Estimation of Production Technologies with Output and Environmental Constraints

Scott E. Atkinson* Rong Luo†

University of Georgia *Renmin University of China*

September 21, 2020

Abstract

Plants in many industries minimize costs subject to constraints on output and emissions, which in turn restrict input choices. We develop a novel cost function approach to estimating their production function parameters, including generation productivity and pollution abatement efficiency terms to measure unobserved heterogeneity. Applying this this technique to a panel of US coal-fired power plants, we model the firm's trade-off of sulfur and heat content to minimize the total costs of coal and pollution control. We find substantial heterogeneity and a marked increase in abatement productivity over time, consistent with the goals of the Acid Rain cap-and-trade program. Counterfactual analysis shows that its initial allocation of emission permits substantially affected costs, due to severe restrictions on permit trading.

KEYWORDS: Cost minimization, Output and Emission Constraints, Transaction Costs, Independence Property, Generation Productivity, Abatement Efficiency.

JEL CODES: D24, L11, L94

*Email: atknsn@uga.edu.

†Email: luorongecon@163.com, corresponding author.

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. 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 must take output orders from parent utilities. None of these power plants can flexibly choose output to maximize profits. 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. However, 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 to produce goods and bads, due to the trade-off between the added expenses of cleaner inputs and pollution abatement.

Coal-fired power plants operate as cost minimizers subject to output and emission constraints. To model these plants, we develop a structural cost function approach to estimating their production technologies. Plants choose between the level 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 reflected in permit prices. 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.

We consider three options for plants that generate more SO_2 than allowed by the initial permit allocation. 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 our econometric model.

We apply this model to a balanced panel of the 80 largest US coal-fired power plants during the 1995-2005 period, where substantial variation exists in production productivity and abatement efficiency. 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.

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), Doraszelski and Jaumandreu (2013), and Gandhi, Navarro, and Rivers (2020)) incorporates an output productivity term and assumes that it follows an AR(1) process for persistence over time. 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 AR(1). The total cost functions depend on the parameters of the Cobb-Douglas production function and the abatement cost function. However, 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 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 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).[2] Third, the estimated unobserved generation productivities and abatement efficiencies vary substantially across plants. Lastly, the two efficiency measures improved substantially during the sample period. In particular, the plants generated 1.7% more electricity in 2005 compared with 1995 conditional on the same inputs, and the FGD O&M cost to remove a unit of SO_2 decreased by 55% from 1995 to

[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.

[2]The sum of the Cobb-Douglas coefficients = 1.06.

2005 on average. The latter 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 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 counterfactual assumes that the spot trading system is non-functioning and evaluates the cost implications of initial permit allocation schemes based on 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 low transaction costs.

We obtain three results which are consistent with the Coase Theorem, but of magnitudes greater than expected. First, each permit allocation method produces significant differences in the demand for sulfur, Btu content, and the abatement costs of FGD plants. Second, 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, costs for non-FGD plants are lowest. Third, allocation based on generation reduces total costs by \$0.4-\$0.6 billion and abatement costs by 35%-101% compared with other allocation methods. We are unaware of any study that estimates the impact of different initial allocation systems on plants' costs under substantial trade impediments. These results should encourage policy makers to carefully examine the cost implications of different initial allocation methods under cap-and-trade systems for coal-fired power plants recently adopted worldwide to control SO_2, nitrogen dioxide (NO_x), and in far greater numbers, carbon dioxide (CO_2), under the counterfactual scenario of a substantial increase in transaction costs.

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

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 factors (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.[3] Regardless, all plants face production decisions made by parent firms, so that they cost minimize subject to constraints on output and emissions.

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

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.[4] By 2009, the 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.[5] However, Reguant and Ellerman (2008), Fowlie and Perloff (2013), and Hahn and Stavins (2011) fail to find significant evidence against the independence property.[6] As an alternative, our counterfactual examines a market where transaction costs are extremely high, namely the ARP after 2009, and estimates the cost effects of different

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

[5]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.

[6]Counter to these results is the study by Fowlie (2010) 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.

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 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.[7]

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 change 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

[7] 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.

data during the period when the ARP permit market was fully functional. Since the Btu/sulfur trade-off is correlated with productivity/efficiency components of the error terms, we mitigate this endogeneity by including terms to measure these effects. Our finding of substantial growth in abatement efficiency over time (which is consistent with the incentives of the ARP) suggests that our methodology has important applications in determining whether the numerous cap-and-trade systems adopted worldwide have generated similar improvements in abatement efficiency over time. Our counterfactual, which finds substantial cost consequences of different permit allocation methods without trading, uses our estimated optimal trade-off between fuel-switching, use of permits, and abatement.

3 Data

We employ a balanced panel of the 80 largest coal-fired power plants (without entry or exit to this set) in the US from 1995 to 2005.[8] 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. Inputs for the coal power plants consist of capital, labor, Btu, and sulfur, which in combination produce electricity in megawatt-hours (mWh) and SO_2. We measure capital as megawatt (MW) generating capacity of the plant, which we adjust as the plant augments existing capacity or reduces it. Labor is the number of full-time employees plus one-half the number of part-time employees. The sulfur content of the coal is the percentage of sulfur in coal. While the power plants in our sample consume coal and either oil or natural gas, 99% of the Btus for each plant comes from coal consumption on average.

We obtained the data from several sources. The Federal Energy Regulatory Commission (FERC) Form 1 provides labor and capital data for private electric power plants,

[8] See the Appendix for our plants and parent utilities.

and the Energy Information Administration (EIA) EIA-412 survey is the source of this data for public power plants. While the US Department of 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 mWh generation by plant. The EPA collected the SO_2 emissions data at the plant level 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, the percent of total plant MW capacity that has installed FGD devices, and the O&M costs of FGD devices. These costs include the variable cost of the feed materials and chemicals, waste disposal, and other costs.

Table 1 shows the summary statistics of the plant-level annual data. Among the 80 plants, only 18 plants employ FGD units throughout the 1995-2005 period. The other 62 plants have either never installed FGD units or installed them for only part of this period. Panel A shows the outputs and inputs of the two types of plants. The median annual generation of an FGD plant is 7.66 million mWh, while that of a non-FGD plant

[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.

is 4.42 million mWh. The median annual SO_2 emission of an FGD plant is about 28,209 tons, while that of a non-FGD plant is 21,641 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 1.48% for FGD plants, but only 0.89% for non-FGD plants, while the median Btu content is 22.52 million/ton for FGD plants and 24.14 million/ton for non-FGD plants.

Table 1: Data Summary Statistics for Plants

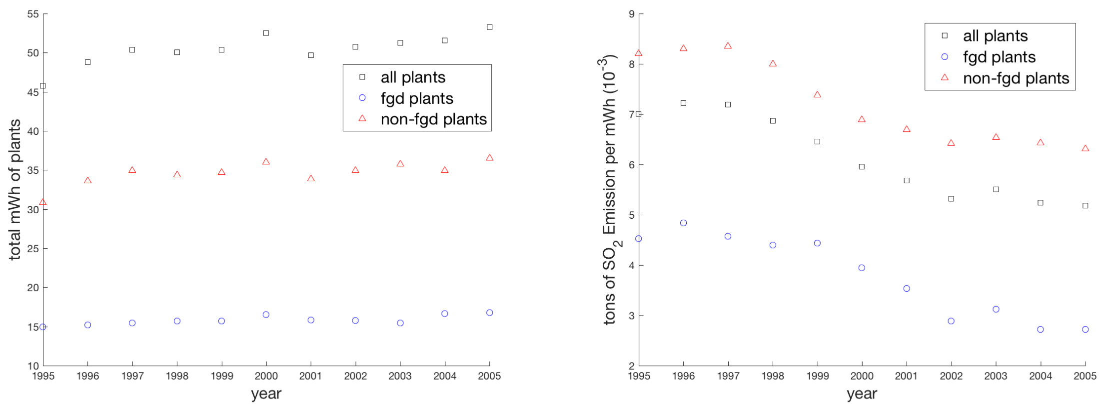
	FGD			Non-FGD		
	median	min	max	median	min	max
A. Outputs and inputs						
Generation (10^6 mWh)	7.66	1.86	20.32	4.42	0.12	22.33
SO_2 emission (tons)	28,209	631	186,470	21,651	3,242	212,377
Generation capacity (mW)	1,620	411	2,600	772	110	3,498
Labor (# employees)	212	64.634	538	129	23.744	580
Coal (10^6 tons)	4.00	0.87	9.14	1.97	0.61	12.31
Sulfur (%)	1.48	0.33	3.95	0.89	0.13	3.79
Btu (10^6/ton)	22.52	15.45	24.64	24.14	16.21	26.35
B. Input prices						
Yield (%)	7.55	5.38	8.95	7.54	5.38	9.77
Wage (10^4\$)	4.36	2.49	9.47	4.34	2.68	8.32
Coal price (\$/ton)	25.62	9.47	53.35	38.07	11.38	141.48
C. Abatement						
Coal abatement (10^6 tons)	2.80	0.12	8.59			
SO_2 removal rate (%)	84.64	57.00	95.00			
FGD O&M costs (10^3\$)	3,793	300	30,015			
O&M costs/(O&M + coal costs) (%)	7.76	0.25	31.07			
Observations	198	198	198	682	682	682

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 \$25.62/ton for FGD plants and \$38.07/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 2.80 million tons, which is less than the coal input, implies that the plants do not abate all coal. The sulfur removal rate represents the fraction of SO_2 abated by the FGD devices, which is a characteristic of the FGD devices. This removal rate varies significantly across plants, with a median of 84.64%. The median annual O&M cost is $3.79 million, while the average share of O&M cost in the plants' total costs of coal and FGD O&M is 7.76%, with a maximum of 31.07%.

From 1995 to 2005, the plant-level heat input from coal has increased slightly, with a mean annual growth rate of 0.96%. The plant-level electricity generation growth rate is close to the heat growth rate. The average electricity generation among the plants in 10^6 mWh increased from 5.71 in 1995 to 6.65 in 2005. The average sulfur and Btu content of all 80 coal-fired power plants has decreased. Generation capital has been slowly increasing for all plants, with an average annual growth rate of 3.66%. The quantity of labor used in generation has been decreasing for all plants, with an annual growth rate of -4.06%.

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. The non-FGD plants generate about twice as much electricity as their FGD counterparts due to a large number of non-FGD plants. Figure 1b plots the average SO_2 emission per mWh generation of the plants. We represent the 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

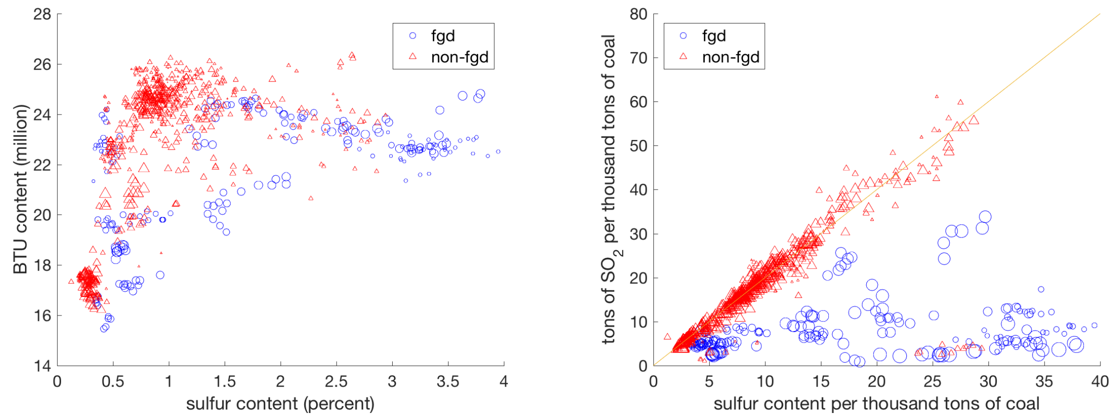


(a) Total generation by plant type (b) SO_2 Emission per mWh Generation

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. To a substantial extent, reserves exist in the lower-triangular portion of this graph.[11]

[11]For example, surface and underground mines in West Virginia have large proven coal reserves with average sulfur content ranging from .52 to 3.4 percent or greater in combination with 26 or greater MMBtu per ton. Surface mines in Wyoming have large proven reserves with average sulfur content ranging from .34 to 2.13 percent in combination with 15 to 23 MMBtu per ton. Underground mines in Utah have two basic combinations of sulfur and Btu content. They have large proven reserves of low-sulfur/high-Btu coal with average sulfur content ranging from .46 to .52 in combination with 23 to 26 MMBtu per ton and lesser (but still substantial) reserves with average sulfur content of .52 or greater and 26 MMBtu per ton or greater. They also have substantial reserves of higher-sulfur/lower-Btu coal with average sulfur content of 2.7 percent in combination with 20 to 23 MMBtu per ton.

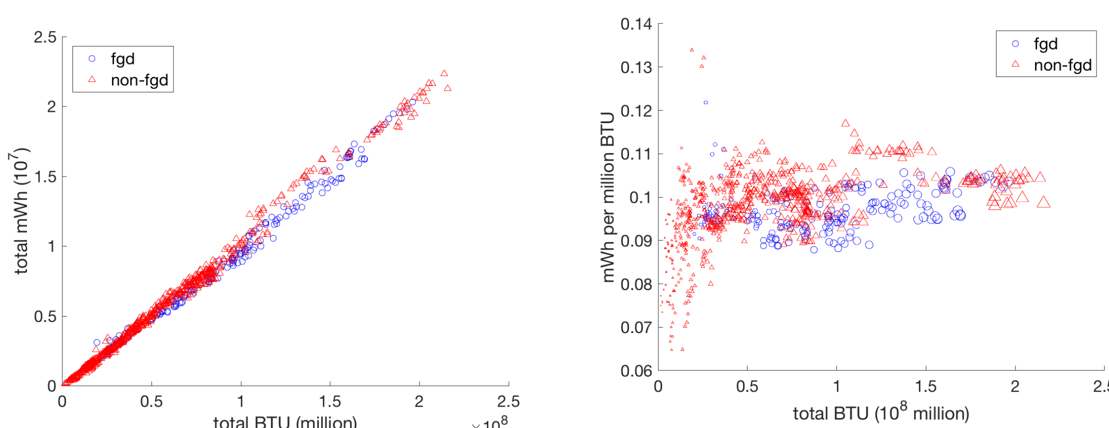
Figure 2: Sulfur Content, Btu Content, and SO_2 Emission



(a) Btu Content and Sulfur Content (b) SO_2 Emission and Sulfur Content

In Figure 2b, we plot the amount of sulfur per thousand tons of coal input against the tons of SO_2 emissions per thousand tons of coal. For the non-FGD plants, a linear relationship exists between the two variables with a slope of approximately 2, as expected from the chemistry of converting sulfur into SO_2. 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.

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 a 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 ranges of Btu and sulfur content. The total variable cost is the sum of the costs of coal and pollution control. FGD plants differ from non-FGD plants in that they have the option to abate SO_2. Each plant faces trade-offs when choosing its sulfur and Btu content, which affects 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.

4.1 Costs of Coal

In year t, plant j first observes its generation capacity, k_{jt}, labor stock, l_{jt}, generation productivity, ω_{jt}^y, 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.

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

$$y_{jt} = e^{(\beta_0+\omega^y_{jt})} k_{jt}^{\beta_k} l_{jt}^{\beta_l} h_{jt}^{\beta_h}, \tag{1}$$

where $(\beta_0, \beta_l, \beta_k, \beta_h)$ are parameters. The term, ω^y_{jt}, captures the plants' heterogeneity due to differences in generation productivity, which are unobserved in the data. Given the production function in (1), the heat needed to produce y_{jt} for a given $(l_{jt}, k_{jt}, \omega^y_{jt})$ is

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

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

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

$$n(b_{jt}; X_{jt}, \omega^y_{jt}) = \frac{h_{jt}(X_{jt}, \omega^y_{jt})}{b_{jt}} = (y_{jt} e^{-(\beta_0+\omega^y_{jt})} l_{jt}^{-\beta_l} k_{jt}^{-\beta_k})^{\frac{1}{\beta_h}} b_{jt}^{-1}. \tag{3}$$

A higher b_{jt} implies that a lower n_{jt} will produce a given y_{jt}. Denote the price of coal 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, $w^c_{jt} = w^c_{jt}(b_{jt}, s_{jt}, f_{jt})$. 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 only 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}:

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

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 its total coal cost and permit cost by choosing b_{jt} and s_{jt}. The plant may face constraints on (b_{jt}, s_{jt}), which can differ across plants. For example, if the plant has signed long-term contracts over time with some coal mines, it can only partially adjust the coal characteristics in a given year. Denote the ranges of (b_{jt}, s_{jt}) by $\left[\underline{b}_{jt}, \bar{b}_{jt}\right]$ and $\left[\underline{s}_{jt}, \bar{s}_{jt}\right]$, where the underbars and overbars indicate lower and upper bounds. A plant chooses (b_{jt}, s_{jt}) to minimize the total variable cost of coal and permits:

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

subject to $b_{jt} \in \left[\underline{b}_{jt}, \bar{b}_{jt}\right]$ and $s_{jt} \in \left[\underline{s}_{jt}, \bar{s}_{jt}\right]$.

We model the optimal choice of (b_{jt}, s_{jt}), conditional on the shipping cost f_{jt}.[12] 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 is reasonable because a plant can purchase large amounts of coal with quite different b_{jt} and s_{jt} either at a given mine or at adjacent mines. The report DOE (1999) for coal reserves and 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 both available in reserves and 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 cost, especially if the distances are roughly the same. From Figure 2a and the information on the reserves of coal, we assume that each plant faces a rectangular choice set of b_{jt} and s_{jt} per year.[13]

Using the coal cost in equation (4) and the permit cost in equation (5), the FOCs for b_{jt} and s_{jt} are

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} n_{jt} + w^c_{jt}(b_{jt}, s_{jt}; f_{jt}) \frac{\partial n_{jt}}{\partial b_{jt}} - \frac{2 p_{jt} s_{jt} n_{jt}}{b_{jt}} - \underline{\mu}^b_{jt} + \bar{\mu}^b_{jt} = 0, \tag{6}$$

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} n_{jt} + 2 p_{jt} n_{jt} - \underline{\mu}^s_{jt} + \bar{\mu}^s_{jt} = 0, \tag{7}$$

where $(\underline{\mu}^b_{jt}, \bar{\mu}^b_{jt})$ and $(\underline{\mu}^s_{jt}, \bar{\mu}^s_{jt})$ are the plant-year Lagrangian multipliers for lower- and upper-bound constraints for b_{jt} and s_{jt}, respectively.[14] 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 (6) gives the increased cost of coal (equal to its

[12] 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.

[13] We assume that the ranges 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 parametric functional form, and no data on the bounds are available.

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

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. From equation (7), 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. For a given level of b_{jt}, the plant would increase s_{jt} until the cost reduction equals the permit expenditure per unit of s_{jt}.

Using the coal cost and the permit cost expressions, the FOC for s_{jt} can be rewritten to show the relationship between the change in the price of coal due to changes in s_{jt} and p_{jt} after dividing by n_{jt}:

$$\frac{\partial w_{jt}^c}{\partial s_{jt}} + 2p_{jt} - \frac{\underline{\mu}_{jt}^s}{n_{jt}} + \frac{\bar{\mu}_{jt}^s}{n_{jt}} = 0. \tag{8}$$

Denote the optimal b_{jt} and s_{jt} by (b_{jt}^*, s_{jt}^*). The total variable cost function for the non-FGD plant is

$$C_{NFGD}(X_{jt}, \omega_{jt}^y; f_{jt}, p_{jt}) = (w_{jt}^c(b_{jt}^*, s_{jt}^*; f_{jt}) + 2p_{jt}s_{jt}^*)\frac{h_{jt}(X_{jt}, \omega_{jt}^y)}{b_{jt}^*}. \tag{9}$$

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 the expenditures on abating SO_2, the cost of purchasing SO_2 permits, and the opportunity cost of using allocated permits. An FGD plant first chooses (b_{jt}, s_{jt}) and then chooses the amount of coal to abate with the FGD devices, n_{jt}^a. We can perform this analysis in this order because the choice of n_{jt}^a depends on b_{jt} and s_{jt}. We first describe, conditional the choice of (b_{jt}, s_{jt}), how the plants choose n_{jt}^a to minimize the pollution control costs.

We then solve the plants' optimization problem of minimizing the total variable cost by choosing $(b_{jt},\, s_{jt})$.

4.3.1 Abatement Costs

Denote the FGD units' sulfur removal rate by $r_{jt} \in [0,1]$, which represents the fraction of SO_2 abated by the FGD devices. The characteristics of the FGD devices determine this rate, so plants do not endogenously choose it. The FGD plants can have different sulfur removal rates. Given the sulfur content and the amount of coal used in FGD units, the amount of sulfur scrubbed by the plant is $n^a_{jt} r_{jt} s_{jt}$.

The plants can have different O&M costs to abate the same amount of sulfur. Denote the unobserved abatement efficiency by ω^a_{jt}. Because the O&M cost includes the cost of the feed materials and chemicals, waste disposal, and other costs, ω^a_{jt} represents the heterogeneity in these costs across plants. 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, $n^a_{jt} r_{jt} s_{jt}$, and the efficiency, ω^a_{jt}. Denote the total abatement cost function by $C^a(n^a_{jt}, s_{jt}, r_{jt}, \omega^a_{jt})$. We assume that abatement cost is a power function of the total amount of sulfur scrubbed:[15]

$$C^a(n^a_{jt}, s_{jt}, r_{jt}, \omega^a_{jt}) = e^{\lambda_0 - \omega^a_{jt}} (n^a_{jt} s_{jt} r_{jt})^{\lambda}. \tag{10}$$

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, $s_{jt}(n_{jt} - n^a_{jt} r_{jt})$, will be converted to $S^e_{jt} = 2 s_{jt}(n_{jt} - n^a_{jt} r_{jt})$ tons of SO_2. The cost of buying emission permits

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

and holding allocated ones to cover unabated emissions is

$$p_{jt}S^{e}_{jt} = 2p_{jt}s_{jt}(n_{jt} - n^{a}_{jt}r_{jt}). \tag{11}$$

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

$$\min_{n^{a}_{jt}} \left\{ e^{\lambda_0 - \omega^{a}_{jt}} (n^{a}_{jt}s_{jt}r_{jt})^{\lambda} + 2p_{jt}s_{jt}(n_{jt} - n^{a}_{jt}r_{jt}) \right\}. \tag{12}$$

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:

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

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

$$n^{a*}_{jt} = \left(\frac{2p_{jt}e^{\omega^{a}_{jt} - \lambda_0}}{\lambda} \right)^{\frac{1}{\lambda-1}} \frac{1}{s_{jt}r_{jt}}. \tag{13}$$

It increases with ω^{a}_{jt} and p_{jt} but decreases with r_{jt}, s_{jt}, and λ. Plugging n^{a*}_{jt} into the pollution control cost equation (12), we obtain the minimized pollution control cost function, conditional on (b_{jt}, s_{jt}). Denote the pollution control cost of an FGD plant by C^{s}_{FGD}. We can write this cost as

$$C^{s}_{\text{FGD}}(b_{jt}, s_{jt}; r_{jt}, p_{jt}, X_{jt}, \omega^{y}_{jt}, \omega^{a}_{jt}) = \Big(\frac{1}{\lambda} - 1 \Big) 2p_{jt} \Big(\frac{2p_{jt}e^{\omega^{a}_{jt} - \lambda_0}}{\lambda} \Big)^{\frac{1}{\lambda-1}} + 2p_{jt}s_{jt} \frac{h_{jt}(X_{jt}, \omega^{y}_{jt})}{b_{jt}}. \tag{14}$$

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 (14). 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

After solving for n_{jt}^{a*} for any given (b_{jt}, s_{jt}), we now explain how an FGD plant chooses (b_{jt}, s_{jt}) to minimize the total variable cost. The total variable cost is the sum of the coal cost and the minimized pollution control cost from above, where the FGD plants also face constraints on (b_{jt}, s_{jt}):

$$\min_{b_{jt}, s_{jt}} w_{jt}^c(b_{jt}, s_{jt}; f_{jt}) n(b_{jt}; r_{jt}, p_{jt}, X_{jt}, \omega_{jt}^y) + C_{\text{FGD}}^s(b_{jt}, s_{jt}; r_{jt}, p_{jt}, X_{jt}, \omega_{jt}^y, \omega_{jt}^a),$$

subject to $b_{jt} \in \left[\underline{b}_{jt}, \bar{b}_{jt}\right]$ and $s_{jt} \in \left[\underline{s}_{jt}, \bar{s}_{jt}\right]$.

The FOCs with respect to b_{jt} and s_{jt} are

$$\frac{\partial w_{jt}^c(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} n_{jt} + w_{jt}^c(b_{jt}, s_{jt}; f_{jt}) \frac{\partial n_{jt}}{\partial b_{jt}} + \frac{\partial C_{\text{FGD}}^s(s_{jt}, b_{jt}; r_{jt}, p_{jt}, X_{jt}, \omega_{jt}^y, \omega_{jt}^a)}{\partial b_{jt}} - \underline{\mu}_{jt}^b + \bar{\mu}_{jt}^b = 0. \tag{15}$$

$$\frac{\partial w_{jt}^c(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} n_{jt} + \frac{\partial C_{\text{FGD}}^s(s_{jt}, b_{jt}; r_{jt}, p_{jt}, X_{jt}, \omega_{jt}^y, \omega_{jt}^a)}{\partial s_{jt}} - \underline{\mu}_{jt}^s + \bar{\mu}_{jt}^s = 0, \tag{16}$$

We derive similar functional forms of the two FOCs as in equations (6) and (7) 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

$$C^s_{FGD}(X_{jt},\omega^y_{jt},\omega^a_{jt};f_{jt},p_{jt},r_{jt}) = \big(w^c_{jt}(b^*_{jt},s^*_{jt};f_{jt})+2p_{jt}s^*_{jt}\big)\frac{h_{jt}(X_{jt},\omega^y_{jt})}{b^*_{jt}}+\Big(\frac{1}{\lambda}-1\Big)2p_{jt}\Big(\frac{2p_{jt}e^{\omega^a_{jt}}}{\lambda}\Big)^{\frac{1}{\lambda-1}}. \tag{17}$$

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 difference 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. First, we estimate the hedonic coal price function in terms of b_{jt} and s_{jt} 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. Second, we estimate the abatement cost function using plant-year-level data on the FGD O&M cost and the sulfur abatement level. Lastly, 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.

5.1 Estimation of the Coal Price Function

We use transaction-level data to estimate the hedonic coal price function. We drop the subscripts to simplify notations in this section. A transaction occurs between a mine (m) and plant (j) during a given month-year (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.[16]

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. We use d^m, d^j, and their interactions to control for the unobserved freight transportation charge, which depends on the distance between the mine and plant and the rail carriers' market power. Thus, f is independent of (b, s). The stochastic hedonic coal price function is

$$\begin{aligned} w^c(b, s; f) = \alpha_0 + \alpha_b b + \alpha_s s + \alpha_{bb} b^2 + \alpha_{ss} s^2 + \alpha_{bs} bs \\ + \alpha_{sA} s\, A^\tau + \sum_{q=1}^{Q} \alpha_q d^q + \sum_{\tau=1}^{\mathcal{T}} \alpha_\tau d^\tau + \sum_{m=1}^{M} \alpha_m d^m \\ + \sum_{j=1}^{J} \alpha_j d^j + \sum_{m=1}^{M} \sum_{j=1}^{J} \alpha_{mj} d^m d^j + \epsilon^w, \end{aligned} \tag{18}$$

where ϵ^w is a coal price shock.

Using coal transaction data, we compute OLS estimates of the stochastic coal price equation (18). 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

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial b_{jt}} = \alpha_b + 2\alpha_{bb} b_{jt} + \alpha_{bs} s_{jt}, \tag{19}$$

$$\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}} = \alpha_s + 2\alpha_{ss} s_{jt} + \alpha_{bs} b_{jt} + \alpha_{sA}\, A_t. \tag{20}$$

[16]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.

Since plants do not report the prices of trades to EPA, we do not observe p_{jt} in the data. After estimating the coal price function, we compute p_{jt} using the FOC for the optimal s_{jt} in equation (8) for both types of plants, which implies that $p_{jt} = -2\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}}$.

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:

$$\omega^a_{jt} = g^a(\omega^a_{jt-1}) + \xi^a_{jt} = \rho^a_0 + \rho^a_1 \omega^a_{jt-1} + \rho^a_2 (\omega^a_{jt-1})^2 + \xi^a_{jt}, \tag{21}$$

where ξ^a_{jt} is the shock to the abating efficiency. Specifically, ξ^a_{jt} represents the shock to the plant-year costs of FGD feed, waste disposal, and related costs. We assume that ξ^a_{jt} is independent of ξ^a_{jt-1}. The parameters to be estimated in this step are λ in equation (10) and $\boldsymbol{\rho}^a = (\rho^a_0, \rho^a_1, \rho^a_2)$ in equation (21). The stochastic version of the logarithm of abatement cost in equation (10) is

$$\ln C^a(s_{jt}, n^a_{jt}, \omega^a_{jt}) = \lambda_0 - \omega^a_{jt} + \lambda(\ln s_{jt} + \ln n^a_{jt} + \ln r_{jt}) + \epsilon^a_{jt}, \tag{22}$$

where ϵ^a_{jt} is an idiosyncratic error. We assume that ϵ^a_{jt} has a zero mean and is uncorrelated with the variables on the right-hand side of equation (22).

Replacing ω^a_{jt} in equation (22) with $\omega^a_{jt} = g^a(\omega^a_{jt-1}) + \xi^a_{jt}$, we get

$$\ln C^a(s_{jt}, n^{a*}_{jt}, \omega^a_{jt}) = \lambda_0 - g^a(\omega^a_{jt-1}) + \lambda(\ln s_{jt} + \ln n^a_{jt} + \ln r_{jt}) - \xi^a_{jt} + \epsilon^a_{jt}. \tag{23}$$

The FOC for abatement level in equation (13) implies that $\omega^a_{jt-1} = (\lambda - 1)\ln(s_{jt-1} n^a_{jt-1} r_{jt-1}) + \ln\left(\frac{\lambda}{2p_{jt-1}}\right)$. Plugging this into equation (23), we replace the unobserved abating efficiency with observed variables. The new error term is $(-\xi^a_{jt} + \epsilon^a_{jt})$.

The efficiency shock ξ^a_{jt} is positively correlated with n^a_{jt} in the abatement cost equation (23). 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 $\boldsymbol{\rho}^a$.[17] Denoting the vector of IV by Z^a_{jt}, the moment conditions are

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

The IVs include the lagged price of coal, lagged price of labor, and the current price of labor. The lagged price of coal is not correlated with ξ^a_{jt} because it is only correlated with the lagged ξ^a_{jt-1}, which is independent of ξ^a_{jt}. The lagged and current prices of labor are also independent of ξ^a_{jt}. This obtains since ξ^a_{jt} contains only the shock to FGD O&M cost, which includes the cost of chemical feed, waste disposal, and other related costs. By construction, ϵ^a_{jt} is uncorrelated with the instruments.

5.3 Estimation of the Production Function Parameters

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

$$\omega^y_{jt} = g^y(\omega^y_{jt-1}) + \xi^y_{jt} = \rho^y_0 + \rho^y_1 \omega^y_{jt-1} + \rho^y_2 (\omega^y_{jt-1})^2 + \xi^y_{jt}, \tag{24}$$

where ξ^y_{jt} is the shock to ω^y_{jt}, due to unanticipated changes in the overall operating productivity of a plant. We assume that ξ^y_{jt} and ξ^y_{jt-1} are independent of each other because ρ^y_1 and ρ^y_2 can capture any persistency in ω^y_{jt}. We also 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 are independent of the production function parameters. Thus, we focus only on

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

the function $\tilde{C}_{jt} (= w^c_{jt} n_{jt} + 2p_{jt} s_{jt} n_{jt})$, whose parameters are those of the Cobb-Douglas 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 (9) and the first two terms in the FGD plants' total costs in (17), we get the equivalent expression,

$$\ln \tilde{C}_{jt} = \ln\left(\frac{w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt} s^*_{jt}}{b^*_{jt}} \right) + \frac{1}{\beta_h} (\ln y_{jt} - \beta_0 - \beta_k \ln k_{jt} - \beta_l \ln l_{jt} - \omega^y_{jt}) + \epsilon^c_{jt}, \tag{25}$$

where we replace h_{jt} using equation (2), and ϵ^c_{jt} is an idiosyncratic plant-year cost shock. Plugging $\omega^y_{jt} = g^y(\omega^y_{jt-1}) + \xi^y_{jt}$ into equation (25) yields

$$\begin{aligned} \ln \tilde{C}_{jt} = \ln\left(\frac{w^c_{jt}(b^*_{jt}, s^*_{jt}; f_{jt}) + 2p_{jt} s^*_{jt}}{b^*_{jt}} \right) + \frac{1}{\beta_h} (\ln y_{jt} - \beta_0 - \beta_k \ln k_{jt} - \beta_l \ln l_{jt} \\ - g^y(\omega^y_{jt-1})) - \frac{1}{\beta_h} \xi^y_{jt} + \epsilon^c_{jt}. \end{aligned} \tag{26}$$

We then take the logarithm of the lagged production function to obtain $\omega^y_{jt-1} = \ln y_{jt-1} - \beta_0 - \beta_k \ln k_{jt-1} - \beta_l \ln l_{jt-1} - \beta_h \ln h_{jt-1}$. We substitute this into equation (26) to replace the unobserved ω^y_{jt-1}.[18]

The estimation of (26) is subject to endogeneity issues. First, if a serial correlation exists in y_{jt}, then a serial correlation also exists in h_{jt}, which makes ξ^y_{jt} and h_{jt-1} correlated. Second, ξ^y_{jt} is also correlated with (b_{jt}, s_{jt}) from their FOCs and the Markov process in equation (24). Therefore, we estimate the parameters in (26) using GMM with IVs. The moment conditions employ the orthogonality between the composite error term and the IVs in Z^c:

$$E\left[Z^c_{jt} \left(-\frac{1}{\beta_h} \xi^y_{jt} + \epsilon^c_{jt} \right) \right] = 0, \tag{27}$$

where Z^c_{jt} includes the logarithm of the lagged price of coal, the lagged capacity and

[18] One can interpret our cost function approach as using the production function as the control function to substitute for the unobserved ω^y_{jt}.

labor inputs, and the current and lagged prices of labor and capital. The instruments are uncorrelated with ξ^y_{jt} and ϵ^c_{jt} for the following reasons. First, ξ^y_{jt} does not affect the lagged price of coal or the lagged and current prices of labor and capital. Second, ξ^y_{jt} is also uncorrelated with lagged values of labor and capacity. Lastly, ϵ^c_{jt} is the idiosyncratic error in the total variable costs, which is uncorrelated with the input prices and lagged inputs.

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 (19) and (20). The marginal prices of b_{jt} and s_{jt} are positive and negative, respectively, for all plants in all years. They are significant at the 0.01-level. We find that the coal price goes up by \$1.79 on average if b_{jt} increases by one million Btu per ton of coal. The price drops by \$1.89 on average if s_{jt} increases by 1%, 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 impact of sulfur content on the price of coal.

Table 2: Estimates of Parameters in the Coal Price Function

Coefficient	Estimates	Coefficient	Estimates
α_s	0.946	α_{bb}	-0.004
	(0.824)		(0.000)
α_b	2.200	α_{bs}	-0.181
	(0.068)		(0.030)
α_{ss}	0.286	α_{sA}	0.048
	(0.037)		(0.007)
R^2	0.754		
Observations	69,674		

Parentheses contain estimated asymptotic standard errors.

The symbol *** indicates significance at the .01 level using a two-tailed t-test.

Table 3: Estimated Marginal Prices of Sulfur and Btu

	Mean
$\frac{\partial w^c_{jt}}{\partial s_{jt}}$	-1.891
	(0.113)
$\frac{\partial w^c_{jt}}{\partial b_{jt}}$	1.786
	(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 permit price was \$105 per ton of SO_2 in 1995, with a standard deviation of \$30. The average permit price drops to \$87 in 2005, with a standard deviation of \$26. The permit prices dropped over the sample period.[19]

Table 4 shows the estimated parameters and standard errors for the abatement cost function (23).[20] The estimate of λ is 2.323, which implies that the marginal cost of

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

[20] We dropped two non-FGD plants due to obvious mistakes in the reported quantity of coal, which leaves us a sample of 18 FGD plants and 60 non-FGD plants in the estimation of the abatement cost function and equation (26).

abating sulfur is increasing. This is very consistent with that obtained by running the Berkenpas, Rubin, and Zaremsky (2007) Integrated Emission Control Model (IECM).[21] 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. The estimated values of ω^a_{jt} vary substantially across plants. The average unit abatement cost for the most efficient plant is only 4% of that for the least efficient plant. The estimates of ρ^a_1 and ρ^a_2 imply that $\frac{\partial \omega^a_{jt+1}}{\partial \omega^a_{jt}} = 0.89$, so there is significant persistency in the plants' abatement efficiencies.

Table 4: Estimates of Parameters in the Abatement Cost Function

	$ln(C^a)$
λ	2.323
	(0.202)
ρ^a_0	-3.622
	(2.364)
ρ^a_1	1.209
	(0.412)
ρ^a_2	0.004
	(0.018)
Observations	170

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

Table 5 shows the estimated parameters and standard errors for equation (26) and the productivity transition equation. Column (1) shows the results using our cost function approach to estimating the Cobb-Douglas production function parameters. We label this column as CM for cost minimization. The estimates of (β_k, β_h) are positive and significant, with $\hat{\beta}_k = 0.184$, $\hat{\beta}_h = 0.906$. The impact of labor on the total generation is insignificant. Plants exhibit slight increasing returns in electricity generation on average, since $\hat{\beta}_k + \hat{\beta}_h + \hat{\beta}_l = 1.06 > 1$, which is consistent with findings in the litera-

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

ture as summarized by Atkinson (2019). The estimated values of ω_{jt}^{y} indicate that the most productive plant can generate four times more electricity than the least productive plant, using the same input bundles.[22] The estimate of ρ_1^y is 0.866 and significant, implying that lagged productivity significantly influences current productivity. More productive plants are consistently more productive over time. The correlation between the estimated ω_{jt}^{y} and ω_{jt}^{a} is 0.28 for the FGD plants.

We compare our results with those from the approaches that directly estimate the production function. We focus on Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015), who invert the intermediate input function to control for ω_{jt}^{y}.[23] The intermediate input is the total heat input, h_{jt}. The results from LP are in column (2), and those from ACF are in column (3). The results show that the LP and ACF approaches underestimate β_k and overestimate β_h relative to the CM approach. These differences arise because LP and ACF assume that the inverted control function for ω_{jt}^{y} is independent of output, while our cost-minimization approach assumes that ω_{jt}^{y}

[22]That is, $\frac{exp(max\{\omega_j^y\})}{exp(min\{\omega_j^y\})} = 4.00$, where $exp(\omega_j^y)$ is the plant average generation efficiency.

[23]The LP approach assumes that capital is pre-determined and labor is variable in each period. To implement this approach, we use their two-step estimation method. In the first step, the estimation equation is

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_h h_{jt} + g_t(k_{jt}, a_{jt}, h_{jt}) + \eta_{jt} = \beta_l l_{jt} + \phi_t(k_{jt}, a_{jt}, h_{jt}) + \eta_{jt},$$

where $g_t(k_{jt}, a_{jt}, h_{jt})$ is the inverted variable input function, which we use to obtain unobserved ω_{jt}^{y}. We approximate the ϕ_t with third-order polynomials and estimate the coefficients using GMM. The moment conditions are $E\left[\eta_{jt} z_{1jt}\right] = 0$, where $z_{1jt} = (1, k_{jt}, a_{jt}, h_{jt})$ and higher order polynomials of them. In the second step, we replace ω_{jt}^{y} by its transition function. Denote $x_{jt} = (k_{jt}, l_{jt}, h_{jt})$. The production function becomes

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_h h_{jt} + f(\omega_{jt-1}, x_{jt-1}) + \xi_{jt} + \eta_{jt},$$

where the transition function is

$$f(\omega_{jt-1}, x_{jt-1}) = \rho_0 + \rho_1 \omega_{jt-1} + \rho_2 \omega_{jt-1}^2 + \rho_3 \omega_{jt-1}^3 + x'_{jt-1}\rho.$$

We estimate the parameters using GMM, and the moment conditions are $E\left[(\xi_{jt} + \eta_{jt})(\theta_0) z_{2jt}\right] = 0$, where $z_{2jt} = (1, k_{jt}, l_{jt-1}, a_{jt-1}, h_{jt-1})$ and higher order polynomials of them, and a_{jt} is vintage.

Our implementation of the ACF approach is similar to that of the LP approach but assumes that labor is also pre-determined. In the first step, the inverted intermediate input function also depends on labor. That is, we replace productivity by $g_t(k_{jt}, l_{jt}, a_{jt}, h_{jt})$. The second step is similar to the LP method.

depends on output when we invert the production function to control for ω^y_{jt}.

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

	(1) CM	(2) LP	(3) ACF
β_k	0.184	0.008	0.022
	(0.092)	(0.021)	(0.020)
β_l	-0.029	-0.024	-0.038
	(0.031)	(0.018)	(0.012)
β_h	0.906	1.057	1.052
	(0.081)	(0.014)	(0.012)
ρ^y_1	0.866	0.846	0.954
	(0.016)	(0.012)	(0.032)
N	760	760	760

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

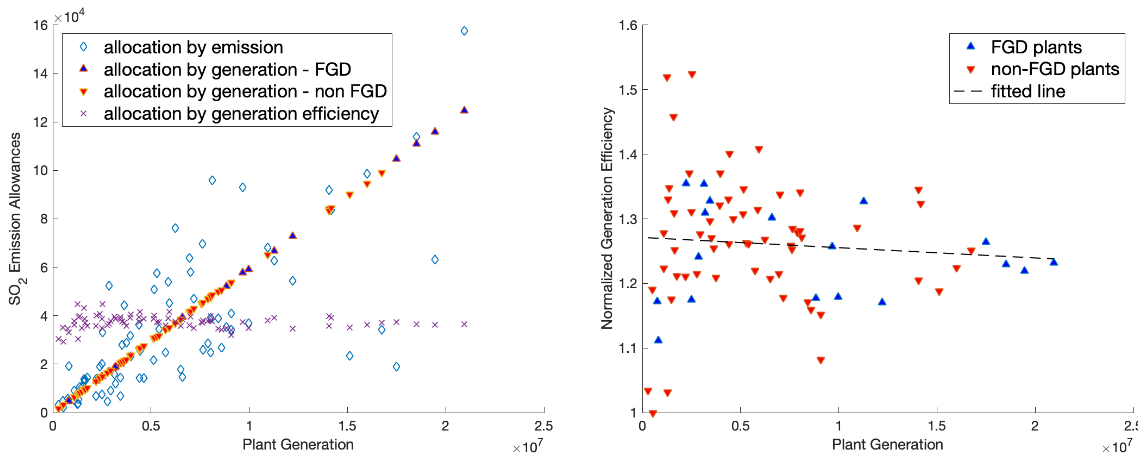
7 Different Allocations under a Non-functioning Permit System

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. 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 to initial permit allocations based on total emissions, emission rates, total output, and abatement efficiency.[24] In this analysis, we compare three different schemes to allocate SO_2 emission permits across plants. Under the first scenario, permits are allocated in amounts equal to the observed emissions of plants as in the

[24]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.

data (where plants trade permits under negligible transaction costs). Under the second scenario, permits are proportional to each plant's annual electricity generation.[25] Under the third scenario, permits are proportional to the generation efficiency of each plant.[26] In all scenarios, the aggregate amount of permits equals the aggregate annual emissions of all plants in the data. We assume that the permit system is non-functioning so that plants do not trade permits.[27]

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 allocated permits against the average annual plant generation in Figure 4a. The horizontal axis is a plant's average yearly generation during the sample period, and the vertical axis is its average allowances. Each point on the graph represents a plant. The scatter points with light blue diamond markers are the allowances with the first method.

[25]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.

[26]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.

[27]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.

The points with solid triangular markers are allowances with the second method. The cross-markers represent allowances with the third method. The blue upward-pointing and red downward-pointing triangular points are for FGD plants and non-FGD plants, respectively.

Figure 4a shows that the FGD plants have higher generation than non-FGD plants on average. Under the first two methods, plants with higher generation receive more permits, but considerable dispersion exists in allowances for a given generation level with the first method. The distribution of permits is close to uniform except for low-generation plants. This is because small plants are more efficient than large plants, as shown in Figure 4b, which plots the average generation efficiency against generation by plant.[28] The dashed fitted line has a negative slope, implying that large plants have smaller ω_{jt}^{y} on average.

Table 6 compares the total permits 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. Columns (1) and (4) are the same because the first scenario is for allocation based on emissions 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 emissions.

Table 6: Total Permits of Plants (all years)

	(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
Total permits (10^7)	3.20	3.20	3.20	3.20
– FGD	0.66	1.34	0.72	0.66
– Non-FGD	2.55	2.17	2.48	2.55

Under each allocation method, denote the new SO_2 allowances for plant j in year t by

[28] 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.

$\tilde{S}^e_{jt}$. 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.[29] 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:

$$\min_{b_{jt}, s_{jt}} \left\{ w^c_{jt}(s_{jt}; b_{jt}, f_{jt}) n(b_{jt}; X_{jt}, \omega^y_{jt}) + e^{\lambda_0 - \omega^a_{jt}} (n^a_{jt} s_{jt} r_{jt})^{\lambda} \right\}, \tag{28}$$

subject to the emission constraint that $2\, s_{jt}\, (n(b_{jt}, X_{jt}, \omega^y_{jt}) - n^a(s_{jt}; b_{jt}, X_{jt}, \omega^y_{jt}, \omega^a_{jt}) r_{jt}) \leq \tilde{S}^e_{jt}$. 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 scenarios and compare the aggregate and average costs across the three allocation methods.[30] Columns (1) to (3) in Table 7 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, FGD plants would sell permits and purchase higher-sulfur-content coal if the marginal abatement cost were lower than the permit price. Non-FGD plants who buy the permits would also purchase higher-sulfur-content coal since they

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

[30]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.

have more permits. Across columns (1)-(3), the average sulfur content of FGD plants is the highest in column (2) at 1.85%. This is because they receive the most permits when based on generation, as shown in Table 6. 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 7: Impacts of Different Emission Permit Allocation Mechanisms

	(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 (%)	1.15	1.10	1.09	1.24
– FGD	1.80	1.85	1.82	2.01
– Non-FGD	0.84	0.73	0.74	0.87
Average Btu content (10^6/ton)	21.94	22.12	22.33	22.44
– FGD	21.39	21.33	21.35	21.78
– Non-FGD	22.21	22.49	22.80	22.76
Average coal price ($/ton)	29.56	30.00	30.41	30.44
– FGD	24.31	24.13	24.21	24.92
– Non-FGD	32.07	32.81	33.37	33.07
Total coal consumption (10^9 tons)	2.52	2.51	2.48	2.39
– FGD	0.85	0.85	0.85	0.80
– Non-FGD	1.68	1.66	1.64	1.58
Coal costs all years ($$10^{10}$)	7.11	7.15	7.18	6.96
– FGD	1.98	1.97	1.97	1.94
– Non-FGD	5.13	5.18	5.20	5.02
Average coal abatement (%)	76.71	56.64	70.26	79.53
Abatement costs all years ($$10^9$)	1.51	0.75	1.01	1.63
Total costs all years ($$10^{10}$)	7.26	7.22	7.28	7.13
– FGD	2.13	2.05	2.08	2.10
– Non-FGD	5.13	5.18	5.20	5.02

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.

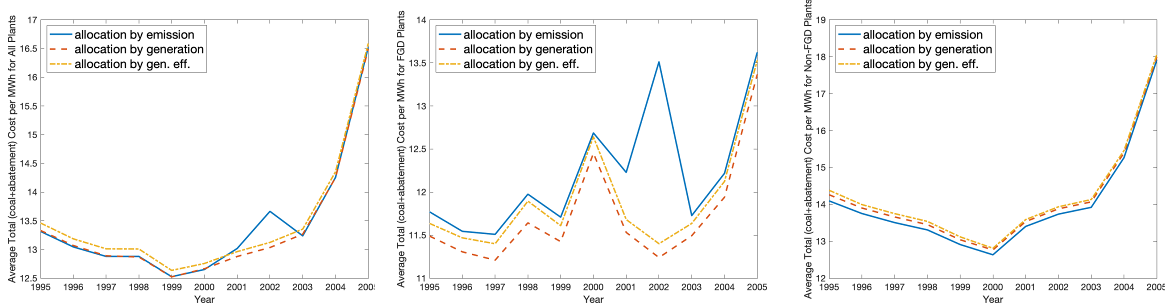
The average Btu content in columns (1)-(3) is lower than in column (4). This is because of the lower sulfur content in columns (1)-(3) than in column (4), which causes plants to lower Btu content to reduce coal costs. As in the data, FGD plants choose lower Btu and higher sulfur content coal than non-FGD plants. Hence, coal prices are significantly lower for FGD plants in all scenarios, slightly above \$24 per ton, while the prices for non-FGD plants are more than \$32 per ton. Because of the lower Btu content, the total coal consumption is greater in columns (1)-(3) than in column (4). Across the three scenarios, the coal costs of FGD plants are the lowest when allocation is based on generation, in which case they receive the most permits. The coal costs of non-FGD plants are the lowest when based on emissions, in which case they receive the most permits.

The allocation methods have dramatic impacts on the FGD plants' abatement rate and costs. The coal abatement rates are 76.71%, 56.64%, and 70.26% in columns (1)-(3). The abatement rates in the counterfactual scenarios are lower than that in the data. This is probably due to the fact that lower sulfur content coal is available in the counterfactual. The FGD plants' total abatement costs are \$1.51, \$0.75, and \$1.01 billion in columns (1)-(3), respectively. FGD plants incur the lowest abatement costs in column (2). Thus, compared with column (2), the allocation methods in columns (1) and (3) could increase the abatement costs by 101% and 35%, respectively.

The FGD plants' total costs of coal and abatement are the lowest in column (2) because their abatement costs are the lowest in this case. Their total costs are lower in columns (2)-(3) than in column (1) because they receive more permits in the former two cases than in the latter case. Among the three counterfactual scenarios, the non-FGD plants have the lowest total costs when allocation is based on emissions. The total costs of all plants under the three methods are \$72.6, \$72.2, and \$72.8 billion, respectively. The costs are the lowest in column (2) due to the significantly lower total costs of FGD plants with this allocation method. For non-FGD plants, total costs of coal and abatement in columns (1)-(3) are higher than in column (4) mainly because they receive fewer permits in columns (2) and (3) than in the data.[31]

[31]The total costs for all plants in column (1) are greater than in column (4) for both types of plants. This is because that we assume one ton of sulfur is transformed into two tons of SO_2 and thus requires

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



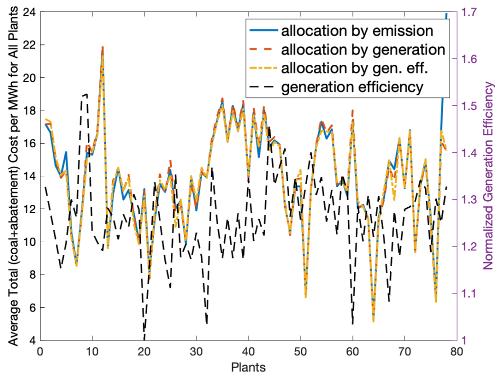
(a) All Plants (b) FGD Plants (c) Non-FGD Plants

Figure 5 shows the average total costs per mWh generation under the three allocation methods. Figure 5a shows that the average cost is the lowest in most years with the first method. The stark increase since 2000 with the first method is because of fewer permits (emissions) for the FGD plants due to high permit prices caused by supply disruptions and speculation in the permit trading market.[32] The allocation methods have different impacts on the FGD and non-FGD plants. Figure 5b shows that the average total costs for FGD plants are lowest with the second method. Figure 5c shows that non-FGD plants have the lowest costs with the first method.

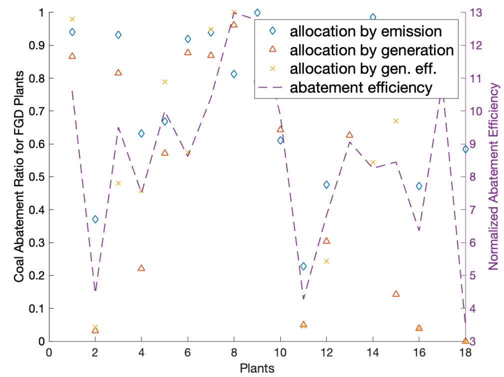
permits for exactly two tons of emissions. However, in the data, the ratio between the total amounts of sulfur and SO_2 can be somewhat smaller or greater than two, as shown for the non-FGD plants in Figure 2b. This makes the emission constraints in column (1) slightly different from the emission constraints in column (4). This factor causes the differences in total costs between columns (1) and (4).

[32]See section C of the Appendix for the comparison of FGD plants' total permits by year across the three methods. From 2000 to 2005, a number of unforeseen rail supply disruptions and speculation about the dramatic reduction in permits under CAIR pushed permit prices and clean coal prices to very high levels. See section B of the Appendix for more details on the permit prices.

Figure 6: Average Costs and ω_{jt}^{y}, Abatement Ratio and ω_{jt}^{a}



(a) Average Total Costs and Efficiency by Plant



(b) Abatement Percentage and Efficiency

To see the impact of the unobserved ω_{jt}^{y} and ω_{jt}^{a} on the plants, we plot the average total cost against generation efficiency and the abatement percentage against abatement efficiency in Figure 6a and Figure 6b. In Figure 6a, the horizontal axis lists the 78 plants. The left vertical axis is the average total cost per mWh, and the right vertical axis is the normalized generation productivity $exp(\omega_{jt}^{y})$.[33] The dashed curve plots the normalized $exp(\omega_{jt}^{y})$. The other curves represent the average costs under the three allocation methods. We find small but negative correlations between the average total costs and generation productivities.[34]

In Figure 6b, the horizontal axis lists the 18 FGD plants. The left vertical axis is the coal abatement percentage, and the right vertical axis is the normalized abatement efficiency $exp(\omega_{jt}^{a})$. The dashed curve plots the normalized $exp(\omega_{jt}^{a})$. The three types of markers represent the abatement percentages under the three methods. We 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 .76, .85, and .94 for the three allocation methods, respectively.

[33] We first compute the average $exp(\omega_{jt}^{y})$ for each plant across the 11 years, and we use the minimum of these averages to normalize the plants' $exp(\omega_{jt}^{y})$ in all years. The normalization of $exp(\omega_{jt}^{a})$ is similar.

[34] The correlations between the average total costs and the normalized $exp(\omega_{j}^{y})$ by plant are -.08, -.10, and -.12 for the three allocation methods, respectively.

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 that 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 80 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 emission permits, non-FGD plants can only employ the latter strategy. Assuming a Cobb-Douglas 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 the endogeneity with 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 improve 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 caused by 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 demand for sulfur and Btu content and 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 a reduction of \$0.4-\$0.6 billion in total costs of all plants and a reduction of 35%-101% in FGD plants' abatement costs during the sample period.

We can adapt our methodology to analyze the effects of output and emission regulations for other pollutants generated by cost-minimizing plants and firms. Currently, this would most likely involve greenhouse gas emissions from the many existing and newly-constructed coal-fired power plants worldwide, which are among the largest sources of these pollutants. While numerous greenhouse gas cap-and-trade systems exist worldwide, one of the most active 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. Our methodological approach, which separately estimates output productivity and abatement efficiency, can also be employed to calculate the growth in abatement efficiency of coal-fired power plants over time under cap-and-trade programs worldwide.

References

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.*

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 Permit Prices

From the coal transaction-level data, we know that every plant bought coal from multiple mines in a year, so every plant has some flexibility in choosing (b_{jt}, s_{jt}), which implies that $\underline{b}_{jt} < \bar{b}_{jt}$ and $\underline{s}_{jt} < \bar{s}_{jt}$. Thus, at least one constraint for b_{jt} and one for s_{jt} are not binding, so at least one $\mu^b_{jt} \in \{\underline{\mu}^b_{jt}, \bar{\mu}^b_{jt}\}$ and at least one $\mu^s_{jt} \in \{\underline{\mu}^s_{jt}, \bar{\mu}^s_{jt}\}$ are zero. More specifically, we now argue that $\underline{\mu}^s_{jt} = \bar{\mu}^s_{jt} = 0$, while $\underline{\mu}^b_{jt}$ and $\bar{\mu}^b_{jt}$ cannot both be zero.

There are four possible combinations for the values of the Lagrangian multipliers in the FOCs for (b_{jt}, s_{jt}). First, none of the four constraints is binding. That is, $\underline{\mu}^s_{jt} = \bar{\mu}^s_{jt} = \underline{\mu}^b_{jt} = \bar{\mu}^b_{jt} = 0$. In this case, the two FOCs imply very different values for p_{jt}, which makes this case improbable. Second, the two constraints for b_{jt} are not binding, and one constraint for s_{jt} is binding. That is, $\underline{\mu}^b_{jt} = \bar{\mu}^b_{jt} = 0$, and either $\underline{\mu}^s_{jt} = 0$ or $\bar{\mu}^s_{jt} = 0$. Plugging $\underline{\mu}^b_{jt} = \bar{\mu}^b_{jt} = 0$ into equation (7), we can get p_{jt}, but these permit prices are dramatically different from the yearly average permit prices in the permit auction data. Third, the two constraints for s_{jt} are not binding, and one constraint for b_{jt} is binding. That is, $\underline{\mu}^s_{jt} = \bar{\mu}^s_{jt} = 0$, and either $\underline{\mu}^b_{jt} = 0$ or $\bar{\mu}^b_{jt} = 0$. Plugging $\underline{\mu}^s_{jt} = \bar{\mu}^s_{jt} = 0$ into equation (8), we obtain estimated permit prices that are reasonably close to auction prices. Lastly, one constraint for b_{jt} is binding, and one constraint for s_{jt} is binding. This implies that the plant chooses corner solutions for both b_{jt} and s_{jt}, but this is unlikely given the fact that the plants are buying coal from multiple mines with a variety of coal characteristics each year.

Therefore, the third case is the most consistent with the data. This finding also aligns with the variation of b_{jt} and s_{jt} in the data. The standard deviation of b_{jt} across plants is only 1.93% of the average b_{jt}, but the standard deviation of s_{jt} across plants is 13.13% of the average s_{jt}. Thus, plants are more likely to face binding constraints on b_{jt} than s_{jt}. Thus, the FOC for s_{jt} implies that $p_{jt} = -2\frac{\partial w^c_{jt}(b_{jt}, s_{jt}; f_{jt})}{\partial s_{jt}}$. We use this to compute the plant-year permit prices. Table B.1 shows the average estimated permit price and the EPA auction price of permits by year.

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

year	Estimates	EPA Auction Price
1995	105.55	130
1996	97.30	66
1997	93.11	107
1998	90.37	108
1999	86.60	201
2000	76.49	126
2001	80.31	174
2002	83.31	161
2003	83.43	172
2004	84.90	260
2005	86.71	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 in 2005 and 2006 due to hurricanes, Powder River Basin train derailments, the CAIR, and speculation, none of which are captured in our cost model, pushed permit price to very high levels.

C Counterfactual Permit Allocation Methods

Table C.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 C.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