition file within the Form 767 file folders for each year, available from the Energy Information Administration of the Department of Energy. We do not include the year subscript t because r_{jt} changed almost imperceptibly during the sample period. See Table C.1 in Appendix section C.

[13]See Table C.1 in the Appendix for more details.

average for all plants with squares, for non-FGD plants with triangles, and for FGD plants with circles. The non-FGD plants emit about 40-80% more SO_2 per MWh of electricity than the FGD plants. We see that SO_2 emissions in thousands of short tons per MWh have fallen substantially for both types of plants. The decline is almost 50% for FGD plants and 25% for non-FGD plants.

Figure 1: Aggregate Generation and SO_2 Emission per MWh Generation

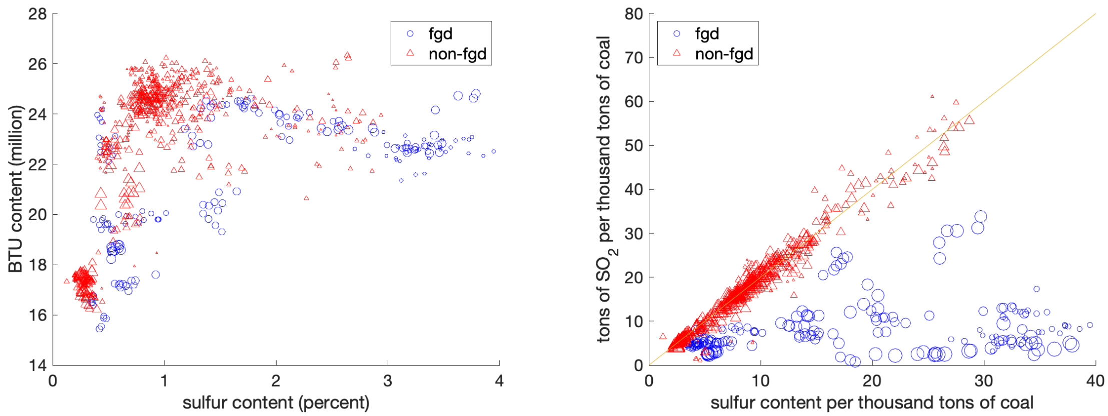


(a) Total generation by plant type

(b) SO_2 Emission per MWh Generation

In Figure 2a, we plot the plant-year average Btu content against sulfur content weighted by the quantity of coal. This indicates a wide variety of Btu-sulfur combinations. We represent the non-FGD plants with triangles and the FGD plants with circles, where the size of each indicates the magnitude of coal purchases in logarithms. Plants possess considerable flexibility in trading off these two characteristics of coal. The ranges of sulfur and Btu content are greater for FGD plants since they can employ FGD devices.

Figure 2: Sulfur Content, Btu Content, and SO_2 Emission



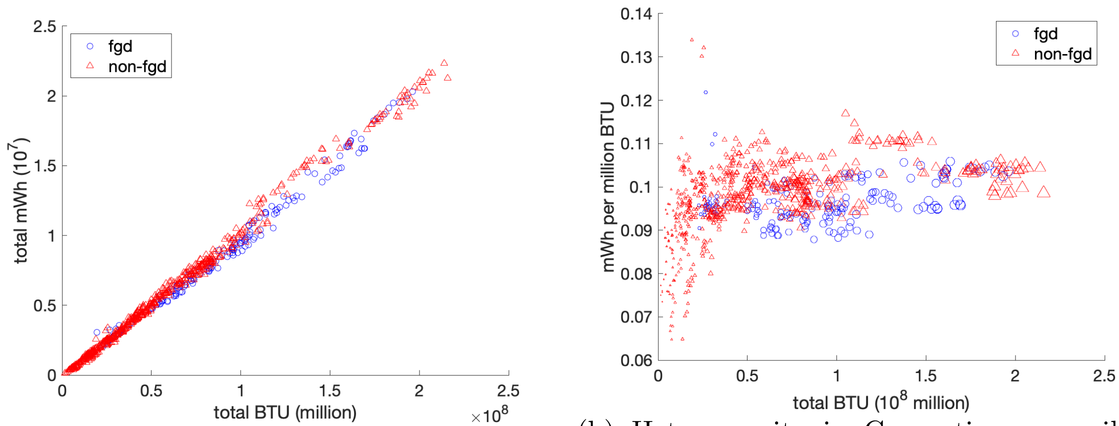
(a) Btu Content and Sulfur Content

(b) SO_2 Emission and Sulfur Content

In Figure 2b, we plot the tons of SO_2 emissions per thousand tons of coal against the amount of sulfur per thousand tons of coal input. For the non-FGD plants, a linear relationship exists between the two variables with a slope of approximately two, as expected from the chemistry of converting sulfur into SO_2.[14] The FGD plants exhibit a wide range of differences in the percent of emissions per ton of coal input relative to sulfur content. This occurs since plants differ substantially in the percent of coal burned in units with FGD devices, as well as the removal rates of FGD devices shown in Table 1. This data also raises the possibility that plants have heterogeneous abatement efficiencies.

[14]When we regress the amount of SO_2 emitted on the amount of sulfur in the coal for the non-FGD plants, the coefficient is 1.94 and significant at the 0.01 level.

Figure 3: kWh Generation versus Btu Consumption



(a) kWh generation versus Btu consumption

(b) Heterogeneity in Generation per million Btu

In Figure 3a, we graph the total generation against total heat input in Btu for non-FGD and FGD plants. It shows a close to linear relationship between generation and heat input, which indicates that heat is a key input for the plants. Nonetheless, for a given total Btu, there exists a great variation in electricity generation per million Btu across plants and years, as shown in Figure 3b. For small plants with total Btu less than 0.5 on the graph, the heterogeneity in MWh per million Btu is dramatic, varying by as much as 200%. For larger plants, this measure can vary by as much as 50%. Therefore, the plants have heterogeneous generation efficiencies.

4 Cost-Minimization Problem by Coal Plants

In this section, we model the cost minimization problems of the FGD and non-FGD plants. Both types of plants choose the Btu and sulfur content to minimize total variable costs subject to constraints on generation and the ranges of Btu and sulfur content. Total variable cost is the sum of the costs of coal and pollution control. FGD plants have the option to abate SO_2, which non-FGD plants lack. Each plant faces trade-offs when choosing its sulfur and Btu content, both of which affect the price of coal and the amount of SO_2 generated. We solve the constrained cost-minimization problems and derive the total variable cost function for each type of plant.

By assuming that the plants are cost minimizing, we rule out two types of behavior. First, we assume that the plants do not intentionally waste inputs in order to increase costs.[15] Second, we assume that the plants cannot flexibly choose output to maximize profits. This is a reasonable assumption because the electricity demand in a service area is largely determined by regulated electricity prices and the demand function, not by the plants' choices of output. More importantly, all of the plants belong to multi-plant companies, and the parent companies allocate the generation across plants.

4.1 Costs of Coal

In year t, plant j first observes its generation capacity, k_{jt}, labor stock, l_{jt}, generation productivity, ω^y_{jt}, and target output, y_{jt}. It then chooses the quality of coal to produce the target output and meet emission restrictions. The two key quality characteristics of coal are the Btu content per ton of coal, b_{jt}, and the sulfur content per ton of coal, s_{jt}. They affect the cost of coal for a plant, since the coal price increases with b_{jt} and decreases with s_{jt}. Additionally, given the output level, the amount of coal a plant needs depends on b_{jt}. The higher b_{jt}, the less coal the plant consumes.

Although plants minimize costs to generate target outputs, we do not assume that y_{jt} is exogenous. Since the parent company allocates the total generation across plants

[15] Abito (2020) finds *utilities* operating from 1988-99 and consuming on average 93% of fuel as coal have purposely incurred higher than efficient costs during the period of regulatory review. This increased inefficiency, measured by increased heat rate (MMBtu/MWh), is arguably due to utilities changing the efficient distribution of load across their plants during rate review by allocating more production to less efficient plants. The resulting increase in variable costs convinces regulators to approve higher allowed electricity prices to cover these costs and guarantee a specified return on capital. During the subsequent period of regulatory lag, utilities reduce the heat rate through efficient allocation of load and earn rents. However, this does not imply that their *plants* waste resources and thereby do not minimize variable costs. There are two basic but easily detectable ways a plant could temporarily increase its heat rate by wasting resources. First, a plant could dump or bury coal. However, this would violate environmental standards. Second, a plant could generate and immediately vent steam, without producing electricity. However, steam venting is reserved for emergencies, such as when transmission lines suddenly fail. Non-emergency venting would be immediately detected by regulators.

based on their relative efficiencies, more efficient plants will generate more output. This implies a positive correlation between y_{jt} and ω^y_{jt}. We do not model the plants' choices of (l_{jt}, k_{jt}) and assume that the plants choose them at least one period ahead.[16]

Let the total heat input (Btu consumed) be h_{jt} and assume that the plant's non-stochastic production function for electricity has a Leontief form,

$$y_{jt} = e^{\beta_0} \min\left\{e^{\omega^y_{jt}} l^{\beta_l}_{jt} h^{\beta_h}_{jt}, k^{\beta_k}_{jt}\right\}, \tag{1}$$

where $(\beta_0, \beta_l, \beta_k, \beta_h)$ are parameters and ω^y_{jt} captures the plant's heterogeneity due to differences in generation productivity, which are unobserved in the data. We assume this Leontief form due to the complementarity between capacity and heat. When the capacity is fully utilized, the plant cannot increase generation by only increasing heat or labor inputs. When the capacity is not fully utilized, the plant cannot increase generation by only increasing the capacity. Given the production function in (1) and conditional on $y_{jt} \le e^{\beta_0} k^{\beta_k}_{jt}$, the heat needed to produce y_{jt} for a given (l_{jt}, ω^y_{jt}) is

$$h_{jt} = h(y_{jt}, l_{jt}, \omega^y_{jt}) = h(X_{jt}, \omega^y_{jt}) = (y_{jt} e^{-(\beta_0 + \omega^y_{jt})} l^{-\beta_l}_{jt})^{\frac{1}{\beta_h}}, \tag{2}$$

where $X_{jt} = (y_{jt}, l_{jt})$.[17] The total heat decreases as ω^y_{jt} increases or as y_{jt} decreases.

For any Btu content, the tons of coal, n_{jt}, required to produce y_{jt} by plant j is the total amount of heat divided by the Btu content. That is,

$$n_{jt} = n(b_{jt}; X_{jt}, \omega^y_{jt}) = \frac{h(X_{jt}, \omega^y_{jt})}{b_{jt}} = (y_{jt} e^{-(\beta_0 + \omega^y_{jt})} l^{-\beta_l}_{jt})^{\frac{1}{\beta_h}} b^{-1}_{jt}. \tag{3}$$

A higher b_{jt} implies that a lower n_{jt} will produce a given y_{jt}. Denote the price of coal

[16] For our plants, over time capital (capacity) increases very little, while labor declines more substantially. We assume that plants choose both inputs at least one period ahead, since expanding capacity is a multi-year process and increasing the labor force requires training and certification.

[17] When the plants face output targets, the heat input demand function depends on the specified output levels. If we invert the heat input function (as a function of ω^y_{jt} and y_{jt}) and plug it into the production function as in the LP and ACF approaches, the output, y_{jt}, would be on both sides of the equation, which would make running a regression problematic.

per ton delivered to the plant by w^c_{jt}. It is a function of the Btu content, the sulfur content, and the transportation cost per ton, f_{jt}, from the mine to the plant. That is, $w^c_{jt} = w^c_{jt}(b_{jt}, s_{jt}, f_{jt})$. This coal price function is plant-year specific, and the subscript jt captures the plant's bargaining power in the coal market and the time fixed effects. Thus, the total cost of coal is the price of coal times the quantity of coal, which is given by

$$w^c_{jt}(b_{jt}, s_{jt}, f_{jt})n(b_{jt}; X_{jt}, \omega^y_{jt}). \tag{4}$$

4.2 Cost Minimization by Non-FGD Plants

A non-FGD plant does not abate SO_2 but must hold permits for all SO_2 emissions. Its pollution control costs include the expenditures to purchase permits and the opportunity cost of any permits allocated by the EPA. Since Figure 2b shows that the total weight of SO_2 is approximately two times the weight of total sulfur for non-FGD plants, we assume that all sulfur is transformed into SO_2. That is, the amount of SO_2 generated and emitted is $S^e_{jt} = 2s_{jt}n_{jt}$ tons. There is also an opportunity cost to use the allocated permits because a plant can trade them. Thus, the cost of permits to cover S^e_{jt} tons of SO_2 emission is the permit price, p_{jt}, times S^e_{jt}:

$$C^s_{NFGD}(b_{jt}, s_{jt}; X_{jt}, p_{jt}, \omega^y_{jt}) = p_{jt}S^e_{jt} = 2p_{jt}s_{jt}n_{jt} = 2p_{jt}s_{jt}\frac{h(X_{jt}, \omega^y_{jt})}{b_{jt}}, \tag{5}$$

where the subscript NFGD denotes non-FGD and the superscript s denotes SO_2 control cost. This cost increases with s_{jt} and decreases with b_{jt}. A higher s_{jt} increases the amount of sulfur, the SO_2 produced, and thus pollution control (permit) costs, while a higher level of b_{jt} reduces the total amount of coal consumed, the SO_2 produced, and thus the pollution control costs.

The non-FGD plant minimizes the sum of these permit costs and its total coal cost by choosing b_{jt} and s_{jt}. Constraints on shipment-level (b, s) are specified in either long-term, medium term, short term, or spot market contracts between plants and coal mines. Specifically, coal contracts are written with a lower bound on b_{jt}, the good input, and an upper bound on s_{jt}, the bad input. Under the terms of all contracts, if b_{jt} is less than a lower bound, monetary penalties apply and the plant may refuse delivery. Further, if s_{jt} exceeds an upper bound, similar conditions apply.[18]

We include these constraints in the plant's choice of (b_{jt}, s_{jt}) to minimize the total variable cost of coal and permits:

$$\min_{b_{jt}, s_{jt}} w^c_{jt}(b_{jt}, s_{jt}; f_{jt}) n(b_{jt}; X_{jt}, \omega^y_{jt}) + C^s_{\text{NFGD}}(b_{jt}, s_{jt}; X_{jt}, p_{jt}, \omega^y_{jt}), \tag{6}$$

subject to $b_{jt} \geq \underline{b}_{jt}$ and $s_{jt} \leq \bar{s}_{jt}$.[19] The optimal choice of (b_{jt}, s_{jt}) is conditional on the shipping cost f_{jt}.[20] The assumption is that, for the given f_{jt}, each plant can choose from a continuum of (b, s). Thus, we can model the choice of (b, s) as a continuous one and derive the FOCs. This assumption allows a plant to purchase coal with different b_{jt} and s_{jt} either at a given mine or at adjacent mines.[21] EIA Forms 423 and 923 (1995-2005) for coal deliveries provide evidence that a wide variety of coal with varying b_{jt} and s_{jt} is sold by mines in each major coal-producing state. Therefore, a plant can choose over a wide variety of different local combinations of (b, s) for the same shipping cost. A plant can even buy and mix coal from different regions at the same shipping

[18] A typical contract is that between the supplier Knight Hawk Coal, LLC and two electric utilities, Louisville Gas and Electric Company and Kentucky Utilities Company in 2019. Section 6 of that contract on Quality specifies the lower bound on b_{jt} ($10.9*10^6$ Btu/ton) and the upper bound on s_{jt} (3.0%). See http://psc.ky.gov. Another similar example is between Wabash River Energy, Ltd. and Midwest Mining Company, LLC in 2004, which specifies a minimum value for b_{jt} and a maximum value for s_{jt}. See https://www.sec.gov/Archives/edgar/data1/. The contract between Alliance Coal, LLC and Tennessee Valley Authority, 2009 redacts the exact upper and lower bound values. See https://www.sec.gov/Archives/edgar/data2/. The Southern Company's Master Coal Purchase and Sale Agreement leaves blanks for specific values of lower and upper bounds of b_{jt} and s_{jt}, respectively. See https://www.southerncompany.com/content/.

[19] We assume that the bounds of b_{jt} and s_{jt} are independent of each other for model tractability. If the bounds for (b, s) are functions of each other, then we would need to assume a parametric functional form, and limited data on the bounds are available.

[20] We do not model the choice of f_{jt} because the shipping cost per ton of coal and the distance for each coal shipment are confidential.

[21] In Section B of the Appendix, we provide evidence that a plant buys coal with different (b, s) from the same mine in each year, using transaction-level data.

cost, especially if the distances are roughly the same.

In Appendix B we present the Lagrangian expression for minimizing total variable costs subject to the constraints on b_{jt} and s_{jt}. From the Lagrangian expression, using the permit cost in equation (5), the FOCs for b_{jt} and s_{jt} are

$$\frac{\partial w_{jt}^c(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} n_{jt} + w_{jt}^c(b_{jt}, s_{jt}; f_{jt}) \frac{\partial n_{jt}}{\partial b_{jt}} - \frac{2 p_{jt} s_{jt} n_{jt}}{b_{jt}} - \underline{\mu}_{jt}^b = 0, \tag{7}$$

$$\frac{\partial w_{jt}^c(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} n_{jt} + 2 p_{jt} n_{jt} + \bar{\mu}_{jt}^s = 0, \tag{8}$$

where $\underline{\mu}_{jt}^b$ and $\bar{\mu}_{jt}^s$ are the plant-year Lagrangian multipliers for lower- and upper-bound constraints for b_{jt} and s_{jt}, respectively.[22] The plant equates the marginal cost of using a higher b_{jt} coal to its marginal savings. Assume that the firm increases b_{jt}, then the first term in equation (7) gives the increased cost of coal (equal to its increased price times quantity), while the second term gives the savings in terms of the price of coal times the reduced quantity of coal due to a higher b_{jt}. The third term gives the savings in terms of the reduced cost of permits as b_{jt} increases. The Lagrangian multiplier, $\underline{\mu}_{jt}^b$, measures the marginal impact of increasing $\underline{b}_{jt}$ on the total variable costs. When the bound on b_{jt} is not binding, increasing $\underline{b}_{jt}$ has no impact on the costs. When the bound is binding, increasing b_{jt} allows the plant to choose a higher b_{jt}, which can reduce the total variable costs. From equation (8), the plant equates the marginal cost of using a lower s_{jt} to its marginal savings. Assume that s_{jt} increases, then the first term is the reduced cost of coal (equal to its reduced price times quantity), while the second term is the expenditure on permits per unit of sulfur. The Lagrangian multiplier, $\bar{\mu}_{jt}^s$, is the marginal impact of increasing $\bar{s}_{jt}$ on the total variable costs.

In Appendix B, we provide clear-cut empirical evidence on which constraints are binding. Data on plant-mine-specific transactions show that the lower-bound constraint on b_{jt} is binding, while the upper-bound constraint on s_{jt} is not. Actual values of b_{jt} are

[22] See Appendix B for the explanation on how we deal with the Lagrangian multipliers for the constraints on sulfur and Btu content.

at a boundary while those for s_{jt} are quite disperse. This makes intuitive sense, since the mine does not wish to provide more of the good input than necessary, while the plants can always cover excess SO_2 emissions with permits. Therefore, the Lagrange multiplier (unknown) for the lower bound of b_{jt}, $\underline{\mu}^b_{jt}$, is positive. This means that we cannot directly use the FOC for b_{jt} in equation (7) to solve for the permit price. However, since the upper bound on s_{jt} is not binding, $\bar{\mu}^s_{jt} = 0$. Thus, we can use the FOC for s_{jt} to solve for p_{jt}. Plugging $\bar{\mu}^s_{jt} = 0$ into equation (8) and dividing by n_{jt}, we obtain the relationship between the marginal price of sulfur and p_{jt}:

$$\frac{\partial w^c_{jt}}{\partial s_{jt}} + 2p_{jt} = 0, \tag{9}$$

which implies that $p_{jt} = -\frac{\partial w^c_{jt}}{2\partial s_{jt}}$. Denote the optimal b_{jt} and s_{jt} by (b^*_{jt}, s^*_{jt}). The total variable cost function for the non-FGD plant is

$$C_{NFGD}(X_{jt}, \omega^y_{jt}; f_{jt}, p_{jt}) = (w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt}s^*_{jt})\frac{h(X_{jt}, \omega^y_{jt})}{b^*_{jt}}. \tag{10}$$

4.3 Cost Minimization by FGD Plants

The total pollution control cost function for FGD plants is very different from that for non-FGD plants. For an FGD plant, total pollution control cost includes not only the opportunity cost of using allocated permits and the cost of purchasing SO_2 permits, but also the expenditures on abating SO_2. FGD plants must choose (b_{jt}, s_{jt}) and the amount of coal to abate with the FGD devices, n^a_{jt}, in order to minimize costs. They solve a two-stage optimization problem. First, for any (b_{jt}, s_{jt}), the plants choose an optimal n^a_{jt} to minimize the pollution control costs. Second, the plants' minimize the total variable cost by choosing the optimal (b_{jt}, s_{jt}), conditional on the optimal n^a_{jt} (which is a function of (b_{jt}, s_{jt})).

4.3.1 Abatement Costs

Given the sulfur content and the amount of coal used in FGD units, the amount of sulfur scrubbed by the plant is $n^a_{jt}r_js_{jt}$, where r_j is the removal rate of SO_2. Given $n^a_{jt}r_js_{jt}$, the plants can have different O&M costs to abate. This heterogeneity is due to differences in the cost of FGD labor, feed materials and chemicals, waste disposal, and other related costs across plants. To represent the heterogeneity in these costs, we denote the unobserved abatement efficiency by ω^a_{jt}. If a plant pays a higher cost, then the plant has a lower ω^a_{jt}. Therefore, the total abatement cost depends on the amount of controlled sulfur, $n^a_{jt}r_js_{jt}$, and the efficiency, ω^a_{jt}. Denote the total abatement cost function by $C^a(n^a_{jt}, s_{jt}, r_j, \omega^a_{jt})$. We assume that abatement cost is a power function of the total amount of sulfur scrubbed:[23]

$$C^a(n^a_{jt}, s_{jt}, r_j, \omega^a_{jt}) = e^{\lambda_0 - \omega^a_{jt}}(n^a_{jt}s_{jt}r_j)^{\lambda}. \tag{11}$$

The constant λ_0 measures the average log abatement cost to remove one unit of sulfur for all FGD plants. The parameter λ determines the monotonicity of the marginal abatement cost. If $\lambda > 1$, then the marginal cost of abatement increases with the abatement level.

After controlling the SO_2, the remaining sulfur in the coal, $s_{jt}(n_{jt} - n^a_{jt}r_j)$, will be converted to $S^e_{jt} = 2s_{jt}(n_{jt} - n^a_{jt}r_j)$ tons of SO_2. The cost of buying emission permits and holding allocated ones to cover unabated emissions is

$$p_{jt}S^e_{jt} = 2p_{jt}s_{jt}(n_{jt} - n^a_{jt}r_j). \tag{12}$$

[23] We justify this functional form in Section 6.

4.3.2 Trade-off between Abating and Emitting SO_2 for FGD Plants

For a given (b_{jt}, s_{jt}), the FGD plant chooses the amount of coal used in FGD units (n^a_{jt}) to minimize the pollution control cost. Its pollution control cost is the sum of the cost of using scrubbers and the cost of permits. The minimization problem for the pollution control cost is

$$\min_{n^a_{jt}} \left\{ e^{\lambda_0 - \omega^a_{jt}} (n^a_{jt} s_{jt} r_j)^{\lambda} + 2p_{jt} s_{jt} (n_{jt} - n^a_{jt} r_j) \right\}. \tag{13}$$

When $\lambda > 1$, the marginal cost of abatement increases with n^a_{jt}, which determines a unique optimal abatement level by equating the marginal abatement cost to the permit price. Thus, the optimal amount of coal that should be scrubbed, n^{a*}_{jt}, must satisfy the FOC:

$$e^{(\lambda_0 - \omega^a_{jt})} \lambda (n^a_{jt})^{\lambda - 1} (s_{jt} r_j)^{\lambda} - 2p_{jt} s_{jt} r_j = 0.$$

We use this FOC to solve for n^{a*}_{jt} as

$$n^{a*}_{jt} = \left(\frac{2p_{jt} e^{\omega^a_{jt} - \lambda_0}}{\lambda} \right)^{\frac{1}{\lambda - 1}} \frac{1}{s_{jt} r_j}. \tag{14}$$

It increases with ω^a_{jt} and p_{jt} but decreases with r_j, s_{jt}, and λ. Plugging n^{a*}_{jt} into the pollution control cost equation (13), we obtain the minimized pollution control cost function, conditional on (b_{jt}, s_{jt}). Denote the pollution control cost of an FGD plant as

$$C^s_{\text{FGD}}(b_{jt}, s_{jt}; r_j, p_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt}) = \left(\frac{1}{\lambda} - 1 \right) 2p_{jt} \left(\frac{2p_{jt} e^{\omega^a_{jt} - \lambda_0}}{\lambda} \right)^{\frac{1}{\lambda - 1}} + 2p_{jt} s_{jt} \frac{h(X_{jt}, \omega^y_{jt})}{b_{jt}}. \tag{15}$$

The last term is the cost that plants with FGD units would incur if they performed no control, but instead covered all emissions with permits. However, they can reduce their total control costs below this level by abating sulfur. This cost savings is given by the first term in (15). It is the difference between the cost of control for the abated SO_2 and the expenditure that would have been incurred on SO_2 permits if these emissions were unabated. If $\lambda > 1$, then this difference is negative. That is, the pollution control costs by an FGD plant are less than what they would have been if it had relied solely on permits.

4.3.3 Total Variable Cost Function for FGD Plants

An FGD plant chooses (b_{jt}, s_{jt}) to minimize the total variable cost, which is the sum of the coal cost and the minimized pollution control cost from above:

$$\min_{b_{jt}, s_{jt}} w^c_{jt}(b_{jt}, s_{jt}; f_{jt}) n(b_{jt}; r_j, p_{jt}, X_{jt}, \omega^y_{jt}) + C^s_{\text{FGD}}(b_{jt}, s_{jt}; r_j, p_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt}), \tag{16}$$

subject to $b_{jt} \geq \underline{b}_{jt}$ and $s_{jt} \leq \bar{s}_{jt}$.The FOCs for the corresponding Lagrangian with respect to b_{jt} and s_{jt} are

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} n_{jt} + w^c_{jt}(b_{jt}, s_{jt}; f_{jt}) \frac{\partial n_{jt}}{\partial b_{jt}} + \frac{\partial C^s_{\text{FGD}}(s_{jt}, b_{jt}; r_j, p_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt})}{\partial b_{jt}} - \underline{\mu}^b_{jt} = 0. \tag{17}$$

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} n_{jt} + \frac{\partial C^s_{\text{FGD}}(s_{jt}, b_{jt}; r_j, p_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt})}{\partial s_{jt}} + \bar{\mu}^s_{jt} = 0, \tag{18}$$

which are similar to the two FOCs for the non-FGD plants.

Denote the optimal b_{jt} and s_{jt} by (b^*_{jt}, s^*_{jt}). The total variable cost function for FGD plants is

$$C^s_{FGD}(X_{jt}, \omega^y_{jt}, \omega^a_{jt}; f_{jt}, p_{jt}, r_j) = \left(w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt}s^*_{jt}\right) \frac{h(X_{jt}, \omega^y_{jt})}{b^*_{jt}} + \left(\frac{1}{\lambda} - 1\right) 2p_{jt} \left(\frac{2p_{jt}e^{\omega^a_{jt}}}{\lambda}\right)^{\frac{1}{\lambda-1}}. \tag{19}$$

The first term is the total cost of coal and permits if the FGD plant does not use FGD devices to abate sulfur. The second term is the savings in the pollution control cost if the plant runs the FGD devices.

5 Econometric Model and Estimation

Our estimation of the model parameters consists of three steps. In the first step, we estimate the hedonic coal price function using coal transaction-level data for all plants in our sample. Using the estimated price function, we compute the marginal prices of b_{jt} and s_{jt} for each plant-year observation. We then compute the permit prices, which are used in steps two and three. In the second step, we estimate the abatement cost function using plant-year-level data on the FGD O&M cost, the sulfur abatement level, and permit prices. In the third step, we estimate a cost function with plant-year-level data, which yields the estimates of the production function parameters. Before estimating the last two steps, we add the transition functions of the unobserved productivities and efficiencies to help identify the parameters, which is analogous to the traditional production function literature.

To reiterate, our sample is the 76 largest coal-fired power plants in the US from 1995-2005. Throughout our sample period, the set of the 76 largest coal-fired power plants does not change. This obviates the need to model sample selection as in Olley and Pakes (1996).

5.1 Estimation of the Coal Price Function

We use transaction-level data to estimate the hedonic coal price function, where a transaction occurs between a mine (m) and plant (j) during a given month-year (τ). For simplicity we drop the subscripts for m, j, and τ. We sometimes observe multiple transactions between m and j in a given τ, and we use them as unique observations. The delivered price of w^c depends on b, s, plant fixed effects (d^j), mine fixed effects (d^m), month-year dummies (d^τ), the total annual allowances for the US in that year (A^τ), the contract–type dummies (d^q), and the freight transportation charges per ton of coal, f.[24]

We assume that plants face a coal price function, $w^c(b,s;f) = \overline{w}^c(b,s) + f$, where $\overline{w}^c(b,s)$ is the mine-mouth price of coal and f is the transportation charge per ton of coal. The transportation charge is unobserved and depends on the distance between the mine and plant in addition to the rail carriers' market power, so f is independent of (b,s). We use d^m, d^j, and their interactions to control for f.

The stochastic hedonic coal price function is

$$\begin{aligned} w^c(b,s;f) = {} & \alpha_0 + \alpha_b b + \alpha_s s + \alpha_{bb} b^2 + \alpha_{ss} s^2 + \alpha_{bs} bs \\ & + \alpha_{sA} s\, A^\tau + \sum_{q=1}^{Q} \alpha_q d^q + \sum_{\tau=1}^{\mathcal{T}} \alpha_\tau d^\tau + \sum_{m=1}^{M} \alpha_m d^m \\ & + \sum_{j=1}^{J} \alpha_j d^j + \sum_{m=1}^{M} \sum_{j=1}^{J} \alpha_{mj} d^m d^j + \epsilon^w, \end{aligned} \tag{20}$$

where ϵ^w is a coal price shock.

Using coal transaction data, we compute OLS estimates of the stochastic coal price equation (20). After estimating this function, we aggregate the transaction-level data to obtain plant-year average b_{jt} and s_{jt}, weighted by the coal quantity. We evaluate the plant-year marginal prices of b_{jt} and s_{jt} as

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} = \alpha_b + 2\alpha_{bb} b_{jt} + \alpha_{bs} s_{jt}, \tag{21}$$

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} = \alpha_s + 2\alpha_{ss} s_{jt} + \alpha_{bs} b_{jt} + \alpha_{sA}\, A_t. \tag{22}$$

Since plants do not report the prices of SO_2 permit trades to EPA, we do not observe p_{jt} in the data. After estimating the coal price function, we compute p_{jt} using the

[24] Total allowances include both the new allowances issued in that year and the allowances banked from previous years. There are three types of coal purchase contracts depending on the length of the contract and whether it is new. Type C contracts have a duration of at least one year. Type NC contracts are new or renegotiated where deliveries are first made during the reporting month. Type S are for the spot-market purchases with a duration of less than one year.

FOC for the optimal s_{jt} in equation (9) for both types of plants, which implies that $p_{jt} = -\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{2\partial s_{jt}}$.

5.2 Estimation of the Abating Cost Function

To capture the potential persistence in the FGD plants' abatement efficiencies, we assume that ω^a_{jt} follows a Markov process:

$$\omega^a_{jt} = g^a(\omega^a_{jt-1}) + \xi^a_{jt} = \rho^a_0 + \rho^a_1 \omega^a_{jt-1} + \rho^a_2 (\omega^a_{jt-1})^2 + \xi^a_{jt}, \tag{23}$$

where $g^a(\omega^a_{jt-1}) = \rho^a_0 + \rho^a_1 \omega^a_{jt-1} + \rho^a_2 (\omega^a_{jt-1})^2$ is a second-order expansion of ω^a_{jt-1}, and ξ^a_{jt} is the shock to the abating efficiency. Specifically, ξ^a_{jt} represents the shock to the plant-year costs of FGD labor, feed, waste disposal, and related costs. We assume that ξ^a_{jt} is independent of ξ^a_{jt-1}. The stochastic version of the logarithm of abatement cost in equation (11) is

$$\ln C^a(s_{jt}, n^a_{jt}, \omega^a_{jt}) = \lambda_0 - \omega^a_{jt} + \lambda(\ln s_{jt} + \ln n^a_{jt} + \ln r_j) + \epsilon^a_{jt}, \tag{24}$$

where ϵ^a_{jt} is an idiosyncratic error that represents the measurement error in the O&M cost data. This error term does not affect the identification of the abatement cost function and transition function of ω^a_{jt}, since we assume that ϵ^a_{jt} has a zero mean and is uncorrelated with the variables on the right-hand side of equation (24). The parameters to be estimated in this step are λ in equation (11) and $\boldsymbol{\rho}^a = (\rho^a_0, \rho^a_1, \rho^a_2)$ in equation (23).

Replacing ω^a_{jt} in equation (24) with $\omega^a_{jt} = g^a(\omega^a_{jt-1}) + \xi^a_{jt}$, we get

$$\ln C^a(s_{jt}, n^a_{jt}, \omega^a_{jt}) = \lambda_0 - g^a(\omega^a_{jt-1}) + \lambda(\ln s_{jt} + \ln n^a_{jt} + \ln r_j) - \xi^a_{jt} + \epsilon^a_{jt}. \tag{25}$$

The FOC for the abatement level in equation (14) implies that

$$\omega^a_{jt-1} = \lambda_0 + (\lambda - 1)\ln(s_{jt-1}n^a_{jt-1}r_j) + \ln(\frac{\lambda}{2p_{jt-1}}). \tag{26}$$

Plugging this into equation (25), we replace the unobserved abating efficiency with observed variables. The new error term is $(-\xi^a_{jt} + \epsilon^a_{jt})$.

Equation (25) contains two endogenous variables. The efficiency shock ξ^a_{jt} is positively correlated with n^a_{jt} in the abatement cost equation (25). This is because a plant with a higher abating efficiency abates more coal to reduce the pollution control cost. Further, ξ^a_{jt} is also correlated with s_{jt} because more efficient plants choose higher sulfur content coal. Thus, we use GMM with IVs to estimate λ and $\boldsymbol{\rho}^a$.[25] Denoting the vector of IVs by Z^a_{jt}, the moment conditions are

$$E[Z^a_{jt}(-\xi^a_{jt} + \epsilon^a_{jt})] = 0.$$

The vector $Z^a_{jt} = (1, \log(w^c_{jt-1}), \log(w^c_{jt-1})^2, \text{yield}_{jt}, \log(r_j))$ includes the constant and four IVs, which are the log of the lagged price of coal and its square, the current yield of capital, and the log of the removal rate. We use these IVs for the following reasons. First, w^c_{jt-1} is not correlated with ξ^a_{jt} because ξ^a_{jt} does not affect the sulfur and Btu content in period $t-1$ or the lagged coal price. This instrument should be relevant since it is correlated with the endogenous variable, s_{jt}, because a high coal price in year $t-1$ gives the plant incentives to choose higher s_{jt} in order to reduce coal costs. Second, the current cost of capital is valid since it is uncorrelated with the current shock ξ^a_{jt}. This instrument is also relevant since as it increases, the cost of running the scrubbers and the depreciation rate of FGD capital increase. Thus, the plants have the incentive to reduce the abatement level, n^a_{jt}. This leads to a negative correlation between n^a_{jt} and the cost of capital, which is supported by the data.[26] Third, r_j is a characteristics baked

[25]Notice that the constants λ_0 and ρ^a_0 are not separately identified.

[26]When we regress n^a_{jt} on the set of IVs, the coefficient for $yield_{jt}$ is negative and significant.

into an FGD device when it is manufactured. In the data, r_j was virtually constant over time.[27] Thus, r_j is not correlated with the current shock ξ^a_{jt}. Meanwhile, r_j is positively correlated with n^a_{jt}, because plants with higher removal rates have greater incentives to abate. By construction, ϵ^a_{jt} is uncorrelated with the instruments. These IVs pass a commonly-used weak-IV test.[28]

Let Z^a denote the N^a-by-5 matrix of Z^c_{jt}, where N^a is the number of observations for FGD plants. The weighting matrix of the moment conditions is the inverse of $(Z^{a'}(Z^a Z^{a'})^{-1} Z^a)$. To simplify the notation, let $\theta^a = (\lambda, \rho^a_0, \rho^a_1, \rho^a_2)$, and let the composite error term be $\eta^a_{jt} = (-\xi^a_{jt} + \epsilon^a_{jt})$. The GMM objective function is

$$Q^a(\theta^a) = \frac{1}{N^a}(Z^a \eta(\theta^a))' * (Z^{a'}(Z^a Z^{a'})^{-1} Z^a)^{-1} * (Z^a \eta(\theta^a)), \tag{27}$$

where $\eta(\theta^a) = \{\eta_{jt}(\theta^a)\}_{jt}$ is a N^a-by-1 vector of the composite error terms. The GMM algorithm searches for the θ^a that minimizes this objective function.

5.3 Estimation of the Production Function Parameters

In this step, we use the total variable cost function to estimate the Leontief coefficients in the production function, $\beta = (\beta_0, \beta_l, \beta_h)$. To capture the persistence in the plants' generation efficiencies, we also assume that ω^y_{jt} follows a Markov process:

$$\omega^y_{jt} = g^y(\omega^y_{jt-1}) + \xi^y_{jt} = \rho^y_0 + \rho^y_1 \omega^y_{jt-1} + \rho^y_2 (\omega^y_{jt-1})^2 + \xi^y_{jt}, \tag{28}$$

where ξ^y_{jt} is the shock to ω^y_{jt}, due to unanticipated changes in the operating productivity of a plant. We assume that ξ^y_{jt} and ξ^y_{jt-1} are independent of each other, and estimate the parameters in the productivity transition function, $\boldsymbol{\rho}^y = (\rho^y_0, \rho^y_1, \rho^y_2)$.

In the previous step, we estimated the abatement cost function parameters, and they

[27] See Table C.1 in section C of the Appendix.

[28] We regress each of the two endogenous variables on the set of IVs. The F-values are 28.54 and 14.95 for s_{jt} and n^a_{jt}, respectively.

are independent of the production function parameters. Thus, we focus only on the function $\tilde{C}_{jt}(= w^c_{jt}n_{jt} + 2p_{jt}s_{jt}n_{jt})$, the cost of burning coal plus the cost of permits to cover all emissions. The parameters in this function are those of the Leontief production function. This cost term, $\tilde{C}_{jt}$, has the same expression for the two types of plants. Taking the logarithm of the non-FGD plants' total costs in equation (10) and the first two terms in the FGD plants' total costs in (19), we get the equivalent expression for both types of plants,

$$\ln \tilde{C}_{jt} = \ln\left(\frac{w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt}s^*_{jt}}{b^*_{jt}}\right) + \frac{1}{\beta_h}(\ln y_{jt} - \beta_0 - \beta_l \ln l_{jt} - \omega^y_{jt}) + \epsilon^y_{jt}, \quad (29)$$

where we replace h_{jt} using equation (2). The error term, ϵ^y_{jt}, is the idiosyncratic plant-year error term that captures measurement error in the coal costs. It does not affect the current sulfur and Btu content choices of the plants. We assume that it has a zero mean. Rearranging this expression we obtain

$$\ln \mathcal{C}_{jt} = \ln \tilde{C}_{jt} - \ln\left(\frac{w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt}s^*_{jt}}{b^*_{jt}}\right) = \frac{1}{\beta_h}(\ln y_{jt} - \beta_0 - \beta_l \ln l_{jt} - \omega^y_{jt}) + \epsilon^y_{jt}. \quad (30)$$

Plugging $\omega^y_{jt} = g^y(\omega^y_{jt-1}) + \xi^y_{jt}$ into equation (30) yields

$$\ln \mathcal{C}_{jt} = \frac{1}{\beta_h}(\ln y_{jt} - \beta_0 - \beta_l \ln l_{jt} - g^y(\omega^y_{jt-1})) - \frac{1}{\beta_h}\xi^y_{jt} + \epsilon^y_{jt}. \quad (31)$$

Taking the logarithm of the lagged production function yields $\omega^y_{jt-1} = \ln y_{jt-1} - \beta_0 - \beta_l \ln l_{jt-1} - \beta_h \ln h_{jt-1}$. We substitute this into equation (31) to replace the unobserved ω^y_{jt-1}. One can interpret our approach as using the production function as the control function to substitute for the unobserved ω^y_{jt}.

The estimation of (31) is subject to endogeneity issues. If the parent firm allocates more generation to the more productive plants after observing ω^y_{jt}, then y_{jt} is positively correlated with ξ^y_{jt}. In this case, the more productive plants may need to hire more employees, which leads to a positive correlation between l_{jt} and ξ^y_{jt}. Therefore, we estimate the parameters in (31) using GMM with IVs. The moment conditions employ the orthogonality between the composite error term and the IVs in Z^y_{jt}:

$$E\left[Z^y_{jt}\left(-\frac{1}{\beta_h}\xi^y_{jt}+\epsilon^y_{jt}\right)\right]=0. \tag{32}$$

The vector $Z^y_{jt}=(1,\log(l_{jt-1}),\log(k_{jt-1)},\log(l_{jt-1})^2,\log(k_{jt-1})^2,\log(w^c_{jt-1}),\log(h_{jt-1}),$ $\text{wage}_{jt})$, which are the logs of the lagged values of labor input, capital input, labor squared, capital squared, coal price, and heat input, and current price of labor. These IVs are valid and relevant for the following reasons. First, l_{jt-1} and k_{jt-1} are uncorrelated with ξ^y_{jt} because they are determined before period t, but they are correlated with l_{jt} due to the persistency in labor and the complementarity between these two inputs. Second, w^c_{jt-1} affects the current sulfur and Btu content choice, but is uncorrelated with the current shock, ξ^y_{jt}. Third, the lagged heat input is uncorrelated with ξ^y_{jt}, but is positively correlated with the current labor input, because larger plants hire more labor. Next, the current wage rate is correlated with the lagged wage rate, which affects the current labor input since we assume that l_{jt} is predetermined before period t. At the same time, the current labor price is uncorrelated with the current shock, ξ^y_{jt}. Lastly, ϵ^c_{jt} is the idiosyncratic error in the total variable costs, which is uncorrelated with the instrumental variables. These IVs pass a commonly-used weak-IV test.[29]

Let Z^y denote the N^y-by-8 matrix of Z^y_{jt}, where N^y is the number of observations. The weighting matrix of the moment conditions is the inverse of $(Z^{y'}(Z^yZ^{y'})^{-1}Z^y)$. To simplify the notation, let $\theta^y=(\beta_h,\beta_l,\beta_0,\rho^y_1,\rho^y_2)$ and let the composite error term be $\eta^y_{jt}=(-\frac{1}{\beta_h}\xi^y_{jt}+\epsilon^y_{jt})$. The GMM objective function is

$$Q^y(\theta^y)=\frac{1}{N^y}(Z^y\eta(\theta^y))'*(Z^{y'}(Z^yZ^{y'})^{-1}Z^y)^{-1}*(Z^y\eta(\theta^y)), \tag{33}$$

[29] We regress each of the two endogenous variables on the set of IVs. The F-values are 3328 and 1513 for y_{jt} and l_{jt}, respectively.

where $\eta(\theta^y) = \{\eta_{jt}(\theta^y)\}_{jt}$ is a N^y-by-1 vector of the composite error terms. For each observation, $\eta_{jt}(\theta^y) = \ln \mathcal{C}_{jt} - \frac{1}{\beta_h}(\ln y_{jt} - \beta_0 - \beta_l \ln l_{jt} - g^y(\omega^y_{jt-1}))$. The GMM algorithm searches for the θ^y that minimizes the objective function.

We do not estimate the second term involving capacity in the Leontief production function of equation (1), because excess capacity existed for our sample of plants. The average utilization rate (also known as the capacity factor) was .62, with a standard deviation of .16.[30] Given this excess capacity, we cannot equate actual output to the second term involving capacity in the Leontief production function. More importantly, this implies that β_k does not affect the plants choices of coal characteristics and pollution control. Thus, we do not estimate β_k and focus on estimating (β_h, β_l) in this paper.

6 Estimation Results

Table 2 presents the estimation results of the coal price function. Table 3 shows the marginal prices of b_{jt} and s_{jt} computed using equations (21) and (22). The marginal prices of b_{jt} and s_{jt} are positive and negative, respectively, for all plants in all years and are significant at the 0.01 level. We find that the coal price goes up by \$1.78 on average if b_{jt} increases by one million Btu per ton of coal. The price drops by \$1.88 on average if s_{jt} increases by 1 percentage point, for example, from 1% to 2%. The estimate of α_{sA} is positive and significant. It means that, as the country-level annual emission permits increase, the marginal price of s_{jt} becomes less negative, implying that relaxing the regulations on SO_2 emissions reduces the negative impact of sulfur content on the price of coal.

[30]The utilization rate is the actual production of electricity in MWh divided by the product of the MW name-plate capacity and the number of hours in a year (8760).

Table 2: Estimates of Parameters in the Coal Price Function

Coefficient	Estimate	Coefficient	Estimate
α_s	1.079	α_{bb}	-0.003^{*}
	(0.392)		(0.002)
α_b	2.149^{***}	α_{bs}	-0.180^{***}
	(0.099)		(0.019)
α_{ss}	0.276^{***}	α_{sA}	0.046^{***}
	(0.030)		(0.002)
Observations	846		

Parentheses contain bootstrap estimated standard errors obtained using the pairs bootstrap. See Appendix D for more details. Henceforth, the symbols *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 levels using a two-tailed t-test, respectively.

Table 3: Estimated Marginal Prices of Sulfur and Btu

	Mean
$\frac{\partial w^c_{jt}}{\partial s_{jt}}$	-1.882^{***}
	(0.113)
$\frac{\partial w^c_{jt}}{\partial b_{jt}}$	1.783^{***}
	(0.048)

We compute estimated standard errors (in parentheses) using the Delta method.

Using the marginal prices of sulfur and the FOC for the two types of plants, we compute the permit prices by plant and year. The results show that the average estimated permit price fell substantially over the sample period. The average estimated price was \$99 per ton of SO_2 in 1995, with a standard deviation of \$28, but only \$81 in 2005, with a standard deviation of \$23.[31]

Table 4 shows the estimated parameters and standard errors for the abatement cost function (25).The estimate of λ is 2.01 and significant at the .1 level, which implies that the marginal cost of abating sulfur is increasing. This is very consistent with the marginal cost estimates obtained by running the Integrated Emission Control Model

[31]In Appendix B, we compare our estimates of the permit prices with the EPA auction prices.

(IECM) by Rubin, Berkenpas, and Zaremsky (2007).[32] The result that $\lambda > 1$ is consistent with the fact that the FGD plants use scrubbers to control some generated SO_2 and use permits for the rest, as shown in Figure 2b. This is because $\lambda > 1$ implies that the marginal abatement cost increases with the amount of sulfur removed. The estimated values of ω_{jt}^a vary substantially across plants. The average unit abatement cost for the most efficient plant is only 4% of that for the least efficient plant.[33] The estimates of ρ_1^a and ρ_2^a imply that significant persistency exists in the plants' abatement efficiencies.

Table 4: Estimates of Parameters in the Abatement Cost Function

	$ln(C^a)$
λ	2.01*
	(1.14)
ρ_0^a	-14.70
	(27.69)
ρ_1^a	5.26***
	(1.53)
ρ_2^a	-0.32
	(0.57)
Observations	160

Table 5 shows the estimated parameters and standard errors for step three of the estimation. To obtain estimated standard errors we employ the pairs bootstrap estimation of the coal price function (20) and (31), after we substitute the productivity transition equation.[34] The estimate of β_h is positive (=1.08) and significant at the 0.01 level. The impact of labor on total generation is insignificant. Plants exhibit slight increasing returns in labor and heat, conditional on having enough capacity, since

[32]In the IECM model, we use 80 data points for different values of b_{jt} and s_{jt}, MW capacity levels, control levels, and regions of the US. The ranges of these values are representative of our data.

[33] That is, $\frac{exp(min\{\omega_j^a\})}{exp(max\{\omega_j^a\})} = 0.04$, where ω_j^a is the average abatement efficiency of plant j during the 11 years.

[34]See Appendix D for details on the bootstrap.

$\hat{\beta}_h + \hat{\beta}_l = 1.05 > 1$, which is consistent with findings in the literature as summarized by Atkinson (2019). The estimated values of ω^y_{jt} indicate that the most productive plant can generate four times more electricity than the least productive plant, using the same input bundles.[35] The estimate of ρ^y_1 is 1.404 and significant at the 0.01 level, implying that lagged productivity significantly influences current productivity.[36]

Table 5: Estimates of Parameters in the Production Function and the Transition Function

β_0	-3.780***
	(0.929)
β_l	-0.031
	(0.153)
β_h	1.083***
	(0.065)
ρ^y_1	1.404***
	(0.403)
N	760

The standard errors are computed using the pairs bootstrap method.

We compute $\hat{\omega}^a$ using equation (26) and $\hat{\omega}^y$ using $\omega^y_{jt} = \ln y_{jt} - \beta_0 - \beta_l \ln l_j - \beta_h \ln h_j$ from the Leontief production function. The mean of $(\hat{\omega}^a, \hat{\omega}^y)$ across plants and years are 6.66 and 0.16, respectively.[37] We then calculate the changes of the generation efficiency and abatement efficiency over time. The results show that generation efficiency was very stable over time, with an average 0.2% yearly decrease in $exp(\hat{\omega}^y_{jt})$ during the sample period. However, abatement efficiency improved significantly, with an average 10% yearly reduction in $exp(-\hat{\omega}^a_{jt})$. This lead to a 37% decrease in unit abatement costs in 2005 compared to 1995. The correlation between the estimated ω^y_{jt} and ω^a_{jt} is 0.25 for the FGD plants.

[35] That is, $\frac{exp(max\{\omega^y_j\})}{exp(min\{\omega^y_j\})} = 4.00$, where ω^y_j is the average generation efficiency of plant j.

[36] In the estimation, we only keep the first-degree term because adding the second-degree term makes ρ^y_1 and ρ^y_2 both insignificant.

[37] Notice that the parameters, β_0 and ρ^y_0 are not separately identified. Thus, the estimates of ω^y_{jt} are only identified up to a constant.

7 Impacts of the Initial Permit Allocations

With the estimated model, we analyze an important policy issue: how the initial permit allocations for the SO_2 trading system affect the plants' costs, in the presence of high transaction costs imposed by the trade restrictions as introduced in section 2. While a power plant's historic emission rate was largely the criteria for allocations under the US Acid Rain Program, other cap-and-trade systems have considered or adopted initial permit allocations based on total emissions, emission rates, total output, and abatement efficiency.[38] In this analysis, we compare three different schemes to allocate SO_2 emission permits across plants. Under the first scenario, in each year, we allocate permits based on the plants' observed emission shares in 1995. Under the second scenario, permits are proportional to each plant's annual electricity generation.[39] Under the third scenario, permits are proportional to the generation efficiency of each plant.[40] In all scenarios, the aggregate amount of permits equals the aggregate annual emissions of all plants in the data.

We first check the heterogeneity of the estimated abatement efficiency within and across states by regressing ω^a_{jt} on the state dummies and plant-year-level FGD device characteristics. Table 6 shows the results with and without the state dummies. By comparing the R^2, we find that most of the heterogeneity lies in the cross-state differences. This means that the plants have similar abatement efficiencies within each state. The prohibition of interstate permit trading leaves little room for the plants to

[38]For example, the California CO_2 system allocates permits principally based on emissions and to a lesser degree on abatement efficiency. The EU provides substantial emission-based allocations to the industrial and airline sectors. Spain has allocated CO_2 permits to coal-fired power plants based on emissions with a quadratic term rewarding abatement efficiency. See Reguant and Ellerman (2008) for details. New Zealand rewards higher-polluting, trade-exposed industries with an allocation scheme based on emission rates. The emission rate is the total emissions divided by total output.

[39]Plant j's amount of permits in year t is $\frac{y_{jt}}{\sum_{j'} y_{j't}}$ times the total number of permits in year t.

[40]Plant j's number of permits in year t is $\frac{exp(\omega^y_{jt})}{\sum_{j'} exp(\omega^y_{j't})}$ times the total number of permits in year t. Notice that, although ω^y_{jt} is not separately identified from β_0, $\frac{exp(\omega^y_{jt})}{\sum_{j'} exp(\omega^y_{j't})}$ is identified because $\frac{exp(\omega^y_{jt})}{\sum_{j'} exp(\omega^y_{j't})} = \frac{exp(\beta_0+\omega^y_{jt})}{\sum_{j'} exp(\beta_0+\omega^y_{j't})}$. We do not analyze an allocation based on abatement efficiency because the non-FGD plants do not have this efficiency measure.

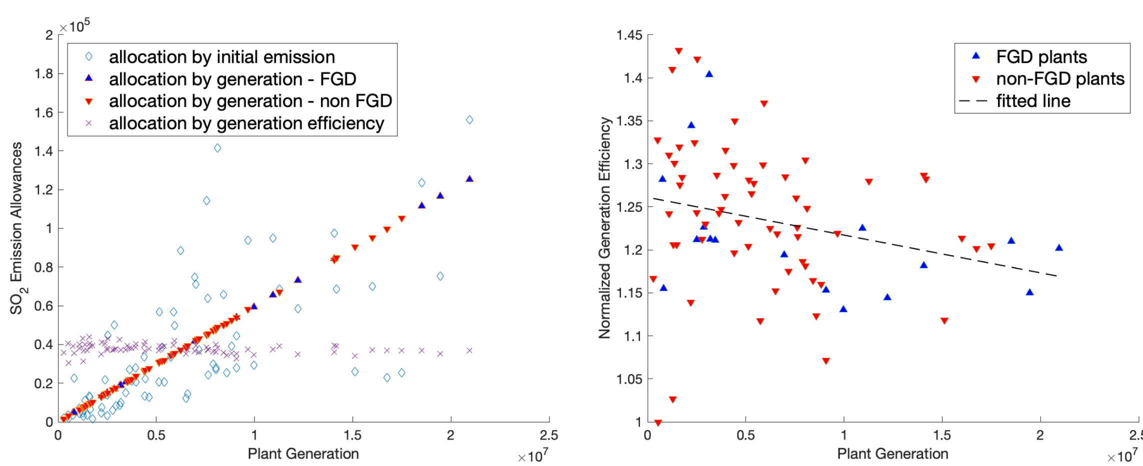
trade permits within states to improve cost efficiency. We first assume that the permit system is non-functioning so that plants do not trade any permits.[41] In Appendix E, we model the case of intra-state trading, finding that when we allow the same ranges of b_{jt} and s_{jt} as in the original counterfactual, intra-state permit markets do not clear in some states.

Table 6: Heterogeneity in $\hat{\omega}^a$ across States

	(1) With state dummies	(2) Without state dummies
FGD capital	0.612	0.084
	(0.123)***	(0.152)
FGD labor	-0.147	0.259
	(0.101)	(0.148)*
Removal rate	-2.727	8.032
	(1.217)**	(1.532)***
R^2	0.83	0.26
N	176	176

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure 4: Different Permit Allocations and Generation Efficiencies



(a) Different Allocations (b) Generation Efficiencies

The three methods imply very different distributions of allowances. We plot the

[41]Since we do not have data on transaction costs of the plants, we cannot estimate a transaction cost function. This implies that we cannot model different levels of transaction costs in this counterfactual analysis.

allocated permits on the vertical axis against the average annual plant generation on the horizontal axis in Figure 4a. Each point on the graph represents a plant. The scatter points with light blue diamond markers are allowances based on initial emissions. The points with solid triangular markers are allowances based on generation. The blue upward-pointing and red downward-pointing triangular symbols are for FGD plants and non-FGD plants, respectively. The x symbols represent allowances based on generation efficiency.

Figure 4a shows that, under the first two methods, plants with higher generation receive more permits, but with the first method, a considerably higher dispersion exists in allowances for a given generation level. The distribution of permits under the third method is close to uniform except for low-generation plants. This is because small plants are more efficient on average than large plants, as shown in Figure 4b, which plots the average generation efficiency against generation by plant.[42] The dashed OLS fitted line has a negative slope, implying that large plants have smaller $\hat{\omega}^y_{jt}$ on average. However, the dispersion is much higher for small plants.

Table 7 compares the total permit allocations for the two types of plants. Columns (1) to (3) list the total permits for the three counterfactual allocation methods, and column (4) is emission in the data. As shown in Figure 4a, among the three methods, the FGD plants receive the most permits when allocation is based on generation. The non-FGD plants receive the most permits when allocation is based on production efficiency.

Table 7: Total Permits of Plants (all years)

	(1) allocation by initial emissions w/o trading	(2) allocation by generation w/o trading	(3) allocation by $exp(\omega^y)$ w/o trading	(4) allocation in data with trading
Total permits (10^7)	3.20	3.20	3.20	3.20
– FGD	0.72	1.02	0.68	0.65
– Non-FGD	2.48	2.28	2.52	2.55

[42]The vertical axis is a plant's average normalized $exp(\omega^y)$. In each year, we normalize $exp(\omega^y)$ by its minimum of all plants. The average is over the 11 years from 1995 to 2005.

Under each allocation method, denote the new SO_2 allowances for plant j in year t by $\tilde{S}^e_{jt}$. We fix the shipping charges at their estimated values, since substantial flexibility exists in the choice of (b_{jt}, s_{jt}) for a given shipping charge, as described in Section 4.2. We constrain the simulated (b_{jt}, s_{jt}) to lie within plus or minus two standard deviations of their data values, which cover 95% of observed (b_{jt}, s_{jt}) for each plant.[43] Due to non-tradability, each plant will use all of its permits and minimize the total costs of coal and abatement. For FGD plants, the total cost is the sum of coal cost and abatement cost. An FGD plant only abates the SO_2 that exceeds its allowances. The new cost-minimization problem is to choose (b_{jt}, s_{jt}) to minimize the sum of coal cost and abatement cost to generate target levels of electricity and SO_2 emissions:

$$\min_{b_{jt}, s_{jt}} \left\{ w^c_{jt}(s_{jt}; b_{jt}, f_{jt}) n(b_{jt}; X_{jt}, \omega^y_{jt}) + e^{\lambda_0 - \omega^a_{jt}} (n^a_{jt} s_{jt} r_j)^\lambda \right\}, \tag{34}$$

subject to the emission constraint that $2\, s_{jt}\, [n(b_{jt}, X_{jt}, \omega^y_{jt}) - n^a(s_{jt}; b_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt}) r_j] \leq \tilde{S}^e_{jt}$. The generation constraint is embedded in the coal demand function, $n(b_{jt}; X_{jt}, \omega^y_{jt})$. For a non-FGD plant, the total cost is the cost of coal. The emission constraint is $2 s_{jt}\, n(b_{jt}, X_{jt}, \omega^y_{jt}) \leq \tilde{S}^e_{jt}$. This is binding when the plant minimizes cost because it cannot trade any unused permits.

We solve for each plant's cost-minimizing (b_{jt}, s_{jt}) under the three allocation scenarios and compare the aggregate and average costs across the three allocation methods.[44] Columns (1) to (3) in Table 8 show results for the three counterfactual scenarios, and column (4) shows the results from the estimated model. Compared with column (4), all plants choose lower sulfur content in columns (1)-(3) where permits are non-tradable. When plants can trade permits as in column (4), FGD plants would sell permits and purchase higher-sulfur-content coal if the marginal abatement cost were lower than the

[43] We compute the standard deviations of b_{jt} and s_{jt} by plant using the 11 years of data.

[44] We use two methods to solve for (b_{jt}, s_{jt}), a derivative-based and a grid-search method. In the grid-search method, we discretize the range of b_{jt} (s_{jt}) to 500 grid points, so there are 250,000 combinations of (b_{jt}, s_{jt}). The two methods give very close results, and we present the results with the grid-search method in the paper.

permit price. Non-FGD plants who buy the permits would also purchase higher-sulfur-content coal since they have more permits in column (4). Across columns (1)-(3), the average sulfur content of FGD plants is the highest in column (2) at 1.79%. This is because, compared across the three methods, they receive the most permits when based on generation, as shown in Table 7. The average sulfur content of non-FGD plants is the highest in column (1), at 0.84%, because they receive the most permits in this scenario.

Table 8: Impacts of Different Emission Permit Allocation Mechanisms with No trading versus Trading in Data

	(1) allocation by emissions w/o trading	(2) allocation by generation w/o trading	(3) allocation by $exp(\omega^y)$ w/o trading	(4) allocation in data with trading
Average sulfur content (%)	1.14	1.07	1.08	1.24
– FGD	1.76	1.79	1.78	2.04
– Non-FGD	0.84	0.74	0.75	0.87
Average Btu content (10^6/ton)	21.95	21.98	22.23	22.44
– FGD	21.04	21.03	21.03	21.77
– Non-FGD	22.37	22.42	22.79	22.76
Average coal price ($/ton)	29.39	29.45	29.96	30.40
– FGD	23.26	23.22	23.22	24.71
– Non-FGD	32.26	32.35	33.11	33.05
Total coal consumption (10^9 tons)	2.51	2.51	2.48	2.37
– FGD	0.84	0.84	0.84	0.79
– Non-FGD	1.67	1.67	1.64	1.58
Coal costs all years ($$10^{10}$)	9.82	9.83	9.88	9.70
– FGD	2.66	2.66	2.66	2.67
– Non-FGD	7.16	7.17	7.22	7.04
Average coal abatement (%)	72.96	53.52	70.35	80.00
Abatement costs all years ($$10^9$)	2.04	1.12	1.48	2.28
Total variable costs all years ($$10^{10}$)	10.03	9.94	10.03	9.93
– FGD	2.86	2.77	2.81	2.82
– Non-FGD	7.16	7.17	7.22	7.04
Average variable cost per MWh ($)	13.36	13.24	13.35	13.22
– FGD	11.95	11.55	11.70	11.74
– Non-FGD	14.03	14.03	14.13	13.77

The sulfur content, Btu content, and coal price are averages for all the plants in all years weighted by generation. The coal costs, abatement costs, and total costs are the total values of all plants from 1995 to 2005. Coal abatement is the average coal abatement percentage weighted by generation of the FGD plants. The variable costs include the coal costs and abatement costs, not the permit costs. The costs are measured in 2005 dollars using the firm yield as the discount rate.

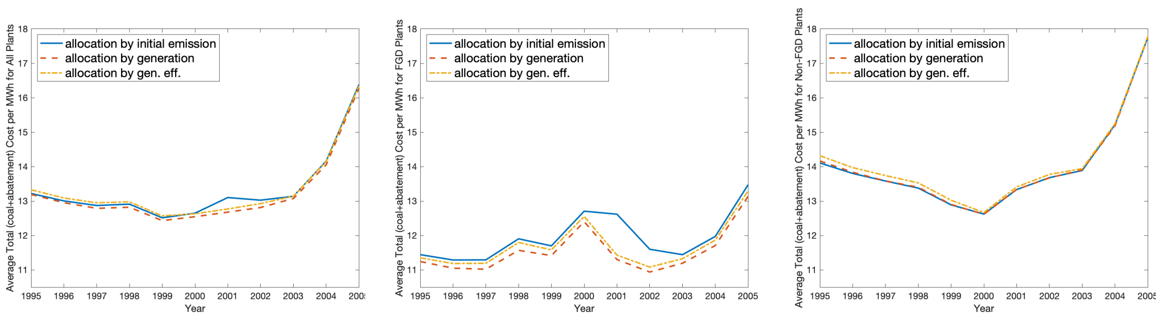
The average Btu content in columns (1)-(3) is lower than in column (4). This is because of the lower sulfur content in columns (1)-(3) than in column (4), which causes plants to lower Btu content to reduce coal costs. In all scenarios, FGD plants choose lower Btu and higher sulfur content coal than non-FGD plants. Hence, coal prices are significantly lower for FGD plants, about $23 per ton, while the prices for non-

FGD plants are about \$32 per ton. Because of the lower Btu content, the total coal consumption is greater in columns (1)-(3) than in column (4). The coal costs of FGD plants are close across the three scenarios. The coal costs of non-FGD plants are the lowest when based on initial emissions.

The allocation methods have dramatic impacts on the FGD plants' abatement rates. The coal abatement rates are 72.96%, 53.52%, and 70.35% in columns (1)-(3), which are considerably lower than the 80% rate in the data. This is due to the lower sulfur content coal utilized in the counterfactual. The choice of allocation method significantly affects total abatement costs. The FGD plants' total abatement costs are \$2.04, \$1.12, and \$1.48 billion in columns (1)-(3), respectively. FGD plants incur the lowest abatement costs in column (2). Compared with column (2), the allocation methods in columns (1) and (3) would increase the abatement costs by 82% and 32%, respectively.

The choice of allocation method substantially affects total variable costs as well. These costs for FGD plants are the lowest in column (2) because their abatement costs are the lowest in this case, due to their higher level of generation-based allocations. For non-FGD plants these costs are lowest In column (1), when allocation is based on emissions. The sum of total variable costs for both types of plants under the three methods are \$100.3, \$99.4, and \$100.3 billion, respectively, so that costs are the lowest in column (2) due to the significantly lower total variable costs of FGD plants with this allocation method. By allocating permits based on initial emissions or generation efficiency, total variable costs for both types of plants would be \$.9 billion higher than if allocation is based on generation. Hence, allocation by total plant generation is the safest method to guard against a non-operational permit trading system.

Figure 5: Average Total (Coal and Abatement) Costs under Different Allocation Methods



(a) All Plants (b) FGD Plants (c) Non-FGD Plants

Figure 5 shows the average total costs per MWh generation under the three allocation methods over time. Figure 5a shows that the average cost for all plants is the lowest in most years with the second method. The allocation methods have different impacts on the FGD and non-FGD plants. Figure 5b shows that average total costs for FGD plants are lowest with the second method. Figure 5c shows that average total costs for non-FGD plants are lowest with the first method.

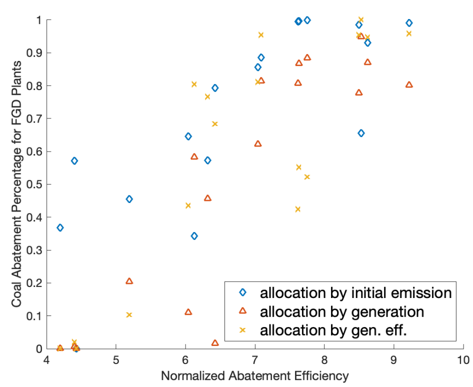


Figure 6: Abatement Percentage and $\hat{\omega}^a_{jt}$

To see the impact of the unobserved ω^a_{jt} on the plants, in Figure 6 we plot the abatement percentage against normalized, plant-level abatement efficiency.[45] The three types of markers represent the abatement percentages under the three methods. We

[45] We first compute the average $exp(\omega^a_{jt})$ for each plant across the 11 years, and we use the minimum of these averages to normalize the plants' $exp(\omega^a_{jt})$ in all years.

find that plants with higher abatement efficiencies abate higher percentages of coal. The correlations between the coal abatement percentages and normalized $exp(\omega_j^a)$ by plant are .80, .88, and .84 for the three allocation methods, respectively.

Although the allocation methods have substantial impacts on the plants, they do not affect the consumer surplus in our model. This is because the total generation of electricity and its price are the same across the allocation methods. In addition, since the total amount of permits is also the same, consumers do not experience different total levels of SO_2 emissions. While this paper does not consider the plants' revenue and thus profits, our results do imply that permit allocation affects plant's cost efficiencies. Since the FGD O&M costs include FGD labor costs, the results about the abatement costs in Table 8 indicate that FGD plants may reduce FGD labor input, when the allocation is based on generation.

We also consider a somewhat less restrictive trading system, one where plants are allowed to make only intra-state permit trades. This corresponds closely to the actual permit trading system of the ARP that evolved after 2006. See Appendix F for an extensive explanation of how we model this counterfactual. Allowing only intra-state trades restricts possible trades so much that some trading markets cannot clear when facing the same constraints on (b_{jt}, s_{jt}) as in the previous counterfactual. This result is consistent with the collapse of the permit market after 2006. However, markets can clear when the constraints are relaxed to the ranges that covers all (b, s) in the data. In this case the demand for low-sulfur coal would increase, and the allocation based on emissions would yield the lowest total variable costs. However, we believe that this scenario is highly improbable due to the likely scarcity of low-sulfur coal at current prices.

8 Conclusions

In many industries, firms act as cost minimizers subject to constraints on production of good and bad outputs, which place important restrictions on input choices. For instance, electric power plants generate two outputs (electricity and pollution) and typically face constraints on both. Since most existing approaches to estimating production functions deal with one good output and identify a single productivity term, we cannot apply them to such firms. We develop a structural model which assumes that firms minimize the costs of producing goods and controlling bads, subject to constraints on each. These firms are heterogeneous in both the productivity of the good outputs and the efficiency of controlling the bad outputs. Since these heterogeneities are correlated with input choices, our model also includes terms to measure generation productivity and abatement efficiency. By solving the cost-minimization problems, we derive the cost functions which allow identification of the production function parameters.

We apply this methodology to a balanced panel of the 76 largest US coal power plants from 1995 to 2005. In the model, plants endogenously choose the sulfur and Btu content of coal to minimize the sum of coal and pollution control costs, subject to output targets and emission constraints. While FGD plants choose between abating emissions using FGD devices and covering unabated emissions using permits, non-FGD plants can only employ the latter strategy. Assuming a Leontief production function, we solve the plants' constrained cost–minimization problems and derive their total variable cost functions, which contain the production function parameters.

Our estimation consists of three steps. The first step estimates the endogenous coal price as a function of the sulfur and Btu content. The second step estimates the abatement cost function for FGD plants. Finally, the last step uses the cost functions to estimate the production function parameters. We estimate the last two steps using GMM and deal with endogeneity using IVs. We find that the coal price increases with Btu content and decreases with sulfur content. The FGD plants have increasing marginal abatement costs and exhibit moderately increasing returns to scale, which are consistent with observed behavior. The estimated unobserved generation productivity and abatement efficiency differ substantially among plants and both improved during the sample period. The dramatic growth of the latter measure is consistent with the primary goal of the ARP.

Using the estimated model, we examine the implications of three allocation methods for SO_2 emission permits without permit trading: allocations based on emissions, total generation, and generation efficiency. This is motivated by the high transaction costs for permit trading due to a series of restrictions by the courts and EPA on trades immediately after our sample period. We find three important results. First, different permit allocation methods result in different demands for sulfur and Btu content, as well as significantly different abatement costs for the FGD plants. Second, the allocation methods have different impacts on FGD versus non-FGD plants. Third, when the permit trading system is not functioning, allocation by generation is more cost-efficient than allocation based on the other two allocation methods, with comparative reductions of \$.9 billion in total variable costs for all plants and 32%-82% in the FGD plants' abatement costs during the sample period. That is, allocation of permits based on a plant's generation is the safest hedge against a non-functioning permit system.

We also consider the case of permit trading only within states. When solving for the equilibrium state-level permit prices, we find that state-level permit markets do not clear in all states when the plants face the same constraints on (b_{jt}, s_{jt}) as in the first counterfactual. This is consistent with the collapse of the permit market after 2006. When the constraints are relaxed to the ranges of all (b, s) in the data, the permit markets clear for all states. In this case, the demand for low-sulfur coal by the plants would increase, and the allocation based on emissions would yield the lowest total variable costs. However, such a scenario is highly unlikely to occur without substantial increases in the supply of low-sulfur coal at current prices.

We can also adapt our methodology to analyze the effects of output and emission regulations for other pollutants generated by cost-minimizing plants and firms. Worldwide, hundreds of newly-constructed and existing coal-fired power plants, major polluters of SO_2 and CO_2, are governed by cap-and-trade systems. One could obtain estimates of growth in output productivity and abatement efficiency over time under these systems and run our counterfactual analyses to help determine the cost consequences of different initial allocations of permits with high transactions costs. One of the most active greenhouse gas cap-and-trade systems in the US is the California Greenhouse Gas Cap-and-Trade Program, in operation since 2013. This system's initial allocations are based primarily on emissions and to a lesser extent on generation efficiency. The EU's Emissions Trading System for CO_2 also provides substantial allocations to airlines and industry, primarily based on emissions. The initial allocation method will affect costs of production and abatement if the courts or governments within these systems erect barriers to permit trading, as has happened with the US ARP. Administrators of other systems could consider the impact of initial allocations on the cost-effective solution, in the event that future transaction costs become substantial.

References

Abito, J. M. (2020): “Measuring the welfare gains from optimal incentive regulation,” *The Review of Economic Studies*, 87(5), 2019–2048.

Ackerberg, D. A., K. Caves, and G. Frazer (2015): “Identification properties of recent production function estimators,” *Econometrica*, 83(6), 2411–2451.

Atkinson, S. E. (2019): “Panel Data in Energy Economics,” in *Panel Data Econometrics*, pp. 495–519. Elsevier.

Atkinson, S. E., and J. H. Dorfman (2005): “Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution,” *Journal of Econometrics*, 126(2), 445–468.

Atkinson, S. E., D. Primont, and M. G. Tsionas (2018): “Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions,” *Journal of Econometrics*, 204(2), 131–146.

Baltagi, B. H., and J. M. Griffin (1988): “A general index of technical change,” *Journal of political Economy*, 96(1), 20–41.

Carlson, C., D. Burtraw, M. Cropper, and K. L. Palmer (2000): “Sulfur dioxide control by electric utilities: What are the gains from trade?,” *Journal of political Economy*, 108(6), 1292–1326.

Coase, R. (1960): “The Problem of Social Cost,” *Journal of Law and Economics*, 3, 1–44.

Doraszelski, U., and J. Jaumandreu (2013): “R&D and productivity: Estimating endogenous productivity,” *Review of Economic Studies*, 80(4), 1338–1383.

Ellerman, A. D., P. L. Joskow, R. Schmalensee, E. M. Bailey, and J.-P. Montero (2000): *Markets for clean air: The US acid rain program*. Cambridge University Press.

Fowlie, M. (2010): “Emissions trading, electricity restructuring, and investment in pollution abatement,” *American Economic Review*, 100(3), 837–69.

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

Gandhi, A., S. Navarro, and D. A. Rivers (2020): “On the identification of gross output production functions,” *Journal of Political Economy, forthcoming.*

Gangadharan, L. (2000): “Transaction costs in pollution markets: an empirical study,” *Land Economics*, pp. 601–614.

Gollop, F. M., and M. J. Roberts (1983): "Environmental regulations and productivity growth: The case of fossil-fueled electric power generation," *Journal of political Economy*, 91(4), 654–674.

——— (1985): "Cost-minimizing regulation of sulfur emissions: Regional gains in electric power," *The Review of Economics and Statistics*, pp. 81–90.

Hahn, R. W. (1984): "Market power and transferable property rights," *The Quarterly Journal of Economics*, 99(4), 753–765.

Hahn, R. W., and R. N. Stavins (2011): "The effect of allowance allocations on cap-and-trade system performance," *The Journal of Law and Economics*, 54(S4), S267–S294.

Hanemann, M. (2009): "The role of emission trading in domestic climate policy," *The Energy Journal*, 30(Special Issue# 2).

Kolstad, C. D., and M. H. Turnovsky (1998): "Cost functions and nonlinear prices: estimating a technology with quality-differentiated inputs," *Review of Economics and Statistics*, 80(3), 444–453.

Levinsohn, J., and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 70(2), 317–341.

Montero, J.-P. (1998): "Marketable pollution permits with uncertainty and transaction costs," *Resource and Energy Economics*, 20(1), 27–50.

Montero, J.-P., J. M. Sanchez, and R. Katz (2002): "A market-based environmental policy experiment in Chile," *The Journal of Law and Economics*, 45(1), 267–287.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Reguant, M., and A. D. Ellerman (2008): "Grandfathering and the endowment effect: An Assessment in the context of the Spanish National Allocation Plan," *Center for Energy and Environmental Policy Research, Cambridge, Massachusetts.*

Rubin, E., M. Berkenpas, and C. Zaremsky (2007): "Integrated Environmental Control Model (IECM) with carbon capture and storage. Final Report of Contract No," Discussion paper, DE-AC21-92MC29094, Prepared by Carnegie Mellon University for US Department ....

Schmalensee, R., and R. N. Stavins (2013): "The SO 2 allowance trading system: the ironic history of a grand policy experiment," *Journal of Economic Perspectives*, 27(1), 103–22.

Stavins, R. N. (1995): "Transaction costs and tradeable permits," *Journal of Environmental Economics and Management*, 29(2), 133–148.

Wooldridge, J. M. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104(3), 112–114.

Appendix

A List of Coal-fired Power Plants and Firms in Our Sample

PLANT	FIRM	PLANT	FIRM
Barry	Alabama Power Co	Riverbend	Duke Energy Corp
Gorgas	Alabama Power Co	Muskingum River	Ohio Power Co
Colbert	Tennessee Valley Authority	W S Lee	Duke Energy Corp
Widows Creek	Tennessee Valley Authority	McMeekin	South Carolina Electric&Gas Co
Cholla	Arizona Public Service Co	Wateree	South Carolina Electric&Gas Co
Cherokee	Public Service Co of Colorado	Williams	South Carolina Electric and Gas
Comanche	Public Service Co of Colorado	Bull Run	Tennessee Valley Authority
Valmont	Public Service Co of Colorado	Cumberland	Tennessee Valley Authority
Lansing Smith	Gulf Power Co	Gallatin	Tennessee Valley Authority
Bowen	Georgia Power Co	John Sevier	Tennessee Valley Authority
Hammond	Georgia Power Co	Johnsonville	Tennessee Valley Authority
Mitchell	Georgia Power Co	Kingston	Tennessee Valley Authority
Joppa Steam	Electric Energy Inc	Carbon	PacifiCorp
Tanners Creek	Indiana Michigan Power Co	Clinch River	Appalachian Power Co
Bailly	Northern Indiana Pub Serv Co	Glen Lyn	Appalachian Power Co
Cayuga	PSI Energy Inc	Bremo Bluff	Virginia Electric & Power Co
R Gallagher	PSI Energy Inc	Chesterfield	Virginia Electric & Power Co
F B Culley	Southern Indiana Gas & Elec Co	Chesapeake	Virginia Electric & Power Co
Kapp	Interstate Power	John E Amos	Appalachian Power Co
Riverside	MidAmerican Energy Co	Kanawha River	Appalachian Power Co
LaCygne	Kansas City Power & Light Co	Philip Sporn	Central Operating Co
Big Sandy	Kentucky Power Co	Rivesville	Monongahela Power Co
E W Brown	Kentucky Utilities Co	Mt Storm	Virginia Electric & Power Co
Ghent	Kentucky Utilities Co	Pulliam	Wisconsin Public Service Corp
Green River	Kentucky Utilities Co	Weston	Wisconsin Public Service Corp
Cane Run	Louisville Gas & Electric Co	Dave Johnston	PacifiCorp
Mill Creek	Louisville Gas & Electric Co	Naughton	PacifiCorp
Paradise	Tennessee Valley Authority	James H Miller Jr	Alabama Power Co
Shawnee	Tennessee Valley Authority	R M Schahfer	Northern Indiana Pub Serv Co
Monroe	Detroit Edison Co	A B Brown	Southern Indiana Gas & Elec Co
St Clair	Detroit Edison Co	Welsh	Southwestern Electric Power Co
High Bridge	Northern States Power Co	Harrington	Southwestern Public Service Co
Asheville	Carolina Power & Light Co	Tolk	Southwestern Public Service Co
Lee	Carolina Power & Light Co	Pawnee	Public Service Co of Colorado
L V Sutton	Carolina Power & Light Co	Mountaineer	Appalachian Power Co
G G Allen	Duke Energy Corp	Belews Creek	Duke Energy Corp
Buck	Duke Energy Corp	Jim Bridger	PacifiCorp
Cliffside	Duke Energy Corp	Huntington	PacifiCorp
Dan River	Duke Energy Corp	Gen J M Gavin	Ohio Power Co
Marshall	Duke Energy Corp	North Valmy	Sierra Pacific Power Co

B Obtaining Permit Prices

We use the Kuhn-Tucker conditions of the plants' constrained cost minimization problems to obtain the permit prices. Given the constraints $s_{jt} \leq \bar{s}_{jt}$ and $b_{jt} \geq \underline{b}_{jt}$, the Lagrangian function for a non-FGD plant is

$$\mathcal{L} = w^c_{jt}(b_{jt}, s_{jt}; f_{jt})n(b_{jt}; X_{jt}, \omega^y_{jt}) + C^s_{\text{NFGD}}(b_{jt}, s_{jt}; X_{jt}, p_{jt}, \omega^y_{jt}) + \underline{\mu}^b_{jt}(\underline{b}_{jt} - b_{jt}) + \bar{\mu}^s_{jt}(s_{jt} - \bar{s}_{jt}).$$

Thus, the Kuhn-Tucker conditions are

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} n_{jt} + w^c_{jt}(b_{jt}, s_{jt}; f_{jt}) \frac{\partial n_{jt}}{\partial b_{jt}} - \frac{2 p_{jt} s_{jt} n_{jt}}{b_{jt}} - \underline{\mu}^b_{jt} = 0, \tag{B.1}$$

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} n_{jt} + 2 p_{jt} n_{jt} + \bar{\mu}^s_{jt} = 0, \tag{B.2}$$

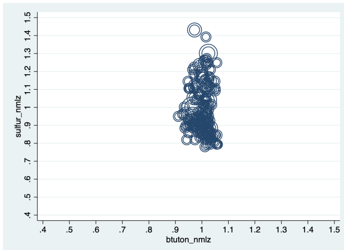
$$\underline{\mu}^b_{jt}(\underline{b}_{jt} - b_{jt}) = 0, \tag{B.3}$$

$$\bar{\mu}^s_{jt}(s_{jt} - \bar{s}_{jt}) = 0. \tag{B.4}$$

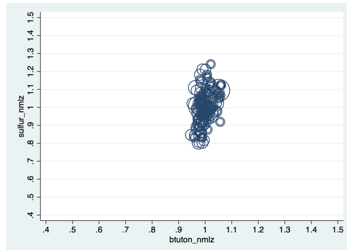
$$\underline{\mu}^b_{jt} \geq 0, \bar{\mu}^s_{jt} \geq 0. \tag{B.5}$$

To determine whether the constraints are binding, we present diagrams of normalized sulfur and Btu content for all transactions between a given plant and a given mine in a given year. We normalize the transaction-level sulfur and Btu content by dividing each variable by its respective mean among the transactions for a given {plant, mine, year}, so that the resulting variables are unit free. Figure B.1 shows the results for the largest four plants based on the number of yearly coal transactions. For each plant, we show its transactions with the largest three mines from which it purchased coal, with different years when possible. These figures show that $\underline{b}_{jt}$ is binding but that $\bar{s}_{jt}$ is not. The normalized values for b_{jt} lie on a straight line, while those for s_{jt} are highly disperse. Thus, we treat only the lower bound on the plants' Btu content as binding.

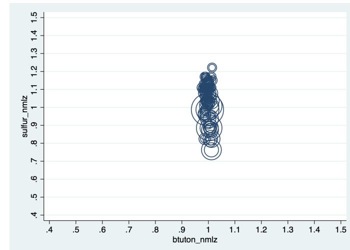
Figure B.1: Transaction-Level Sulfur and Btu Content by Plant-Mine-Year



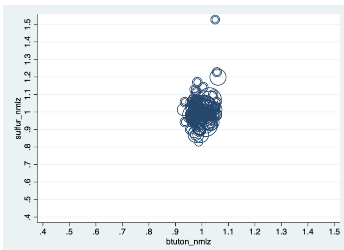
(a) Belews Creek, Mine#16, 2001



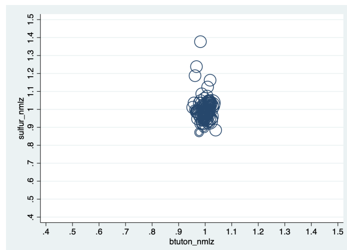
(b) Belews Creek, Mine#15, 2004



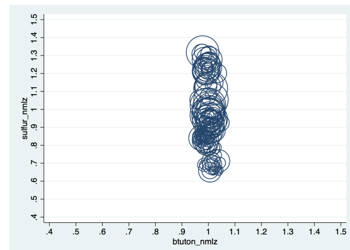
(c) Belews Creek, Mine#7, 2001



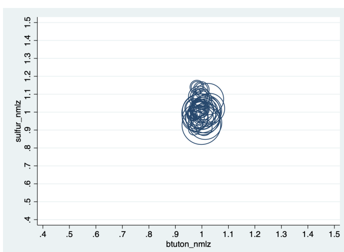
(d) Marshall, Mine#15, 2002



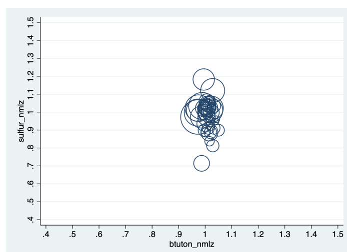
(e) Marshall, Mine#12, 2004



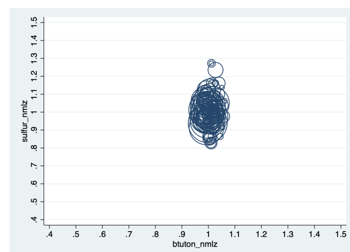
(f) Marshall, Mine#7, 1996



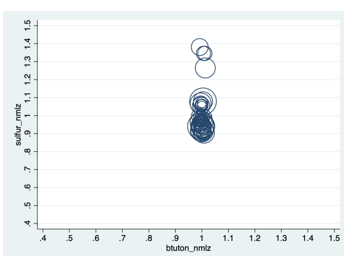
(g) Amos, Mine#13, 1998



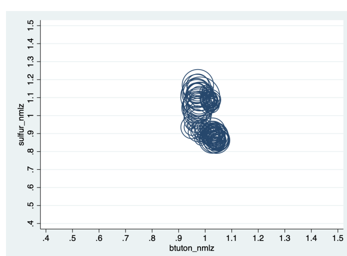
(h) Amos, Mine#14, 2003



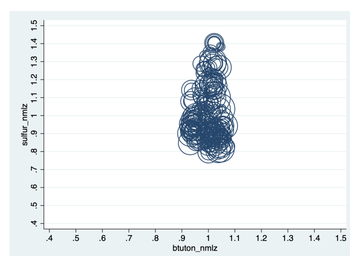
(i) Amos, Mine#12, 2004



(j) Bowen, Mine#2, 2002



(k) Bowen, Mine#5, 1995



(l) Bowen, Mine#8, 2002

Since the lower bound on b_{jt} is binding ($\underline{b}_{jt} - b_{jt} = 0$), by the Kuhn-Tucker conditions, $\underline{\mu}^b_{jt} > 0$. Because $\underline{\mu}^b_{jt}$ is unknown, we cannot use the FOC for b_{jt} to solve for p_{jt}. However, since the upper bound on s_{jt} is not binding ($s_{jt} - \bar{s}_{jt} \leq 0$), $\bar{\mu}^s_{jt} = 0$ and we can use its FOC to solve for one unknown, p_{jt}, with one equation. From equation (B.2), we use $p_{jt} = -\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{2\partial s_{jt}}$ to compute each plant-year permit price. Table B.1 shows the average estimated permit price and the EPA auction price of permits by year.

Table B.1: Average Plant-Year SO_2 Permit Prices ($/ton)

year	Estimates	EPA Auction Price
1995	99.42	130
1996	91.00	66
1997	86.97	107
1998	83.59	108
1999	80.18	201
2000	71.52	126
2001	74.12	174
2002	77.05	161
2003	77.67	172
2004	79.27	260
2005	80.66	690

Our estimates are close to the auction prices until 2001, when we begin to substantially underestimate actual prices. This occurs because from 2001 to 2005, a number of unforeseen supply disruptions pushed permit prices to very high levels. These included hurricanes, Powder River Basin train derailments, the CAIR, and speculation, none of which our model captures.

C Removal Rate of FGD Devices

The removal rate, r_j, is the percentage of SO_2 that is abated, measured as "percent removal of SO_2 at 100% generation load", as stated in the variable definitions on the Directory of Form 767 files of the DOE EIA. Thus, r_j is baked into the FGD when it is manufactured and any extremely small variations from this number are due to measurement or reporting error that is uncorrelated with r_j. Table C.1 shows the variation in r_j within plants from 1995 to 2005 for the 16 FGD plants. The removal rates did not change during the 11 years for 11 FGD plants. For the other five plants, the standard deviations are extremely small. For these reasons, we use r_j as an IV.

Table C.1: Removal Rate of FGD Plants

plantid	mean	std. dev.
50	0.817	0.002
113	0.883	0.004
1012	0.950	0
1241	0.800	0
1356	0.950	0
1363	0.866	0.002
1364	0.900	0
1378	0.842	0
3399	0.950	0
4158	0.570	0
4162	0.700	0
6085	0.900	0
6137	0.875	0.001
8066	0.875	0.001
8069	0.800	0
8102	0.950	0

D Bootstrapping

Our estimation of the model consists of three stages. The first stage estimates the hedonic price function, equation (20), the second stage estimates the abatement cost function, equation (25), and the third stage estimates the log of the production function, equation (31). From the first stage, we obtain estimates of p_{jt} which feed into stages two and three. However, the estimated parameters from stage two do not feed into stage three, so we do not need to bootstrap stages two and three together. Ideally, we would like to bootstrap stages one and two together and stages one and three together.

There are two generally available bootstrap methods, the pairs and the wild. The wild requires that the fitted right-hand side of the original equation is held constant during all bootstrap replications. The only source of variation from one draw to the next comes from randomly drawing the residuals, multiplying them by a Rademacher random variable, and then adding this to the fitted right-hand side to get a new bootstrap value for the left-hand-side variable. Thus, this method is not available to bootstrap stage one jointly with stage two, since the fitted right-hand side of stage two will vary with each draw of the wild bootstrap, due to the changing values of $\hat{p}_{jt}$ generated from the first stage.

The pairs method also is not available to estimate stage one jointly with stage two using the disaggregated plant-mine transaction-level data. The data for the first stage includes FGD and non-FGD plants, while the second stage uses only the FGD plants. Randomly drawing a set of plants and using all of their plant-mine transactions in the first stage, clustering on plant ids, means that a fixed sample size of FGD plants cannot be guaranteed for the second stage on each bootstrap replication. That is, in one replication all 16 FGD plants may be drawn in the first step and in the next replication only 13 FGD plants may be drawn in this step. The second stage will inherit different numbers of FGD plants. Further, a fixed sample size cannot be guaranteed for the first step on each bootstrap replication, since the number of transactions differs across plants. If the sample size is not constant for each replication of the bootstrap in the estimation of each step, the pairs bootstrap is invalid.

We turn now to the feasibility of bootstrapping stage one jointly with stage three. We cannot use the wild method, since the fitted right-hand side will not remain constant across bootstrap draws for the reasons given above. Further, we cannot employ the pairs with the disaggregated plant-mine transactions data, since the number of transactions varies across plants. This would cause the sample size for the estimation of stage one, clustering on plant id's, to vary with each bootstrap replication, again invalidating the pairs method. However, we can use the pairs method to bootstrap stage one jointly with stage three if we first aggregate our plant-mine transaction data to the plant level before estimating stage one. Now with this balanced panel, the set of firms and years for both steps would be identical and the bootstrap sample size would be the same on each bootstrap replication. Before beginning the pairs bootstrap, the data sets for both equations must first be incorporated into one large data set. We randomly select a set of plants, with replacement, before each bootstrap replication begins, clustering on the plant id's, so that all years for each chosen plant are in each bootstrap sample. We perform 100 bootstrap replications. Computed bootstrap standard errors are reported in Tables 2 and 5.

The following is our procedure for aggregating the transaction-level data for (b, s, w^c) to plant-year data, $(b_{jt}, s_{jt}, w^c_{jt})$, in step one. To obtain aggregate data for b and s we compute weighted averages, where weights are the coal quantity. To obtain aggregate data for the price of coal, we first estimate the coal price function in (20). We then obtain a "net coal price", which is free of transportation costs, by subtracting from w^c at the transaction-level all the terms involving transaction-level dummy variables, (d^q, d^τ, d^m, d^j). We make this adjustment before aggregating, because averages of these dummies are not meaningful. From this net coal price, we aggregate the transaction-level w^c weighted by the coal quantity to obtain the plant-year averages, w^c_{jt}.

E Counterfactual Permit Allocation Methods

Table E.1 shows the total number of permits for the FGD plants by year for the three counterfactual allocation methods. Among the three allocation methods, FGD plants' total permits are the lowest in column (1) after 2000, because their emissions fell dramatically due to high permit prices caused by supply disruptions and speculation in the permit trading market, as already discussed.

Table E.1: Total Permits of FGD Plants under Counterfactual Scenarios by Year (10^6)

	(1) allocation by emissions w/o trading	(2) allocation by generation w/o trading	(3) allocation by $exp(\omega^y)$ w/o trading
1995	0.68	1.00	0.68
1996	0.73	1.05	0.74
1997	0.71	1.07	0.76
1998	0.69	1.04	0.74
1999	0.70	0.98	0.69
2000	0.65	0.95	0.66
2001	0.56	0.88	0.60
2002	0.45	0.82	0.59
2003	0.48	0.84	0.60
2004	0.45	0.85	0.59
2005	0.46	0.86	0.60

F Impacts of Permit Allocation Methods under Within-State Trading

Instead of assuming a completely non-functioning permit market, we now consider the scenario where only within-state trading is allowed, which ultimately occurred with the ARP. The three allocation methods are the same as in section 7. The within-state trading implies that the plants in the same state face the permit price, p_{st}, where the subscript s means the state. In this case, the non-FGD plants choose the optimal s_{jt} to minimize the total cost of coal and permits as in the model. The FGD plants choose not only the optimal s_{jt} but also the optimal abatement level given the permit price it faces. In this analysis, we fix the Btu content to be the same as in the data.[46] Thus,

[46]We make this assumption for two reasons. First, adding b_{jt} as an endogenous choice variable increases the dimension and difficulty of solving the constrained cost-minimization problem and the state-level permit market clearing condition simultaneously. Second, given the small standard deviation in b_{jt} within a plant and the binding constraints for $\underline{b}_{jt}$, it is reasonable to assume that b_{jt} is the same as in the data.

the cost-minimization problem for non-FGD plants is:

$$\min_{s_{jt}} C_{NFGD}(s_{jt}; b_{jt}, p_{st}, X_{jt}, \omega^y_{jt}) = \min_{s_{jt}} \{ (w^c_{jt}(s_{jt}; b_{jt}, f_{jt}) + 2p_{st}s_{jt}) \frac{h(X_{jt}, \omega^y_{jt})}{b_{jt}} \},$$

subject to constraints on s_{jt}, so it is restricted to the range of s_{jt} for each plant in all years of the data.[47] Given s_{jt}, the quantity of SO_2 generated/emitted is $2p_{st}s_{jt}n_{jt}$ where $n_{jt} = \frac{h_{jt}}{b_{jt}}$. Let m^a_{jt} be the amount of permits allocated to plant j in year t and m^b_{jt} be the amount of SO_2 permits that plant j needs to purchase in order to cover all of its emissions. We have the following equation:

$$m^b_{jt} = 2s_{jt}n_{jt} - m^a_{jt},$$

where m^b_{jt} can be negative if the plant sells excess permits.

If we restrict the choice of sulfur content to be within two standard deviations of the plant-level average, then the market clearing condition of permits does not hold for many states. This is consistent with the effective collapse of the permit trading system after the implementation of the CSAPR in 2010 which allowed only intra-state trading. In order to satisfy the market-clearing condition for permits, we now allow a much wider range of sulfur levels which are limited by the range of sulfur in all years of the data for all plants.

The problem for an FGD plant is also similar to the problem in Section 4. The plant solves a two-stage optimization problem. For any choice of s_{jt}, the plants chooses the optimal abatement level of n^a_{jt}. The cost-minimization problem is given by

$$\begin{aligned} \min_{s_{jt}} C_{FGD}(s_{jt}; b_{jt}, p_{st}, X_{jt}, \omega^y_{jt}, \omega^a_{jt}) = \min_{s_{jt}} \Big\{ & w^c_{jt}(b_{jt}, s_{jt}) \frac{h(X_{jt}, \omega^y_{jt})}{b_{jt}} \\ & + \min_{n^a_{jt}} \big[e^{\lambda_0 - \omega^a_{jt}} (n^a_{jt} s_{jt} r_j)^\lambda + 2p_{st}s_{jt}(n_{jt} - n^a_{jt} r_j) \big] \Big\}, \end{aligned}$$

[47] We fix the shipping charges at their estimated values, since substantial flexibility exists in the choice of s_{jt} for a given shipping charge, as described in Section 4.2.

subject to the constraints on s_{jt}. These plants would sell permits and purchase higher-sulfur-content coal if the marginal abatement cost were lower than the permit price. The amount of permits a plant needs to purchase is the difference between the SO_2 emission after abating the sulfur and the allocated permits. That is,

$$m^b_{jt} = 2s_{jt}(n_{jt} - n^a_{jt} r_j) - m^a_{jt}.$$

The plants within the same state can trade permits with each other. Their demand for permits and the state-wide aggregate allocated permits will determine the state-level permit price, p_{st}. In equilibrium, all plants' marginal costs of abating SO_2 will be equal to the state-level permit price, and the market for permits will clear. Denote the set of plants in state s by $\Omega_s = \{1, 2, ..., J_s\}$. A state may have both FGD and non-FGD plants. The market clearing condition for permits in state s is

$$\sum_{j \in \Omega_s} m^b_{jt} = 0.$$

For each state, we solve for the equilibrium permit price (p_{st}) and the optimal sulfur content (s_{jt}) for the plants. The equations we use are the market clearing conditions for the state and the FOCs for the plants. The number of unknowns is equal to the number of equations, $(1 + J_s)$.

We solve for each plants' optimal s_{jt} and the permit price in each state under the three scenarios and compare the aggregate and average costs across the three allocation methods. Columns (1) to (3) in Table 8 show the results for the three counterfactual scenarios. The average sulfur content of FGD plants weighted by generation is the lowest in column (3) at 0.70%. This is because they receive the least amount of permits when allocation is based on the generation efficiency, as shown in Table 7. Due to the low sulfur content, the average coal price and coal costs are the highest in column (3). The permit costs include the costs of buying permits and the opportunity cost of using allocated permits. Therefore, permit costs depend on the permit price, the initial allocation, and the emission level of the plants. We find that permit costs in column (2) are highest for FGD plants among the three counterfactual scenarios, due to the highest sulfur content and lowest abatement percentage.

Table F.1: Impacts of Different Emission Permit Allocation Mechanisms — Intra-State Trading Only

	(1) allocation by emissions	(2) allocation by generation	(3) allocation by $exp(\omega^y)$	(4) allocation in data
Average sulfur content (%)	0.84	0.82	0.81	1.24
– FGD	0.79	0.81	0.70	2.04
– Non-FGD	0.87	0.82	0.87	0.87
Average Btu content (10^6/ton)	22.44	22.44	22.44	22.44
– FGD	21.78	21.78	21.78	21.78
– Non-FGD	22.76	22.76	22.76	22.76
Average coal price ($/ton)	31.03	31.11	31.14	30.40
– FGD	26.53	26.52	26.76	24.71
– Non-FGD	33.12	33.24	33.18	33.05
Coal costs all years ($\$10^{10}$)	10.20	10.22	10.24	9.70
– FGD	2.94	2.94	2.96	2.67
– Non-FGD	7.26	7.29	7.27	7.04
Abatement costs (FGD) all years ($\$10^9$)	0.18	0.17	0.12	2.28
Average coal abatement (%)	85.78	76.16	80.02	80.00
Permit costs all years ($\$10^9$)	3.85	3.77	3.63	3.93
– FGD	0.36	0.45	0.38	0.34
– Non-FGD	3.49	3.32	3.25	3.41
Total variable costs all years ($\$10^{10}$)	10.61	10.62	10.62	10.32
– FGD	2.99	3.00	3.01	2.94
– Non-FGD	7.61	7.62	7.60	7.38
Average variable cost per MWh ($)	14.11	14.13	14.13	16.84
– FGD	12.46	12.48	12.53	17.14
– Non-FGD	14.89	14.90	14.87	19.84

The sulfur content and coal price are averages for all the plants in all years weighted by generation. The coal costs, abatement costs, and total costs are the total values of all plants from 1995 to 2005. Coal abatement is the average coal abatement percentage weighted by generation of the FGD plants. The total variable costs include the permit costs, abatement costs, and coal costs. The total variable costs are measured in 2005 dollars with the firm yield as the discount rate.

The average sulfur content of non-FGD plants is the lowest in column (2), at 0.82%, because they receive the least amount of permits when allocation is based on generation. Accordingly, the coal price and coal costs of non-FGD plants are the highest in column (2). The average coal prices for the non-FGD plants are higher than the FGD plants in all columns. This is due to higher Btu content choices of the non-FGD plants. Their sulfur content (and thus the total amount of sulfur in coal) and emissions of SO_2 are the lowest in column (2). The permit costs for non-FGD plants are the lowest in column (3).

The allocation methods have dramatic impacts on the FGD plants' abatement rate and abatement costs. The coal abatement rates are 85.78%, 76.16%, and 80.02% in columns (1)-(3). The FGD plants' total abatement costs are \$0.18, \$0.17, and \$0.12 billion, respectively. These plants incur the lowest abatement costs in column (3) because their sulfur content is the lowest in this case. Thus, compared with column (3), the allocation methods in columns (1) and (2) increase the abatement costs by 133% and 42%, respectively. For FGD plants, the total variable costs are marginally lowest in columns (1) and (2) when the allocation is based on emissions and generation. The average total cost per MWh is lowest in column (1) for FGD plants and overall for both types of plants.

Compared with columns (2) and (3), the total variable costs in column (1) is \$.1 billion lower, and the average variable costs per MWh in column (1) are \$.02 lower for all the plants. For FGD plants, allocation based on emissions is slightly less costly in terms of total and average variable costs. For non-FGD plants, allocation based on generation efficiency is slightly less expensive in terms of total and average variable costs.