Agency Costs in Environmental Regulation: Evidence from Regulated Electric Utilities[1]

Jose Miguel Abito[2]

December 18, 2014

[1]This paper is based on the first chapter of my PhD Dissertation at Northwestern University and was previously circulated under the title "Welfare Gains from Optimal Pollution Regulation." I would like to thank my advisors Aviv Nevo, David Besanko and Robert Porter for all of their help and guidance. I also thank Mark Chicu, Daniel Diermeier, Uli Doraszelski, Jose Espin Sanchez, Matt Gentzkow, Joe Harrington, Igal Hendel, JF Houde, Ginger Jin, Brad Larsen, Matt Masten, Tiago Pires, Mike Powell, Mar Reguant, Min Ren, Bill Rogerson, Stephen Ryan, Jesse Shapiro, Matt Shum, Kosuke Uetake, Mike Whinston, Ali Yurukoglu and seminar participants at Northwestern and at other institutions. Data acquired from SNL Financial was partly funded by the TGS Graduate Research Grant and by the Center for the Study of Industrial Organization (CSIO) at Northwestern University.

[2]Department of Business Economics and Public Policy, Wharton School, University of Pennsylvania. Email: abito@wharton.upenn.edu.

Abstract

Successful implementation of pollution regulation requires redistributing benefits to firms incurring costs. When firms have private information, they may overstate costs and demand higher compensation. I examine the interaction of emissions and price regulation in the electricity industry to study optimal pollution regulation when there are agency costs. I exploit the effect of price regulation on firms' strategic behavior to identify exogenous and endogenous components of abatement costs, and quantify the tradeoff between efficiency and information rent extraction. Due to large information rents, I find substantial welfare gains from distorting emissions upward relative to the "least cost/efficient" level. Assuming a constant marginal damage of \$300 per ton of SO_2 emissions, the least cost solution requires a subsidy of \$1.5 billion per year to implement a cap of about 5 million tons. In contrast, the optimal mechanism reduces the required subsidy by 42%, implying that agency costs account for almost half of the subsidy under a least cost solution.

1 Introduction

Environmental regulation imposes nontrivial costs to polluting firms. To comply with environmental regulation, firms invest in expensive technologies or significantly change their production processes; for example, coal-fired power plants need to install scrubbers or burn cleaner and more expensive forms of coal to comply with emissions regulation. Because of these costs, successful implementation often necessitates compensation of affected firms to decrease the burden of regulation. However, the regulator has to rely on cost information from the firm to determine the necessary compensation that achieves an acceptable policy. If costs are not fully observed or the regulator cannot fully monitor behavior, firms may overstate costs or behave in a way that is inconsistent with cost-minimization to receive higher compensation.

My paper empirically measures the *agency costs* of implementing acceptable policies arising from asymmetric information between the regulator and firms, and investigates how these agency costs affect relative welfare outcomes under various policies. My focus is on sulfur dioxide (SO_2) emissions regulation of price-regulated electric utilities. In this context, agency costs reflect the additional costs of extracting private information from a firm about its *intrinsic* fuel efficiency, and providing incentives for the firm to pursue actions that improve fuel efficiency when these activities are difficult to monitor. The agency costs I estimate are large and form a significant fraction of the overall cost of implementing policies. These large agency costs have both positive and normative implications. From a positive perspective, substantial agency costs may be an important reason why implemented policies deviate from ideal ones, often in favor of certain parties (e.g. delayed implementation of a more stringent policy and concessions given to established firms). From a normative perspective, I estimate that implementing emission levels that *deviate* from the ones that equate marginal damages with true (efficient) marginal abatement costs yields substantial welfare gains.

For coal-fired power plants, the process of burning coal creates both electricity and emissions as outputs. Hence, fuel efficiency is an important factor with respect to the cost of producing electricity and the cost of reducing emissions. I measure agency costs associated with reducing emissions by investigating agency costs associated with fuel efficiency. While realized fuel efficiencies are observed or can be inferred from available data, fuel efficiency is a function of *exogenous* or intrinsic factors that affect overall utility-level productivity (e.g. productivity of individual generating units, ability of current boilers to accept lower sulfur coal), and *endogenous* factors that the utility can directly control but hard for the regulator to monitor (e.g. allocation of output across plants with varying productivities, management of plant maintenance and operations). To reach an acceptable policy,

the regulator has to provide more compensation to intrinsically inefficient firms.[1] However, since compensation depends on realized fuel efficiencies, firms lose incentives to endogenously seek actions consistent with maximizing fuel efficiency. Therefore the key challenge for the regulator (and the econometrician) is to decompose realized fuel efficiencies into these exogenous and endogenous parts.

I propose a novel identification strategy that exploits how *price* regulation (or *rate* regulation) of electric utilities affects fuel efficiency and the cost of reducing emissions. Firms' strategic response to rate regulation allows identification of key structural parameters that determine strategic response to emissions regulation.

Rate regulation is characterized by alternating periods when (i) prices are set with the regulator (rate case), and when (ii) prices are fixed until the next rate case (regulatory lag). These two periods create different incentives for firms to efficiently produce electricity. My paper provides empirical evidence of poor incentives during rate cases, but strong incentives during regulatory lags. I find that during rate cases, fuel efficiency goes down by about 4% to 6% compared to non-rate case years. Since fuel accounts for 75% of operating costs, this reflects a 4% to 6% increase in variable generating costs during the rate case relative to the regulatory lag.

This pattern in cost and fuel efficiency hints on the nature of interaction between rate regulation and firm incentives. To understand the mechanism through which rate regulation affects firm behavior, I adopt the cost and information structure of Laffont and Tirole's (1986) model of natural monopoly regulation. Specifically, I model fuel efficiency as a function of two components: an exogenous, *intrinsic type* component, and an *endogenous effort* component. A firm's type and exerted effort are both private information of the firm. While the regulator observes realized cost, it does not observe the firm's type and effort separately. For the regulator (and for the econometrician as well), knowledge of the firm's type and the cost of exerting effort are essential in designing and evaluating policies.

Laffont and Tirole (1986, 1993) solve for the optimal mechanism that balances the tradeoff between efficiency and information rent extraction in this environment. They propose an indirect mechanism that implements the optimal mechanism. This indirect mechanism is a cost-reimbursement rule which specifies the (net) payment or transfer paid to the firm as a function of reported or observed cost. A net transfer function that is insensitive to costs provides little to no incentives to reduce cost while a sensitive net transfer function provides high-powered incentives. For example, the net transfer function that implements the optimal mechanism is strictly decreasing and convex.

The net transfer function has an analogous object in the rate case that I observe in the data.

[1] For example, concessions are given to firms located in states that heavily rely on the coal industry.

Using panel data on rate cases, I provide evidence for a net transfer function that is unresponsive to contemporaneous and past costs. I exploit this finding in my identification strategy. Since the net transfer function provides zero incentives to exert effort, fuel efficiency during the rate case reveals the distribution of firms' intrinsic types. Next, since the firm becomes the residual claimant during the regulatory lag, the difference in efficiencies during the rate case and during the regulatory lag reveals the level of effort that equates its marginal benefit with its marginal disutility. Inverting this relationship reveals the effort disutility function.

While a stringent policy may be optimal from the point of view of minimizing the sum of emissions damages and abatement costs (i.e. the *least cost* solution), the associated agency costs to implement such a policy may be too large. In this case, the optimal policy will generally be less stringent and involve emission levels that do not equate the marginal damages from emissions with the true (efficient) marginal cost of reducing them (Laffont and Tirole, 1993; Laffont, 1994; Lewis, 1996). Using the estimated distribution of types and disutility of effort, I evaluate welfare under different counterfactual pollution regulatory regimes, and quantify the tradeoff between efficiency and information rent extraction.

When there are no informational problems, the least cost solution is optimal. In contrast, when information rents have to be paid and these rents have welfare costs (e.g. shadow cost of public funds), the least cost solution is no longer optimal. Assuming a constant marginal damage from SO_2 emissions in the range of \$100/ton to \$600/ton, the total annual loss due to distortions in efficiency under the optimal mechanism amounts to \$74 million to \$113 million relative to the least cost solution.[2] However, the optimal mechanism provides *gains* amounting to \$725 million due to lower information rents. This implies *net* welfare gains of \$612 million to \$651 million from the optimal mechanism relative to the least cost solution.[3] If net welfare gains are fully passed on to consumers, this leads to a 2% reduction in the total revenue requirement that each state has to collect via electricity prices.

To implement the US Acid Rain Program (ARP), about \$600 million to \$1.8 billion worth of SO_2 emission permits were grandfathered to electric utilities for free instead of being auctioned (Joskow and Schmalensee, 1998). Because forgone auction revenues could have been "recycled" to replace

[2] Annual aggregate emissions are higher under the optimal mechanism (5.207 million tons) than in the least cost solution (5.128 million tons). With regard to a cap-and-trade regime, this means that the aggregate emissions cap under the optimal mechanism is less stringent than in the least cost solution, leading to damages worth \$8 million to \$47 million. Moreover, since the optimal mechanism reduces the equilibrium effort exerted by inefficient types relative to the least cost solution, annual aggregate abatement costs are \$66 million higher in the optimal mechanism.

[3] Interestingly, the least cost regime yields *lower* welfare when compared to a regime with rate regulation which provides poor efficiency incentives. In other words, the value of information rents under the least cost regime swamps the loss in efficiency from poor incentives.

distortionary taxes (Goulder et al, 1997; Cramton and Kerr, 2002), this type of redistribution carries welfare costs. The social planner ideally will offer compensation that leads to zero economic profits if there were no information asymmetries. In the case of asymmetric information, firms earn positive information rents. Assuming a constant marginal damage from SO_2 emissions of \$300/ton, implementing the least cost solution requires a total subsidy worth about \$1.5 billion (5.128 million tons $\times$ \$300/ton). The net welfare gains I estimate represent a reduction of 42% of this subsidy, implying that firms get almost half of the value of grandfathered permits as information rents in the least cost solution.

Related literature

The paper contributes to three main literatures. First, the paper contributes to the growing empirical literature on environmental regulation and is most related to recent papers studying the design and performance of market mechanisms to control pollution. An important departure of my paper is that I analyze the dimension of abatement costs that can be endogenously affected by regulation through incentive provision, and empirically study optimal pollution regulation in a setting with asymmetric information and distributional concerns (firms receive compensation). I emphasize the role of private information and participation constraints in pollution regulation, firms' strategic response to regulation, and the cost of providing incentives.

In terms of the empirical literature on environmental regulation, my paper is closest to Carlson et al (2000), Fowlie (2010), Fowlie et al (2014) and Fowlie and Muller (2013). Carlson et al (2000) estimate the cost-savings from a market of tradable emission permits and estimate marginal abatement costs from fuel-switching using a multiproduct cost function and similar data from the Acid Rain Program. However, Carlson et al (2000) ignore price regulation in estimating marginal abatement costs which is precisely what I exploit in my identification strategy. In contrast to Carlson et al (2000), Fowlie (2010) brings price regulation at the forefront of the analysis and provides evidence that price regulation induce firms to choose more capital-intensive abatement options in the context of NO_x emissions regulation–consistent with the well-known Averch and Johnson (1962) effect in the regulation literature. I look at a different dimension of abatement choice (efficiency in the fuel-burning process) and study a more subtle but important margin where price regulation can affect this choice. My analysis allows me to identify relevant structural parameters to study firms' strategic use of private information in response to regulation. This allows me to examine the tradeoff between efficiency and information rent extraction, and evaluate a welfare measure that takes into account incentive provision as opposed to only abatement cost-savings and emission reductions. Fowlie et al (2014) and Fowlie and Muller (2013) evaluate welfare under various counterfactual policies, although these papers do not deal with issues arising from both

informational and participation constraints. These papers do not study how firms use their private information strategically to extract information rents, and how different policies perform in light of this behavior.

Second, the paper contributes to the structural empirical regulation literature pioneered by Wolak (1994). Wolak (1994) and Brocas et al (2006) use the normative models of Baron and Myerson (1982) and Besanko (1985) to provide a link between observed behavior and the firm's private information. This approach assumes that the actual regulatory institution can be modeled "as if" the optimal form of regulation was being implemented by the regulator. The optimal mechanism characterizes a mapping between the firm's private information and observed regulatory variables (e.g. price and rate of return) which can then be inverted to identify and estimate the firm's primitives (Perrigne and Vuong, 2011). One issue with using a normative model is that it assumes a highly sophisticated regulator that can design and commit to the optimal mechanism.[4] Moreover, this approach precludes the researcher from asking normative questions since the optimal mechanism is already assumed to be implemented in the data. My approach instead is to model the actual regulatory institution to provide the link between observed behavior and the firm's primitives, similar to Gagnepain and Ivaldi (2002)[5] and Lim and Yurukoglu (2014)[6].

Lastly, the paper contributes to the empirical literature on the effects of regulation, specifically in the context of the electricity industry. Recent literature has focused on efficiency improvements from restructuring, e.g. Fabrizio et al (2007), Davis and Wolfram (2012), H. S. Chan et al (2013), Cicala (2014). Complementing this literature, my paper provides evidence of poor incentive provision during rate cases, but a positive incentive effect from regulatory lags. The pattern in fuel efficiency and cost that I find supports Joskow and Schmalensee's (1986) hypothesis that rate regulation need not be equivalent to a pure cost-of-service regime due to the positive incentive effects of regulatory

[4] Although Perrigne and Vuong (2011) allow observed regulatory variables to deviate from the one specified by the optimal mechanism, this deviation should be unsystematic, i.e. unrelated to the firm's primitives.

[5] Gagnepain and Ivaldi (2002) (see also Gagnepain et al (2013)) do not rely on a normative model and instead exploit variation in actual regulatory regimes to estimate welfare in the French urban transport industry. My paper differs from their identification strategy in two ways. First, the firms in their setting either face a fixed-price or a cost-plus contract. Under the assumption that the assignment to a regulatory regime is exogenous, the variation in regimes in the data allows identification of firms' type and disutility of effort. In my setting, firms face a more complicated regulatory regime in the form of a rate case combined with a regulatory lag. Second, I do not impose distributional assumptions on the type distribution. I establish nonparametric identification of the type distribution by applying the result of Kotlarski (1967) and estimate this distribution by extending the smoothed discrete approximation developed by Hausdorff (1923) and Beran and Hall (1992).

[6] Lim and Yurukoglu (2014) investigate investment incentives in rate regulation. My paper differs from their analysis by focusing on agency costs and its impact on emissions regulation, and by exploiting a different aspect of the regulation (i.e. interaction between the rate case and regulatory lag).

lags (Baumol and Klevorick (1970), Bailey and Coleman (1971), and Pint (1992)). Going beyond this literature, my paper illustrates how we can exploit firms' strategic response to rate regulation to provide a novel identification strategy for answering important and challenging policy questions.

Structure of the paper

The paper is organized as follows. The next section provides a brief overview of the regulatory institutions and the data. Section 3 focuses on rate regulation. I show a striking pattern in cost and fuel efficiency in the data which I then rationalize using a simple model of rate regulation. Section 4 is the main empirical section of the paper and discusses identification, estimation, and results. Section 5 contains the welfare analysis and the final section concludes.

2 Background and data

2.1 Regulatory environment

The electricity industry is arguably the single largest source of harmful emissions. In the US for example, the industry accounts for 38% of CO_2 emissions (Environmental Protection Agency, 2014) and 65% of SO_2 emissions (Environmental Protection Agency, 2001). Because of its significant contribution to pollution, the industry has constantly been the prime target of environment regulations. I focus on the time period covering the first phase (1995-1999) of SO_2 emissions regulation of electric utilities in the US under the Environmental Protection Agency's (EPA) Acid Rain Program (ARP). ARP is a federal cap-and-trade program that establishes a market for SO_2 emission permits. While generally lauded as a success (G. Chan et al, 2012), the legislative history of ARP illustrates that implementation of the program largely hinged on the ability to redistribute the benefits of abatement and compensate affected polluting sources via freely allocated initial permits (Joskow and Schmalensee, 1998; Ellerman et al, 2000 Ch 3; G. Chan et al, 2012; Schmalensee and Stavins, 2012). Joskow and Schmalensee (1998) estimate the value of these free permits to be about \$600 million to \$1.8 billion.

The electric utilities covered by the ARP primarily rely on coal to produce electricity: the average ratio of coal consumption to total fuel burned is about 92%. Coal contains sulfur and SO_2 is released to the atmosphere as a by-product. Sulfur content ranges from about 0.2 pounds per heat input (lbs/MMBtu) to about 7 lbs/MMBtu (Perry et al, 1997). There is a tradeoff between heat and sulfur content: bituminuous coal tends to have a higher heat content but also high sulfur content compared to sub-bituminuous coal. Hence, absent pollution regulation, plants tend to burn

coal with higher sulfur content.[7]

Two primary forms of reducing SO_2 emissions are fuel-switching and installation of a flue-gas desulfurization (FGD) unit, also known as a scrubber. Fuel-switching involves using coal with lower sulfur content or blending different types of coal with varying sulfur contents. In contrast, a plant can install an FGD which is an end-pipe control technology installed near the plant's emission stacks. The plant can still burn high sulfur coal, and the FGD will "scrub" SO_2 from the emissions stream. Compared to installing an FGD, fuel-switching was the more popular abatement method during my sample period (1988-1999). In my sample, there are only 15 plants out of about 150 that newly-installed an FGD. Plants with FGDs represent only 20% of all the plants. This number includes plants that installed FGDs to satisfy SO_2 regulations that were in place before the ARP. The share of abatement from fuel-switching during this period ranged from 54% to 60% (Ellerman and Montero, 2007, Table 5).

Fuel-switching directly affects the marginal cost of producing electricity. Lower sulfur coal produces less heat, and more coal has to be burned to produce an additional quantity of electricity. I focus on fuel-switching as an abatement strategy and measure marginal abatement cost as the increase in the cost of producing electricity for an incremental reduction in emission rates. By focusing on fuel-switching as an abatement strategy, I can look at an electric utility's fuel efficiency and the impact of reducing emission rates on the cost of producing electricity. The key empirical challenge, then, is to identify how asymmetric information and pollution regulation affect fuel efficiency. I exploit a second form of regulation that electric utilities were facing to address this challenge.

Against the backdrop of SO_2 emissions regulation, electric utilities were also facing state-level price regulation in the form of *rate regulation*.[8] While electric utilities were allowed to operate as vertically-integrated monopolists, prices were set by the utility regulator based on information about cost and operations that the firm provides in a quasi-judicial proceeding called the *rate case*. The rate case typically starts with the firm submitting a proposed revenue requirement, which is the amount of revenues the firm is authorized to collect, and from which prices will be based on. The

[7] Distance of the plant from coal mines is another factor that determines coal choice since transportation costs are a significant component of delivered prices. The dirtiest plants in terms of SO_2 are those that are located far from sources of lower sulfur coal. Rail deregulation and falling delivered prices of sub-bituminuous coal from the Powder River Basin (PRB) made this type of coal more competitive. However Ellerman et al (2000, p. 89) note that although the competitiveness of PRB coal led to an overall decrease in contracted prices of coal, long-term contracts continued delivering high sulfur coal.

[8] Almost all power plants under the first phase of ARP were owned and operated by regulated electric utilities during this period. Despite recent attempts to restructure the electricity industry, the share of net generation and emissions of plants owned by electric utilities was still over 85% in 2005 among large plants covered by the first and second phases of the ARP.

regulator then examines the proposal, reviews evidence, hears arguments, and finally decides on the revenue requirement to authorize. Once the rate case concludes, prices based on the authorized revenue requirement remain essentially fixed until the next rate case. This period between two consecutive rate cases is referred to as a *regulatory lag*.[9] The rate case and the regulatory lag are the main features of rate regulation that I focus on in the paper.

2.2 Data

My sample consists of electric utilities which have generating units included in Phase I of the EPA's Acid Rain Program. I use operations and fuel data from the Energy Information Administration's (EIA) Form 767 and the Federal Energy Regulatory Commissions' (FERC) Form 423. I combine these with financial and rate case data from SNL Financial and the Regulatory Research Associates (RRA). Details on data construction can be found in the online appendix.

Table 1: Summary statistics of operations and costs data

Variable		Mean	Std dev	Min	Max
O&M var cost	\$M	328	253	23	1198
Net generation	MwH	2.4×10^7	2.2×10^7	558739	9.9×10^7
Ave O&M cost / net gen	cents / kwH	1.8	0.9	0.5	5.4
Emission rate	lbs/MMBtu	1.82	1.04	0.38	7.22
Nameplate	MW	5010	4897	232	23227
FGD dummy	FGD dummy	0.33	0.47	0	1
Salary	\$000/emp/mo	15.8	8.2	.6	52.3
Price coal	\$/ton	33.11	9.93	12.48	53.80
Price oil	\$/barrel	22.53	4.94	10.06	37.38
Price gas	\$/MMBtu	2.96	0.97	1.38	15.48

Table 1 contains summary statistics for these utilities. The number of firm-year observations is 351. The operating and maintenance (O&M) variable cost measure is the sum of fuel expense and non-fuel O&M expense related to electricity generation. Fuel expense accounts for about 75% of O&M expense, on average. Moreover, on average, coal accounts for about 93% of total fuel consumption (in MMBtu) while about 5% and 3% for oil and natural gas respectively. Average O&M cost per kilowatt-hour of net generation is about 1.8 cents

Table 2 contains rate case summary statistics for these utilities. On average, a rate case lasts

[9] The term regulatory lag is sometimes also used to refer to the duration of the rate case. I use the term to refer to the time between two consecutive cases.

Table 2: Summary statistics of rate case data

Variable		Mean	Std dev	Min	Max
Rate case duration	Years	1.2	0.7	0	3
Regulatory lag	Years	2.3	1.9	0	6
Percent disallow Rev Req	% of Prop Rev Req	4	3	0	12
Proposed RRB	$M	320	406	7	1868
Authorized RRB	$M	296	374	7	1617
Percent disallow RRB	% of Prop RRB	7	4	0	31
Proposed rate base	$M	2580	3272	73	15963
Authorized rate base	$M	2453	3090	66	14485
Proposed ROR	%	10.2	0.9	7.9	12.2
Authorized ROR	%	9.8	1.0	7.4	11.8

just over a year and can extend for 3 years. The number of years from the time a rate case is authorized to a new rate case is proposed (i.e. the regulatory lag) is 2, on average, but can be as long as 6 years. Majority of utilities in my sample experienced no more than 3 rate cases during 1988-1999. On average, about 4% of the proposed revenue requirement is disallowed during the rate case. This disallowance ranges from no disallowance to 12%.

One of the main instruments of rate regulation is the *return on the rate base* (RRB). RRB is the return that the utility gets from its investment, net of operating cost. It is the product of the monetary value of its assets, which is called the rate base, and a rate-of-return (ROR). The firm proposes the rate base and ROR at the beginning of the case, and the regulator authorizes a rate base and ROR at the end. The average RRB disallowance, i.e. difference between proposed RRB and authorized RRB, as measured as a percentage of proposed RRB is 7%, and ranges from 0%, i.e. no disallowance, to as high as 31%. The average RRB disallowance as expressed as a percentage of the total disallowance in the revenue requirement is 80%. Finally, proposed ROR has a mean of 10.2% while authorized ROR has a mean of 9.8%.

3 Rate regulation

The traditional form of price regulation is rate regulation[10], which is conducted via a quasi-judicial proceeding called a rate case. The primary goal in these rate cases is to set the revenue requirement– the total amount to be collected from consumers to compensate the firm for providing electricity.

[10]Rate-of-return regulation and cost-of-service regulation are other names used for this form of regulation.

The revenue requirement is the sum of operating expenses and a return on the assets of the firm. This return is equal to the rate base, which is the value of the firm's investment, multiplied by a rate of return.

The rate case serves as a platform for the firm to provide information about its operating cost and environment to the regulator (public utility commission or PUC), who then decides on what revenue requirement to authorize. The case is typically initiated by the firm although the regulator, urged by consumer groups, can also initiate a case. A hearing takes place where the firm and concerned parties (e.g. consumer interest groups) participate and provide testimony on the rationale of the proposed changes and the potential impacts these may have on consumer welfare. The firm, consumer groups, and the commission staff testify to support their position and to refute opposing arguments. A discovery phase also occurs where data and evidence are presented. If a settlement between concerned parties is not reached, the PUC commissioners decide on the case. The decision consists of the approved revenue requirement which often differs from the initial proposal of the firm (difference is about 4% of proposed revenue requirement, on average)

In theory, the debate and disagreement in rate cases revolve around the three components of the revenue requirement: operating expenses, the rate base, and the rate of return. In practice, major rate cases focus on the determination of the rate base and the rate of return. Reported operating expenses are often passed through as long as these abide established accounting rules (Alt, 2006). For example, a typical expense that is disallowed concerns depreciation of the firm's fixed assets and stems from an improper application of accounting depreciation rules. In terms of the rate base, the PUC may disallow certain assets if they do not satisfy the "used and useful" criterion. For example, the regulator may disallow a firm's investment in plant expansion if this leads to significant overcapacity, *ex-post*.

Finally, the PUC is required to provide a "fair" rate of return on the firm's investment, which ideally should be enough to cover the firm's opportunity cost. Rate cases often involve lengthy debates on how to compute this rate of return. One aspect of the determination of the rate of return that is often overlooked is that the PUC can *potentially* use the authorized rate of return as an incentive device. For example, in a 1990 rate case involving the Florida PUC and Gulf Power, additional reductions to the authorized rate of return were imposed for a two-year period due to "unethical or illegal" activities.

Once the revenue requirement has been determined and authorized, electricity rates are then set for different customer classes. Rates essentially remain fixed until the next rate case, except for some minor adjustments. These adjustments appear as surcharges in a customer's bill and are triggered by an unexpected but observable rise in input cost, e.g. fuel surcharge due to an increase in the price of coal or oil. However, the price of the most relevant fuel (coal) during this time period

remained flat so was unlikely to trigger a significant adjustment. Thus, at least for my sample time period, there were regulatory lags between successive rate cases. During these regulatory lags, rates were fixed and unresponsive to movements in cost driven by factors that are not directly observable.

How does the rate case and regulatory lag affect operating cost and fuel efficiency? Table 3 contains results from regressions of the log of O&M variable costs on the log of output[11] (electricity and emissions), input prices (labor, coal, oil and gas) and capital (nameplate and indicator if the firm has a scrubber), together with indicator variables for whether the observation comes from years when the rate case is ongoing. I construct three indicator variables. The first dummy is equal to one if the observation occurs during the rate case, i.e. from proposed to authorized year, inclusive. The second dummy is equal to one if the observation occurs on the year immediately after the authorization year. Finally the third dummy is equal to one if neither of the two dummies are one. In the regression, the omitted dummy category is the second dummy so dummy coefficients measure the % difference relative to the year after the rate case concludes.

Focusing on the estimates for the rate case dummy, we see that average O&M variable costs are 4% to 6% higher during a rate case compared to the year after, where the regulatory lag takes effect. Moving to the "neither" dummy coefficient estimate, we find no statistically significant differences in O&M variable costs among non-rate case years, which are the years during the regulatory lag. These results hold even when controlling for output, input prices, capital, and year effects, and also when looking at within firm and within firm-rate case variation.

To investigate this further, I examine whether a similar pattern arises for heat rates, which is the amount of fuel burned per unit of electricity produced.[12] I regress the log of heat rate on the log of electricity generated, log of capital, indicator for FGD, and the rate case dummies. State electricity demand is used as an instrument for generated electricity to account for potential simultaneity bias (Fabrizio et al, 2007).[13] Table 4 contains results of this regression.

Consistent with my earlier finding, heat rates are about 4% to 6% higher during rate cases relative to the year after. Interestingly, there are no statistically significant differences in heat rates when I compare the year after the rate case, and the succeeding non-rate case years. Thus, the decrease in fuel efficiency during the rate case is only temporary and disappears right after the case

[11] I include specifications where I use state-level electricity demand as an instrument for electricity output and regional prices for low and high sulfur coal as instruments for emission rates. Low sulfur coal is defined as coal with sulfur content below 1.2 lbs/MMBtu, while high sulfur coal is defined as coal with sulfur content above 3 lbs/MMBtu First stage F-statistics are 164 and 27 for electricity output and emission rates respectively.

[12] One story that can explain the pattern in operating cost suggested by the regressions in table 3 is that firms strategically initiate rate cases when cost is known or anticipated to be temporarily high in order to lock-in higher rates. Thus, this pattern is consistent with asymmetric information on *exogenous* fluctuations on cost, and the main endogenous decision of the firm is just the timing of the rate case. Evidence on heat rates reveal a different story.

[13] First stage F-statistic is 168.

Table 3: Regression results: O&M variable cost and rate case dummies.

log O&M var cost	(1)	(2)	(3)	(4)	(5)
Rate case	0.037	0.054**	0.056***	0.052**	0.052**
	(0.030)	(0.024)	(0.019)	(0.020)	(0.020)
Neither Rate case	−0.001	−0.006	−0.024	−0.023	−0.023
nor Year after	(0.067)	(0.018)	(0.025)	(0.027)	(0.027)
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	No	No
Firm-Rate Case FE	No	No	Yes	Yes	Yes
IV for electricity	No	No	No	Yes	Yes
IV for emission rate	No	No	No	No	Yes
Num. Obs.	351	351	351	293	293

Notes: Standard errors are either clustered at firm or firm-rate case level. Regression via OLS except when indicated. Additional regressors are a dummy for FGD; the logs of electricity output, emission rate, input prices (labor, coal, oil and gas), and nameplate rating. I use log of state electricity demand as an IV for electricity output and regional prices for low ($<$ 1.2 lbs/MMBtu) and high ($>$ 3 lbs/MMBtu) sulfur coal for emission rates. Significance level: * 10%, ** 5%, *** 1%.

(during the regulatory lag).

Short-run variations in heat rates are more likely due to temporary changes in how a firm operates its plants. A higher heat rate means less efficient production since the firm burns more fuel to produce the same amount of electricity. The evidence on heat rates suggests that the pattern in operating cost is likely driven by a firm's *endogenous* "effort" to improve operating efficiency and reduce costs, rather than by exogenous factors.[14]

In order to understand how rate regulation drives this pattern in operating cost and fuel efficiency, we need to put more structure on how rate regulation affects incentives of firms. This structure also allows us to estimate primitives relating to firms' cost, and use these to perform

[14] I used a firm-level measure of heat rates in the regressions in table 4. Firm-level heat rate is a function of individual plant's heat rates and the quantity of electricity produced by each plant. Thus, endogenous effort can either reflect a reallocation of output from less efficient plants to more efficient ones, or a reduction in an individual plant's heat rate, for example, through optimization of combustion and sootblowing processes (Bushnell and Wolfram, 2007). I repeat the heat rate regressions at the plant-level and find support for the view that effort reflects a temporary reallocation of output across plants rather than a reduction of individual plants' heat rates.

Table 4: Regression results: Heat rates and rate case dummies

log heat rate	(1)	(2)	(3)	(4)
Rate case	0.065***	0.051**	0.046*	0.042**
	(0.024)	(0.020)	(0.026)	(0.020)
Neither Rate case	0.003	−0.014	−0.016	−0.012
nor Year after	(0.020)	(0.035)	(0.041)	(0.033)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	No	Yes	No
Firm-Rate Case FE	No	Yes	No	Yes
IV for electricity	No	No	Yes	Yes
Num. Obs.	351	351	293	293

Notes: Standard errors are clustered at firm level. Regression via OLS except when indicated. Additional regressors include a dummy for FGD and the logs of electricity output and nameplate. I use log of state electricity demand as an IV for electricity output. Significance level: * 10%, ** 5%, *** 1%.

counterfactual policy experiments.

3.1 Regulatory incentives

3.1.1 Theory

The main object in a rate case is the revenue requirement which I denote as RR. It is the sum of operating cost, $\widetilde{C}$, and the return on the rate base, $\bar{R}$:

$$RR = \widetilde{C} + \bar{R} \tag{1}$$

The return on the rate base $\bar{R}$ is a utility's profit over and above its operating costs, i.e. a *net transfer*, and is equal to the monetary value of the firm's capital (rate base) multiplied by an allowed rate of return. My focus is on how $\widetilde{C}$ influences $\bar{R}$ and I do not explicitly model the rate base and the rate of return.

I assume the regulator observes operating cost $\widetilde{C}$. However, operating cost is a function of an unobserved type θ that determines the firm's intrinsic fuel efficiency, and an unobserved effort $e \geq 0$ which reflects activities that improve efficiency. Specifically, operating cost is given by

$$\widetilde{C}(\theta) = \exp(\theta - e)\, C(q, s) \tag{2}$$

where q is the quantity of electricity produced and s is a measure of sulfur emissions. Cost is increasing in θ and decreasing in e. The firm's intrinsic type θ is an independent draw from the distribution $\mathcal{F}$. The firm reduces its cost by exerting effort, and this incurs a disutility given by $\psi(e)$, which is strictly increasing and convex.

The cost and information structure follows the standard model of Laffont and Tirole (1986). One can map the revenue requirement to their model and solve for the mechanism that maximizes social welfare

$$W = \int \{V(q(\theta)) - D(s(\theta)) - (1+\lambda)RR(\theta) + \Pi(\theta)\}\, d\mathcal{F}(\theta) \tag{3}$$

subject to incentive compatibility and participation constraints (Laffont, 1994), where $V(q)$ is the surplus from consuming electricity, $D(s)$ is the damage from emissions, λ is the cost of public funds and $\Pi(\theta)$ is firm's profit given by

$$\Pi(\theta) = RR(\theta) - \left[\widetilde{C}(\theta) + \psi(e(\theta))\right]. \tag{4}$$

As shown in Laffont and Tirole (1993), the optimal mechanism yields an optimal net transfer function $\bar{R}^*(\widetilde{C})$ that is strictly decreasing and convex in the observed operating cost. The optimal mechanism can thus be implemented by a menu of cost-reimbursement rules.

For now, suppose we have a one-shot interaction between the regulator and the firm. That is, once the rate case concludes, the game ends. In this case, whether rate regulation provides incentives to the firm to exert effort depends on how $\widetilde{C}$ affects the authorized return $\bar{R}$. For a given (q, s) and $\bar{R}(\cdot)$, a firm of type θ chooses effort to maximize its profit:

$$RR - (\widetilde{C} + \psi(e)) = \bar{R}(\widetilde{C}) - \psi(e) \tag{5}$$

subject to $\widetilde{C} = \exp(\theta - e)\, C(q, s)$. The following first order condition is necessary for a strictly positive level of effort:

$$-\bar{R}'(\widetilde{C}) \cdot \exp(\theta - e)\, C(q, s) = \psi'(e). \tag{6}$$

Since $\psi'(e) > 0$, then $\bar{R}'(\widetilde{C}) < 0$ is necessary to induce a positive level of effort. The optimal mechanism involves a strictly decreasing $\bar{R}^*(\widetilde{C})$ but any other strictly decreasing $\bar{R}(\cdot)$ may induce some other level of positive effort. Corollarily, a sufficient condition for *zero effort* in this one-shot game is for $\bar{R}'(\cdot) \geq 0$.

Actual rate regulation is not a one-shot game that ends with a single rate case. After the rate case, the revenue requirement remains fixed until the next rate case. The reduced-form results on operating cost and heat rates suggest an important role played by the interaction of the rate case with the regulatory lag.

Consider the following three-period model. In period $t = 1$, a rate case is held. Once the rate case concludes, we enter period $t = 2$, which is the regulatory lag. Finally period $t = 3$ is another

rate case and the game ends once the case concludes. Let RR_t and $\bar{R}_t$ be the authorized revenue requirement and return on the rate base in period t, respectively. In general, $\bar{R}_t$ can be a function of current and past costs.

To model the regulatory lag, assume that in period 2, the regulator commits to $RR_2 = RR_1 = \widetilde{C}_1 + \bar{R}_1(\widetilde{C}_1)$, and so the revenue requirement is determined during the previous rate case ($t = 1$) and cannot be changed. Finally, we allow the return in period 3 to depend on current and past cost, i.e. $\bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3)$, hence potentially exhibiting ratchet effects (Laffont and Tirole, 1993). Thus in period 3, $RR_3 = \widetilde{C}_3 + \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3)$.

Given (q, s) and $\bar{R}_t$'s, a firm of type θ chooses (e_1, e_2, e_3) that solve

$$\max \left[\bar{R}_1(\widetilde{C}_1) - \psi(e_1)\right] + \left[\hat{C}_1 + \bar{R}_1(\widetilde{C}_1) - \hat{C}_2 - \psi(e_2)\right] + \left[\bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3) - \psi(e_3)\right] \tag{7}$$

where $\widetilde{C}_t = \exp(\theta - e_t)\, C(q, s)$ for $t = 1, 2, 3$. The following are necessary conditions for an interior optimum:

$$-\left(2\frac{\partial \bar{R}_1(\widetilde{C}_1)}{\partial \widetilde{C}_1} + \frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3)}{\partial \widetilde{C}_1} + 1\right) \cdot \exp(\theta - e_1)\, C(q, s) = \psi'(e_1) \tag{8}$$

$$-\left(\frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_2)}{\partial \widetilde{C}_2} - 1\right) \cdot \exp(\theta - e_2)\, C(q, s) = \psi'(e_2) \tag{9}$$

$$-\frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_2)}{\partial \widetilde{C}_3} \cdot \exp(\theta - e_3)\, C(q, s) = \psi'(e_3) \tag{10}$$

The following proposition follows directly from equations 8, 9 and 10:

Proposition 1 *If*

$$\frac{\partial \bar{R}_1(\widetilde{C}_1)}{\partial \widetilde{C}_1}, \frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_2)}{\partial \widetilde{C}_1}, \frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_2)}{\partial \widetilde{C}_3} \geq 0, \tag{11}$$

$$\frac{\partial \bar{R}_3(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_2)}{\partial \widetilde{C}_2} = 0 \tag{12}$$

then $e_1^* = 0$, $e_2^* > 0$ *and solves*

$$\exp(\theta - e_2^*)\, C(q, s) = \psi'(e_2^*) \tag{13}$$

and finally, $e_3^* = 0$.

Zero effort during the rate case arises since the net transfer function does not punish the firm based on observed operating cost in the current rate case and on cost from past rates cases. What this means is that there is effectively full pass-through of operating costs into the revenue requirement, hence providing no incentives to reduce cost. However, once the rate cases finishes,

rates are fixed and so the firm becomes the residual claimant to cost reductions during the regulatory lag. Proposition 1 also says that costs during the lag do not affect future authorized returns, and thus information on operating cost revealed in the past are not used by the regulator. This implies that effort during the regulatory lag will be set to minimize the sum of current operating cost and disutility of effort.

Proposition 1 provides sufficient conditions for the patterns in cost and heat rates found in the last section to be consistent with (i) zero effort during rate cases, and (ii) a positive level of effort during regulatory lags that equates the marginal benefit of effort with its marginal disutility. This interpretation of the reduced-form results on cost and efficiency will be important for identification of the distribution of exogenous types and the disutility of effort. Observed efficiency during rate cases reveals exogenous efficiency (θ) since effort is zero while observed efficiency during the regulatory lag reflect the combination of exogenous efficiency and an effort level that equates the marginal benefit from effort with its marginal cost (disutility).

3.1.2 Evidence

The goal of this section is to show that the sufficient conditions in proposition 1 hold in the data. Since I observe the authorized return on the rate base in the data (which directly maps to the net transfer function in the Laffont and Tirole (1986, 1993) model), I can directly check whether the conditions in proposition 1 are satisfied. I regress the log of authorized return on the rate base on the logs of O&M variable cost, lagged O&M variable cost, O&M cost from the previous rate case, electricity output, emissions rate and nameplate. Following a companion paper (Abito, 2014), I also include the *proposed* return on the rate base in the regression. In that paper, I model the rate case as a signaling game of auditing and provide evidence that disallowances in the return on the rate base is sensitive to the proposed return but not on cost. This is consistent with a fully separating equilibrium where the proposed return is used as a signal but not operating cost. Here, I focus on the authorized return on the rate base as a net transfer function a la Laffont and Tirole (1986, 1993), and examine its relationship with respect to current and past operating costs.

Table 5 contains the regression results. The first model (I) only includes operating cost during the current rate case. The coefficient on cost implies that a 1% increase in observed cost during the case translates to (only) a 0.02% *increase* in the authorized return. This effect is not statistically different from zero. This is consistent with $\frac{\partial \bar{R}_1}{\partial \tilde{C}_1} = \frac{\partial \bar{R}_3}{\partial \tilde{C}_3} = 0$. Next, the second model (II) includes lagged operating cost, which is the cost during the regulatory lag. Again, the effect is small and positive, but not statistically significant, consistent with $\frac{\partial \bar{R}_3}{\partial \tilde{C}_2} = 0$. Finally, the third model (III) also includes cost from the previous rate case. While coefficients in this model are all negative, these are small in magnitude and not statistically significant. For example, a 1% increase in operating cost

in the current rate case translates to less than 0.01% reduction in the authorized return on the rate base, or a pass-through rate reduction of less than 1/100 of a percent. Cost from the previous rate case has no statistically significant effect on the authorized return, hence consistent with $\frac{\partial \bar{R}_3}{\partial \widetilde{C}_1} = 0$.

Table 5: Relationship between authorized return on the rate base and operating costs

log authorized return on rate base $\bar{R}$	I	II	III
O&M var cost $\widetilde{C}$	0.024	0.043	-0.003
	(0.037)	(0.034)	(0.009)
Lag $\widetilde{C}$	.	0.008	-0.002
	(.)	(0.037)	(0.011)
$\widetilde{C}$ in last rate case	.	.	-0.077
	(.)	(.)	(0.065)
Proposed return	0.623***	0.590***	0.523***
	(0.188)	(0.216)	(0.042)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Num. Obs.	141	128	60

Notes: Standard errors are clustered at firm level. Authorized return on rate base $\bar{R}$ and expense $\widetilde{C}$ are in logs. Regressions only use data from rate case years except for lagged O&M cost. Additional regressors include logs of electricity output, emissions rate and nameplate. Significance level: * 10%, ** 5%, *** 1%.

4 Empirical model

I estimate a multiproduct cost function to provide a measure of (marginal) abatement costs for electric utilities. The cost of reducing emissions is measured as the increase in the cost of producing electricity due to changes in production methods, i.e. fuel-switching. I restrict attention to costs, output, emissions and input choices related to coal, oil and gas plants. The analysis is done at the utility-level since rate regulation and rate cases involve the firm as a whole. Moreover, Ellerman et al (2000, p. 301) remark that compliance decision-making is often made at the utility-level even if pollution regulation *per se* is at the unit-level.

I assume the following stochastic specification for realized O&M variable costs of producing

electricity and emissions. For firm i at year t, realized O&M variable cost is given by

$$\widetilde{C}_{it} = \exp(\omega_{it}) C(q_{it}, s_{it}, p_{lit}, p_{fit}, N_{it}, d_{FGDit}, t; \beta) \exp(\varepsilon_{it}) \tag{14}$$

where

$$\begin{aligned}
\omega_{it} &= \theta_{it} - e_{it} \\
p_{fit} &= (p_{cit}, p_{oit}, p_{git}) \\
\beta &= (\beta_0, \beta_N, \beta_{FGD}, \beta_{YEAR}, \beta_q, \beta_s, \beta_{sd}, \beta_l, \beta_c, \beta_o, \beta_g) \\
C(q, s, p_l, p_f, N, d_{FGDit}, t; \beta) &= \exp(\beta_0 + \beta_{YEAR} t + \beta_{FGD} d_{FGDit}) N^{\beta_N} q^{\beta_q} s^{\beta_s + \beta_{sd} d_{FGD}} p_l^{\beta_l} p_c^{\beta_c} p_o^{\beta_o} p_g^{\beta_g}.
\end{aligned}$$

The term $\exp(\omega) = \exp(\theta - e)$ captures cost efficiency where θ is the firm's intrinsic type and e is unobserved effort. The utility knows θ and chooses e. The function $C(q, s, p_l, p_f, N, d_{FGD}, t; \beta)$ is the *baseline cost function* of the utility where q is net generated electricity, s is the SO_2 emission rate, p_l is the average salary for full-time employees related to electricity generation, p_f is a vector composed of fuel prices[15] for coal, oil and gas, averaged across the utility's plants, N is the sum of nameplate ratings of the utility's plants, and d_{FGD} is a dummy equal to one if the utility has at least one plant with a flue-gas desulfurization (scrubber) unit installed. This baseline cost captures differences in O&M costs that can be explained by differences in input prices, outputs and capital. The vector β contains the parameters of the baseline cost function (including a constant β_0) that need to be estimated. Finally ε is a mean zero stochastic error term that summarizes other factors that affect realized costs. I assume ε is unanticipated by the firm when making its input choices and uncorrelated with the regressors.

I assume the firm's intrinsic type θ_{it} is a draw from the distribution $\mathcal{F}_\theta$. Ideally $\mathcal{F}_\theta$ would be conditioned on variables such as firm's capacity or portfolio of plants, but to make estimation more tractable later, I assume $\mathcal{F}_\theta$ is only a function of the rate case year. Next, the reason why the firm's unobserved type is indexed by t is that I allow θ to change across different rate cases. However I assume that θ remains constant within two consecutive rate cases. Thus, I assume that θ is firm-rate case specific. Let t_τ be the time index (year) for a specific firm's rate case τ. For example, if firm i has three rate cases during the sample period, then $\tau \in \{1, 2, 3\}$ which occurs on years t_1, t_2 and t_3 respectively. Formally, for all i, t and τ,

$$\theta_{it} = \begin{cases} \theta_{it_\tau} & \text{if } t \in [t_\tau, t_{\tau+1}) \\ \theta_{it_{\tau+1}} & \text{if } t = t_{\tau+1}. \end{cases}$$

[15] Fuel prices are either spot or contracted prices. Managerial effort (captured in ω through e) can affect the actual price the firm faces and hence can be an additional source of endogeneity (Cicala, 2014).

I assume for each firm i, θ_{it_1} is a draw from $\mathcal{F}_\theta$, θ_{it_2} is a draw from $\mathcal{F}_{\theta|\theta_{it_1}}$, θ_{it_3} is a draw from $\mathcal{F}_{\theta|\theta_{it_2}}$, etc.

In the next subsection, I discuss identification of the distribution of types $\mathcal{F}_\theta$, the disutility function $\psi(\cdot)$ and the baseline cost function parameters β.

4.1 Identification

There are two interrelated challenges for identification. The first challenge is an endogeneity problem in identifying the cost parameters β. The second challenge involves extracting the distribution of the unobserved type θ from the variation in realized costs that is unobserved by the econometrician, i.e. $\exp(\theta_{it} - e_{it})\exp(\varepsilon_{it})$. The first challenge arises precisely because e_{it} is an endogeneous variable chosen by the firm. Variables that enter the firm's baseline cost affect the level of effort to exert since baseline cost captures the cost reductions from effort. The firm's cost efficiency ($\omega = \theta - e$) in turn affects costs, electricity output (since regulated electricity prices are based on reported expenses), and potentially, input prices (Cicala, 2014). If ω_{it} were observed by the econometrician, then we can directly control for it, and identify the vector β. However ω_{it} is not observed.

Once we have identified the vector of baseline cost parameters β, we then need to identify the distribution of θ. This distribution is a critical input in the counterfactual analysis since this determines the degree of heterogeneity in marginal abatement costs and hence, the size of the gains from the optimal mechanism. The second challenge then, is to extract the distribution of θ from the unobserved variation $\exp(\theta_{it} - e_{it})\exp(\varepsilon_{it})$.

My identification strategy involves two parts. First, to identify the parameters of the empirical model, I use Proposition 1 to pin down ω_{it} for different time periods. This allows me to take different transformations of the data to eliminate ω_{it} from the estimating equations. Second, to identify the distribution of intrinsic types θ, I recast the problem under the framework of measurement error with repeated measurements (e.g. Li and Vuong (1988)) and use the deconvolution result of Kotlarski (1967).

4.1.1 Identification of parameters

If the sufficient conditions in Proposition 1 hold, then the firm does not exert effort during the rate case. Hence $\omega_{it_\tau} = \theta_{it_\tau}$. After the rate case, i.e. at time $t = t_\tau + 1$, the firm exerts effort such that

$$\psi'(e_{it_\tau+1}) = \exp(\omega_{it_\tau+1})\, C(q_{it_\tau+1}, s_{it_\tau+1}, p_{lit_\tau+1}, p_{fit_\tau+1}, N_{it_\tau+1}, d_{FGDit_\tau+1}, t_\tau + 1; \beta),$$

i.e. the marginal disutility of exerting effort is equal to the marginal cost reduction. To determine what ω_{it} is after the rate case, I impose the following functional form[16] for $\psi(\cdot)$:

Assumption 1 *The disutility of effort is given by*

$$\psi(e_{it}, \upsilon_{it}) = \frac{1}{\gamma}\exp(\gamma e_{it} + \upsilon_{it}) - \frac{1}{\gamma}$$

where γ is a parameter and υ_{it}'s are mean zero shocks that are uncorrelated with

$$(q_{it}, s_{it}, p_{lit}, p_{fit}, N_{it}, d_{FGDit}, t)$$

and iid across i and t.

Assumption 1 allows me to express ω_{it} as a linear function of θ_{it}, the log of the baseline cost function $C_{it}(\beta) = C(q_{it}, s_{it}, p_{lit}, p_{fit}, N_{it}, d_{FGDit}; \beta)$ and the shock υ_{it}:

$$\omega_{it} = \frac{1}{1+\gamma}(\gamma\theta_{it} - \ln C_{it}(\beta) + \upsilon_{it}) \tag{15}$$

for $t = t_\tau + 1$. Proposition 1 and the assumption that θ_{it} is constant within rate cases give expressions for realized costs during different "events":

$$\ln \widetilde{C}_{it_\tau} = \theta_{it_\tau} + \ln C_{it_\tau}(\beta) + \varepsilon_{it_\tau} \tag{16}$$

$$\ln \widetilde{C}_{it_\tau+1} = \frac{\gamma}{1+\gamma}(\theta_{it_\tau} + \ln C_{it_\tau+1}(\beta)) + \frac{1}{1+\gamma}\upsilon_{it_\tau+1} + \varepsilon_{it_\tau+1} \tag{17}$$

for all rate cases τ. The first line is the realized cost during rate cases, while the second line is for the year after the case.

Although θ_{it} is constant within rate cases for each firm i, I allow θ_{it} to vary across rate cases. I assume that θ_{it} follows a linear process across two rate cases :

Assumption 2 *For each i and τ, intrinsic types across two rate cases τ and $\tau-1$ evolve according to*

$$\theta_{it_\tau} = \rho\theta_{it_{\tau-1}} + \xi_{it_\tau}$$

where ρ is a parameter and ξ_{it_τ}'s are iid across i and t_τ.

Remark 1 I do not include a constant in the specifications for the disutility function $\psi(\cdot)$ and the evolution of types across rate cases. The reason is that these are not identified when I include a constant β_0 in the baseline cost function. That is, the means of ξ and υ are subsumed in β_0.

[16] Gagnepain and Ivaldi (2002) uses a similar exponential form for the disutility function.

Assumption 2 provides a way to difference out cost efficiency ω_{it}. Using assumption 2, I can quasi-difference equation (16) for two consecutive rate cases. This yields

$$\ln \widetilde{C}_{it_\tau} - \rho \ln \widetilde{C}_{it_{\tau-1}} = \ln C_{it_\tau}(\beta) - \rho \ln C_{it_{\tau-1}}(\beta) + \eta_{1it_\tau}$$

where

$$\eta_{1it_\tau} = \xi_{it_\tau} + \varepsilon_{it_\tau} - \rho \varepsilon_{it_{\tau-1}} \equiv \eta_{1it_\tau}(\beta, \rho).$$

I construct moment conditions

$$E\left[\eta_{1it_\tau}(\beta, \rho) \cdot z_{it_{\tau-1}}\right] = 0 \tag{18}$$

where $z_{it_{\tau-1}} = \left(q_{it_{\tau-1}}, s_{it_{\tau-1}}, s_{it_{\tau-1}} d_{FGDit_{\tau-1}}, p_{lit_{\tau-1}}, p_{fit_{\tau-1}}, N_{it_{\tau-1}}, d_{FGDit_{\tau-1}}, t_{\tau-1}\right)'$. These moment conditions hold because ξ_{it_τ} and ε_{it_τ} are iid across t, and $\varepsilon_{it_{\tau-1}}$ is an unanticipated shock during $t_{\tau-1}$.

Another way to difference out ω_{it} is by looking at observations during and after the rate case. Specifically, consider the following quasi-difference across $t_{\tau+1}$ and t_τ:

$$\ln \widetilde{C}_{it_\tau+1} - \frac{\gamma}{1+\gamma} \ln \widetilde{C}_{it_\tau} = \frac{\gamma}{1+\gamma}\left(\ln C_{it_\tau+1}(\beta) - \ln C_{it_\tau}(\beta)\right) + \eta_{2it_\tau}$$

where

$$\eta_{2it_\tau} = \frac{1}{1+\gamma} \upsilon_{it_\tau+1} + \varepsilon_{it_\tau+1} - \frac{\gamma}{1+\gamma} \varepsilon_{it_\tau} \equiv \eta_{2it_\tau}(\beta, \gamma).$$

From this, I construct the moment condition

$$E\left[\eta_{2it_\tau}(\beta, \gamma) \cdot \ln \widetilde{C}_{it_{\tau-1}}\right] = 0. \tag{19}$$

Since $\ln \widetilde{C}_{it_\tau}$ is correlated with η_{2it_τ} through ε_{it_τ}, I use $\ln \widetilde{C}_{it_{\tau-1}}$ as an instrument for $\ln \widetilde{C}_{it_\tau}$. Realized cost during the previous rate case is uncorrelated with the shock in the current rate case. Moreover, $\widetilde{C}_{it_{\tau-1}}$ will be correlated with $\widetilde{C}_{it_\tau}$ as long as $\rho \neq 0$.

Finally, consider the following quasi-difference across t_τ and $t_{\tau-1}+1$:

$$\frac{\gamma}{1+\gamma} \ln \widetilde{C}_{it_\tau} - \rho \ln \widetilde{C}_{it_{\tau-1}+1} = \frac{\gamma}{1+\gamma}\left(\ln C_{it_\tau}(\beta) - \rho \ln C_{it_{\tau-1}+1}(\beta)\right) + \eta_{3it_\tau}$$

where

$$\eta_{3it_\tau} = \frac{\gamma}{1+\gamma}\left(\xi_{it_\tau} + \varepsilon_{it_\tau}\right) - \rho\left(\frac{1}{1+\gamma} \upsilon_{it_{\tau-1}+1} + \varepsilon_{it_{\tau-1}+1}\right) \equiv \eta_{3it_\tau}(\beta, \gamma, \rho).$$

I construct moment conditions

$$E\left[\eta_{3it_\tau}(\beta, \gamma, \rho) \cdot \begin{pmatrix} 1 \\ \ln \widetilde{C}_{it_\tau+1} \end{pmatrix}\right] = 0. \tag{20}$$

Notice that I have used $\ln \widetilde{C}_{it_\tau+1}$ as an instrument for $\ln \widetilde{C}_{it_{\tau-1}+1}$. Realized cost in $t = t_\tau + 1$ is uncorrelated with past shocks but is correlated with $\widetilde{C}_{it_{\tau-1}+1}$ through the evolution of θ.

The parameters β, γ and ρ are identified as the solution to the moment conditions (18), (19) and (20). Uniqueness of the solution can be seen by taking each equation one at a time. For example, given ρ, equation (18) is linear in β; given β, equation (18) is linear in $\gamma/(1+\gamma)$ which uniquely pins down γ; and given β and γ, equation (20) is linear in ρ.

4.1.2 Identification of type distribution

Given the parameters and using assumption 2, I can rewrite realized cost during two consecutive rate cases as

$$\begin{aligned}\frac{\ln \widetilde{C}_{it_\tau} - \ln C_{it_\tau}(\beta)}{\rho} &= \theta_{it_{\tau-1}} + \frac{\xi_{it_\tau} + \varepsilon_{it_\tau}}{\rho} \\ \ln \widetilde{C}_{it_{\tau-1}} - \ln C_{it_{\tau-1}}(\beta) &= \theta_{it_{\tau-1}} + \varepsilon_{it_{\tau-1}}.\end{aligned}$$

The problem of finding the distribution of θ can be recast in the framework of measurement error with repeated measurements. Let $\left(\xi_{it_\tau} + \varepsilon_{it_\tau}\right)/\rho$ and $\varepsilon_{it_{\tau-1}}$ be the "measurement errors" while $\theta_{it_{\tau-1}}$ is the latent variable. The two measurement errors and the latent variable are all mutually independent and this follows from the assumptions on ξ_{it_τ} and the unanticipated cost shocks. Let ϕ_θ, ϕ_{U_1} and ϕ_{U_2} be the characteristic functions of $\theta_{it_{\tau-1}}$, $\left(\xi_{it_\tau} + \varepsilon_{it_\tau}\right)/\rho$ and $\varepsilon_{it_{\tau-1}}$ respectively. Assuming ϕ_θ, ϕ_{U_1} and ϕ_{U_2} have no real zeros[17], Kotlarski's (1967, Lemma 1) identification result imply[18]

$$\begin{aligned}\phi_\theta(t) &= \exp\left(\int_0^t \frac{\partial \phi_Y(0,t_2)/\partial t_1}{\phi_Y(0,t_2)} dt_2\right) \\ \phi_{U_1}(t) &= \frac{\phi_Y(t,0)}{\phi_\theta(t)} \\ \phi_{U_2}(t) &= \frac{\phi_Y(0,t)}{\phi_\theta(t)}\end{aligned}$$

where $\phi_Y(\cdot,\cdot)$ is the characteristic function of $\left(\frac{\ln \widetilde{C}_{it_\tau} - \ln C_{it_\tau}(\beta)}{\rho}, \ln \widetilde{C}_{it_{\tau-1}} - \ln C_{it_{\tau-1}}(\beta)\right)$. Since characteristic functions uniquely determine the distribution of random variables, we can therefore identify the distribution of $\theta_{it_{\tau-1}}$ from the distribution and characteristic function of

$$\left(\frac{\ln \widetilde{C}_{it_\tau} - \ln C_{it_\tau}(\beta)}{\rho}, \ln \widetilde{C}_{it_{\tau-1}} - \ln C_{it_{\tau-1}}(\beta)\right).$$

[17] Arellano and Bonhomme (2012) provide intuition for this technical requirement. When the characteristic function of the measurement errors are zero at certain points or intervals, the characteristic function of the observed measurements is not informative about the latent variable. Evdokimov and White (2012) replace this assumption with weaker conditions.

[18] See Rao (1992) and Li and Vuong (1998).

4.2 Estimation

To estimate the parameters, I use the sample analog of the moment conditions given by equations (18), (19) and (20). Ideally I would have a single estimating sample to construct the three moment conditions. However these moment conditions taken together require each firm in the sample to have at least two rate cases that are initiated *and* completed in the period 1988-1998. This leaves me with just 26 firms. The vector β contains 11 elements and therefore I need to estimate 13 parameters in total.

To increase the effective number of firms, I treat the same firm in two different rate cases as if they were different firms, e.g. I can define two different "firms" as (firm i, rate case τ) and (firm i, rate case $\tau+1$). Although there is dependence across these two firms, this dependence is fully captured by the dependence between types θ_{it_τ} and $\theta_{it_{\tau+1}}$ across the two rate cases. Thus differencing out θ's essentially gives independent samples (conditional on observables z).

To further alleviate the problem of a small sample size, I construct different samples for each of the moment conditions.[19] Moment condition (18), which identifies the cost function parameters conditional on ρ, only depends on rate case years. I include observations with rate cases initiated on or before 1999 even if they are concluded after 1999 in estimating this moment condition.

Sample selection bias may arise because the timing of the rate case is partly controlled by the firm. Suppose a rate case is initiated by the firm at time $t_{\tau+1}$ when at time $t \in (t_\tau+1, t_{\tau+1})$ realized costs are above some threshold. Whether the firm initiates a rate case at $t_{\tau+1}$ or not depends on the time $t \in (t_\tau+1, t_{\tau+1})$ values of observables, cost efficiencies, unanticipated cost shocks ε, and the unobserved threshold that is unrelated to $\varepsilon_{it_{\tau+1}}$ (otherwise this threshold provides information about $\varepsilon_{it_{\tau+1}}$ hence $\varepsilon_{it_{\tau+1}}$ will be anticipated by the firm). Selection bias arises because cost efficiencies ω_{it} are unobserved by the econometrician and these are correlated across time through the firm's intrinisic type θ_{it}. My estimating equations difference out ω_{it}'s which eliminates sample selection bias of this nature.

To estimate the distribution of $\theta_{it_{\tau-1}}$, I modify the algorithm described in Beran and Hall (1992) which adapts the discrete approximation of Hausdorff (1923).[20] The idea is to approximate

[19] I use a bootstrap procedure that samples over the "firms" to compute standard errors since the moment conditions are based on different samples.

[20] An alternative procedure is to estimate the characteristic function of $\left(\frac{\ln \widetilde{C}_{it_\tau} - \ln C_{it_\tau}(\beta)}{\rho}, \ln \widetilde{C}_{it_{\tau-1}} - \ln C_{it_{\tau-1}}(\beta)\right)$, derive the characteristic function of θ_{it_τ}, and then use an inverse Fourier transform to get the density of θ_{it_τ} (see for example Li and Vuong (1998) and Krasnokutskaya (2011)). Li and Vuong (1998) note that the procedure of Beran and Hall (1992) is a special case of their estimation procedure since all moments of the distribution are used to estimate the distribution. Beran and Hall (1992) instead only use a finite number of moments and apply the discrete approximation of Hausdorff (1923). To the extent that the distribution of θ_{it_τ} can be captured by a finite number of moments, the Beran and Hall (1992) procedure requires less data since this introduces less bias from (implicitly)

Table 6: Parameter estimates

log O&M variable cost	Model		FE	
	Est	SE	Est	SE
log emission rate	-0.706***	0.166	-0.226***	0.035
log emission rate*FGD	0.260***	0.173	0.160***	0.047
log Electricity output	1.178***	0.595	0.436***	0.043
log Price of labor	0.654***	0.332	0.051**	0.027
log Price of coal	0.239***	0.246	0.661***	0.064
log Price of oil	0.052***	0.065	0.207***	0.051
log Price of gas	0.055***	0.050	0.080*	0.044
log Nameplate	-0.643***	0.272	-0.151	0.101
FGD	3.715	3.142	-0.200***	0.048
Year	-0.002***	0.006	Yes	.
Disutility (γ)	88.178***	9.349	.	.
Type evolution (ρ_1)	0.841***	0.180	.	.

Notes: The first two columns contain estimation results from the procedure described in section 4.2. Standard errors are computed using bootstrap, where sampling is over firm-rate case. The last two columns contain OLS estimates with firm and year dummies included. Standard errors are clustered at the firm-level. Significance level: * 10%, ** 5%, *** 1%.

the distribution of $\theta_{it_{\tau-1}}$ by a discrete distribution that is constructed from estimated moments of $\theta_{it_{\tau-1}}$. To use the discrete distribution in the counterfactual welfare simulations, I fit a 5th order polynomial to its cumulative distribution function. I then invert this polynomial, draw a random sample of size 100, and use this as my sample. The algorithm is described in the online appendix.

4.3 Results

Table 6 presents the parameter estimates. The first two columns present results from the procedure described in the previous subsection. The coefficient on the emission rate imply that for a 10% decrease in emission rates, O&M variable cost increases by 7%, with 95% confidence interval of (-0.989,-0.476). If the utility has at least one flue-gas desulfurization unit, the effect of decreasing emission rates goes down by about 40%. To interpret the coefficient on log electricity output,

estimated higher-order moments.

I compute a simple measure of single-output returns to scale using Nelson's (1985) measure for variable cost functions.[21] My estimates imply a returns to scale measure of 1.39, hence exhibiting some returns to scale. For example, recent estimates of returns to scale range from 0.99 to 1.56 (Kleit and Terrell, 2001). Cost elasticities for the firm's variable inputs are 0.65, 0.24, 0.05 and 0.06 for labor, coal, oil and natural gas inputs. The cost elasticity for labor is surprisingly high while the one for coal price is low when we interpret these elasticities as cost shares since, in our sample, fuel accounts for 75% of O&M expense and 95% of fuel consumption comes from coal. The 95% confidence intervals for labor and coal price are (0.034, 0.732) and (0.220, 0.644) respectively, so confidence intervals for these coefficient estimates are quite wide.

The estimated evolution of intrinsic types show strong persistence. The coefficient on the past rate case's intrinsic type is 0.84 with a 95% confidence interval of (0.508, 0.997). An interesting question is whether a firm fixed effect would be sufficient to capture the unobserved heterogeneity in cost efficiencies given the high persistence of intrinsic types across rate cases. The last two columns of table 6 show the estimates from a regression model with firm fixed effects and year dummies. Focusing on the estimates of the coefficients on electricity output and emissions, we see that the estimates from the fixed effect model are severely attenuated. Although the firm fixed effect can capture the variation in cost efficiencies due to variation in intrinsic types across firms, the fixed effect fails to capture the effect of endogenous effort on cost efficiency. The upward bias in the coefficient on emission rates can be explained as follows. Think of effort as an omitted variable and imagine that emission rate is the only regressor. This omitted variable is negatively correlated with cost and negatively related to emission rates (because lower emission rates increase cost, which increases the marginal benefit from exerting effort). Thus there will be upward bias. An upward bias in the coefficient on emission rates leads to underestimated marginal abatement costs (MAC) since

$$MAC = -\frac{\partial}{\partial s}\left[\exp\left(\theta - e\right) C\left(\beta\right)\right] = \left|\beta_s\right| \exp\left(\theta - e\right) C\left(\beta\right) s^{-1}.$$

Figure 1 plots the discrete approximation to the cumulative distribution of intrinsic type θ, and the fitted polynomial. The distribution of θ implies a distribution of MACs and I plot the histogram of MACs in figure 2. In computing these MACs, I assume the following: (i) all firms have emission rate of 2.5 lbs per MMBtu, (ii) observable variables (electricity output, input prices and fuel burned) are at their mean values, (iii) firms do not have FGDs installed (i.e. $d_{FGD} = 0$), and (iii) firms exert optimal positive effort. The emission rate of 2.5 is the implicit emission standard under Phase I of the Acid Rain Program, so figure 2 reflects the distribution of MACs if SO_2 regulation were implemented by a uniform emission standard.

[21] Specifically, $RTS = \frac{1-\beta_N}{\beta_q}$ (Nelson, 1985).

There is considerable heterogeneity in the estimated MACs. The mean MAC is \$301 per ton while the median MAC is \$177. The 75th percentile is \$442 so most of the mass of the distribution is in the sub-\$500. The 90th and 95th percentiles are \$821 and \$1108 respectively, hence there is a nonneglible mass of firms that have MACs above \$800. Butraw (1999) presents a range of MAC estimates from various studies in the literature (both engineering-based and econometrically estimated). These MACs range from \$291 to \$760. Note that these estimates are affected by variation in output and input prices, while the distribution of MACs I estimate are fully generated by the estimated distribution of θ.

In figure 3, I regenerate a distribution of MACs under the assumption that firms do not exert effort. This gives the distribution of MACs during the rate case when prices are being set. If the public utility regulator reports costs to legislatures and the EPA based on information learned during rate cases (or weights rate case information heavily), then this distribution is the relevant one in the design of the emissions policy. Under this no effort scenario, mean MAC is higher by about 9.7% compared to the previous one where firms exerted optimal effort. The disutility costs associated with optimal effort is about \$242,000. However the regulator has to pay information rents on top of covering these disutility costs since firms have private information on its type and effort. One question I investigate in the next section is whether it is welfare-improving to induce firms to exert optimal effort, i.e. the least cost solution. It turns out that doing so entails large information rents to the point that it is generally better to have firms exerting zero effort.

5 Counterfactual welfare

The social planner's responsibility encompasses both pollution and economic regulation. Pollution regulation is concerned with emission rates while economic regulation deals with how the firm will be paid for providing its services. I focus on emission rates as the regulatory variable, taking the quantity of electricity, capital and input prices as exogenously given. A regulatory regime is a direct revelation contract that specifies a bundle $\left(s, \tilde{C}, t\right)$ for each intrinsic type θ. The bundle consists of an emission rate s, a realized and observable cost $\tilde{C}$, and a lump-sum transfer t (= revenue requirement RR). Finally, the mechanism can be implemented as follows: the firm reports its emissions rate and realized cost to the regulator, and then the regulator provides a transfer given this report.

The planner cares about social welfare given by equation (3), which I reproduce here (replacing RR with t):

$$W = \int \{V(q(\theta)) - D(s(\theta)) - (1+\lambda)t(\theta) + \Pi(\theta)\}\, d\mathcal{F}(\theta). \tag{3}$$

Moreover, the planner faces constraints in designing the regime. First, the planner needs to satisfy

individual rationality constraints which require leaving firms with nonnegative economic profits:

$$\Pi(\theta) = t(\theta) - \left[\tilde{C}(\theta) + \psi\left[e(\theta)\right]\right] \geq 0 \tag{21}$$

for all θ, where we can express effort as a function of $\left(s, \tilde{C}\right)$:

$$e(\theta) = \theta - \ln \frac{\tilde{C}(\theta)}{C(s(\theta))}. \tag{22}$$

As in Laffont (1994) and Laffont and Tirole (1986), I assume the social planner observes realized cost but not the firm's type and effort. This informational constraint leads to incentive compatibility constraints

$$\Pi(\theta) \geq t\left(\theta'\right) - \left\{\tilde{C}\left(\theta'\right) + \psi\left[e\left(\theta'\right)\right]\right\} \tag{23}$$

for all θ and $\theta' \neq \theta$. These constraints ensure that a type θ does not have an incentive to pick some other type's bundle.

Let Θ be the random sample of size N of θ's that I have drawn from the estimated distribution of θ. I impose that firms do not have a flue-gas desulfurization unit, set $t = 1995$, and use the mean values of electricity output, input prices, amount of fuel burned and capital to compute $C(s) = \Psi s^{\beta_s}$, where $\beta_s < 0$ and $\Psi = \exp(\beta_{YEAR}1995)N^{\beta_N} q^{\beta_q} p_l^{\beta_l} p_c^{\beta_c} p_o^{\beta_o} p_g^{\beta_g}$.

Given a regulatory regime, I compute welfare as

$$\widetilde{W}(p, \lambda) = \frac{1}{N} \sum_{\theta \in \Theta} \left\{-p \cdot \mathcal{S}(s(\theta)) - (1+\lambda)\, t(\theta) + \Pi(\theta)\right\}.$$

The linear function $\mathcal{S}(s(\theta))$ converts an emission rate $s(\theta)$ to tons of SO_2 emissions using the mean amount of fuel burned. I impose a linear pollution damage function so $p > 0$ represents the constant marginal damage from a ton of pollution.[22] The variable $\lambda > 0$ is the social cost of public funds. I treat (p, λ) as simulation parameters and I compute $\widetilde{W}$ for different combinations of (p, λ). Specifically, I consider $p \in \{200, 250, 300..., 1000\}$ and $\lambda \in \{0.1, 0.2, 0.3, ..., 1\}$.[23] Finally, the welfare metric $\widetilde{W}$ does not include the surplus from electricity consumption and thus I focus on welfare differences.

[22] Allowing for a more complicated nonlinear damage function necessitates sophisticated techniques to estimate marginal damages across sources. Fowlie and Muller (2012) perform welfare analysis for non-uniformly mixed pollutants by utilizing the method for computing marginal damages developed in Muller and Mendelsohn (2009). They do not touch on issues arising in regulation with asymmetric information *and* costly information rents which is my main focus.

[23] The range of emission permit prices during Phase I was about \$60 to about \$400 per ton. Moreover, the range of emission tax rates under the proposed Sulfur and Nitrogen Emissions Tax Act of 1987 (H.R. 2497) is \$300 to \$900 per ton. Thus these choices of constant marginal damages are reasonable approximations of what policy-makers had in mind with respect to marginal damage from SO_2 emissions. As for the cost of public funds, the public finance literature (Laffont, 2005; Ballard, Shoven and Whalley, 1985) estimates this to be around $\lambda = 0.3$ for the US. Higher values of λ reflect greater difficulty to raise taxes (e.g. developing countries).

5.1 Welfare analysis

I compute welfare under the following regulatory regimes.[24]

Full-information (I) The planner observes θ and e so incentive compatibility constraints are not relevant. Define the *first best allocation* as the pair $\left(s^{FB}(\theta), e^{FB}(\theta)\right)$ that solves

$$\begin{aligned} p \cdot \frac{d\mathcal{S}(s(\theta))}{ds} &= (-\beta_s)(1+\lambda)\exp[\theta - e(\theta)] \cdot \Psi \cdot s(\theta)^{\beta_s - 1} \\ \psi'[e(\theta)] &= \exp[\theta - e(\theta)] \cdot \Psi \cdot s(\theta)^{\beta_s}. \end{aligned}$$

The planner pays firms a transfer that is just enough to cover costs. Thus the full-information regime is characterized by

$$\theta \mapsto \left(s^{FB}(\theta), e^{FB}(\theta), \exp\left[\theta - e^{FB}(\theta)\right] \cdot C\left(s^{FB}(\theta)\right) + \psi\left[e^{FB}(\theta)\right]\right).$$

Optimal regulation (II) The planner chooses $\left(s, \tilde{C}, t\right)$ to maximize welfare subject to individual rationality and incentive compatibility constraints. The optimal mechanism is fully characterized in the online appendix and is similar to the mechanism characterized in Proposition 2 of Laffont (1994). Optimal allocations $(s(\theta), e(\theta))$ deviate from $\left(s^{FB}(\theta), e^{FB}(\theta)\right)$ except for the most efficient type because of the planner's desire to reduce information rents. The most inefficient type earns zero profits while the rest earn strictly positive profits.

Uniform emissions standard ($\overline{s} = 2.5$) (III) The planner requires $s(\theta)$ to be equal to the uniform emissions standard $\overline{s} = 2.5$ for all θ. This value of $\overline{s}$ is the emission rate used to allocate free allowances during the first phase of the Acid Rain Program.[25] In computing welfare, I take operating costs as if the firm was in a rate case. The firm's reported costs get fully reimbursed and this leads to the firm exerting zero effort.

Least cost (IV) The planner induces the allocation $\left(s^{FB}(\theta), e^{FB}(\theta)\right)$ as in the full-information regime. This allocation minimizes the sum of damages and abatement costs, hence I refer to it as the least cost regime. Unlike in the full-information regime, the planner has to incentivize firms to reveal their private information, and so has to pay information rents. Specifically, the planner chooses the set of transfers such that individual rationality and incentive compatibility constraints are satisfied given the first best allocation. Note that these transfers depend on type.

[24] I characterize these regimes in terms of $e(\theta)$ instead of $\tilde{C}(\theta)$ using equation (22).

[25] In theory, Phase I allocations were based on historical heat input multiplied by 2.5 lbs/MMBtu. However, actual allocations differ from this formula due to special provisions such as bonus allowances for installing scrubbers.

Status quo (rate regulation with emission taxes) (V) This regime approximates the actual regime in place.[26] I compute welfare under the assumption that rate case information on costs are used in designing the policy. Rate regulation induces the firm to exert zero effort in producing electricity and abatement during the rate case. Next, emissions taxes are set equal to adjusted marginal damage, i.e. $p/(1+\lambda)$ per ton, and the firm picks $s(\theta)$ to minimize the tax payment plus abatement costs (given $e(\theta) = 0$). The planner then fully reimburses whatever cost the firm reports. I assume that tax payments are also fully reimbursed, *ex-post*. That is, the transfer $t(\theta)$ includes the *fixed* amount $p/(1+\lambda) \cdot s(\theta)$ where $s(\theta)$ equates the tax with marginal abatement cost.

Table 7: Welfare results

Regime:	I	II	III	IV	V
Aggregate emissions (in M tons)	5.128	5.207	5.456	5.128	5.407
Cost-savings relative to unif standard and weighted by λ (in \$M)	947	881	0	947	713
Information rents multiplied by λ (in \$M)	0	34	0.03	759	0.03
Information rents multiplied by λ (as % of value of total emissions)	0	1.7	0.002	37	0.002

Notes: I = Full-information, II = Optimal regulation, III = Uniform emission standard, IV = Least cost, V = Status quo

Table 7 contains aggregate emissions, cost-saving and information rent estimates under the various regimes. These numbers are averaged across the simulation parameters. To get aggregate numbers, I compute the implied number of firms such that aggregate emissions in the uniform emission standard regime is equal to the number of freely distributed emission permits in Phase I of the Acid Rain Program. The basic rule for freely allocated permits multiplies the emission rate of 2.5 lbs/MMBtu with historical fuel consumption, and this amounts to about 5.456M tons of SO_2 (Joskow and Schmalensee, 1998). Aggregate emissions in the status quo regime (V) is about 1% lower than the calibrated uniform emissions standard regime, which implies average emissions rate in the former is quite close to 2.5 lbs/MMBtu as one would expect. Aggregate emissions under the full-information (I) and optimal regulation (II) regimes are 5.128M and 5.207M tons respectively, or equivalently about 6% and 5% lower than under the uniform emission standard and the status quo.

The optimal mechanism distorts the first best allocation $\left(s^{FB}(\theta), e^{FB}(\theta)\right)$ in such a way that reduces information rents. Reducing abatement and effort levels required for inefficient types incurs

[26] While the actual regime involves tradable emissions permits market, I approximate this using a regime with emissions taxes.

second order losses in efficiency in exchange for first order gains in lower information rents to efficient types. Figure 4 plots the difference in mean SO_2 emission rates for the full-information and optimal regulation regimes for various combinations of simulation parameters. The gap in emission rates reflects the degree of distortion that the optimal mechanism introduces, and this generally decreases as the marginal damage increases, and increases as the cost of public funds increases. Figure 5 plots the emissions profile (ordered from most efficient to least efficient firm type) under the two regimes, while figure 6 plots the corresponding densities (for both figures, I let $p = 300$ and $\lambda = 0.3$). As expected, abatement is distorted downwards (i.e. more emissions) for less efficient types, relative to first best.

Except for the uniform emission standard regime, all of our regimes allows for different emission rates across different efficiency types. This added flexibility reduces abatement cost relative to the uniform emission standard, conditional on the same level of abatement. The greatest cost savings is achieved by the full-information (I) and the least cost regimes (IV). This amounts to aggregate annual cost-savings of \$947M relative to the uniform emission standard regime. Carlson et al (2000) estimates this cost-savings to be around \$388M.[27]As these cost-savings reflect gains from trade, the larger cost-savings I get may be reflecting more pronounced heterogeneities in my estimated efficiency distribution. Cost-savings goes down by \$66M with the optimal regulation regime, which is expected since the optimal mechanism trades off efficiency for lower information rents. Finally, efficiency distortions from rate regulation leads to reduced cost-savings of about \$234M relative to the least cost regime.

The least cost regime implements the full-information abatement and effort levels, which minimizes the sum of damage and abatement costs. However, because firms have private information about the efficiency types and effort, the planner has to pay information rents to induce firms to reveal their information. When information rents are large, it may not be desirable or even feasible to implement the full-information allocations (Spulber, 1998). Therefore it is important to weigh the benefits of added cost-savings against the increase in information rents when evaluating policies. Table 7 presents aggregate information rents, both in terms of monetary value, and as a percentage of the (negative) value of total emissions. Information rents under the least cost regime amounts to \$759M, which is much larger than information rents under optimal regulation (\$34M). Although there are large gains from trade in the least cost regime, implementing the least cost regime is too costly since large information rents have to be paid. Looking across the simulation parameters, the value of information rents (i.e. multiplied by λ) to implement the least cost regime range from \$114M ($\lambda = 0.1$, $p = 200$) to \$1.7B ($\lambda = 1$, $p = 1000$). Figure 7 plots the magnitude

[27]Carlson et al (2000) estimates cost-savings in Phase I of ARP to be around \$250M. I multiply this number by $1 + \bar{\lambda}$ where $\bar{\lambda} = 0.55$, i.e. cost of public funds averaged across my simulation parameters.

of information rents (i.e. not multiplied by λ) against different values of simulation parameters. These information rents increase with higher marginal damages, and decrease as the cost of public funds go up.

Given that there are large information rents in the least cost regime, are we better off with the status quo? Figure 8 plots the difference in welfare between the status quo and the least cost regime. Except for cases where the cost of public funds is low (e.g. $\lambda = 0.1$), welfare is higher under the status quo compared to the least cost regime. This means that the loss in efficiency due to zero effort in the status quo is well-compensated by the gain from reducing information rents.

I close this section by looking at welfare gains from the optimal mechanism relative to the status quo. Annual aggregate welfare gains range from \$148M ($\lambda = 0.2$, $p = 200$) to \$347M ($\lambda = 1$, $p = 1000$). Figure 9 plots these welfare gains against the simulation parameters. These gains increase as marginal damage increases. However, the relationship between gains and the cost of public funds is nonmonotonic, although the effect of λ is quite small.

6 Conclusion

I empirically examine the interaction of emissions and price regulation in the electricity industry to study optimal pollution regulation when there are informational and distributional constraints. I provide evidence that the form of price regulation that electricity-generating firms were facing distorts their incentives to burn fuel efficiently and increases heat rates by about 4% to 6%, with operating costs increasing by roughly the same percentage. I then exploit price regulation to identify the dimension of abatement cost affected by incentives, and finally evaluate welfare under different counterfactual pollution regulatory regimes.

The welfare analysis reveals the important role that information rents play in evaluating various regulatory regimes. While there are large gains from trade from implementing the least cost allocation, i.e. abatement and effort levels that minimize damages from emissions plus abatement costs, much larger information rents have to be transferred to firms in order to implement this allocation. Under the assumption that public funds are costly to raise, the total welfare cost of such a transfer amounts to \$114M to \$1.7B. Although a regime with rate regulation combined with an emissions tax provides poor incentives for firms to endogenously reduce their costs, this regime generally leads to higher welfare compared to the least cost allocation. The reduction in information rents more than compensates for efficiency losses due to poor effort incentive provision. Finally, total annual welfare gains from the optimal mechanism relative to the regime that combines rate regulation with emissions taxes to be around \$148M to \$347M.

References

Abito, J. M. (2014), "An Auditing Model of Natural Monopoly Regulation," mimeo, Univ of Pennsylvania.

Alt, L. E. (2006), *Energy Utility Rate Setting: A Practical Guide to the Retail Rate-Setting Process for Regulated Electric and Natural Gas Utilities*, Morrisville, NC: Lulu.com.

Arellano, M. and S. Bonhomme (2012), "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," Review of Economic Studies, 79:3, 987-1020.

Averch, H. and L. Johnson (1962), "Behavior of the Firm under Regulatory Constraint," *American Economic Review*, 52:5, 1053-1069.

Bailey, E. E. and R. D. Coleman (1971), "The Effect of Lagged Regulation in an Averch-Johnson Model," Bell Journal of Economics, 2:1, 278-292.

Ballard, C. L., J. B. Shoven and J. Whalley (1985), "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States," *American Economic Review*, 75:1, 128-138.

Baron, D. P. and R. B. Myerson (1982), "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50:4, 911-930.

Baumol, W. J. and A. K. Klevorick (1970), "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion," *Bell Journal of Economics and Management Science*, 1:2, 162-190

Beran, R. and P. Hall (1992), "Estimating Coefficient Distributions in Random Coefficient Regressions," *Annals of Statistics*, 20:4, 1970-1984.

Besanko, D. (1985), "On the Use of Revenue Requirements Regulation Under Imperfect Information," in M. A. Crew (ed.), *Analyzing the Impact of Regulatory Change in Public Utilities*, Lexington, MA: Lexington Books, 39-55.

Brocas, I., K. Chan and I. Perrigne (2006), "Regulation under Asymmetric Information in Water Utilities," *American Economic Review, Papers and Proceedings*, 96, 62-66.

Bushnell, J. and C. Wolfram (2007), "The Guy at the Controls: Labor Quality and Power Plant Efficiency," NBER working paper 13215.

Butraw, D. (1999), "Cost Savings, Market Performance and Economic Benefits of the US Acid Rain Program," in S. Sorrell and J. Skea (eds.), *Pollution for Sale: Emissions Trading and Joint Implementation*, Edward Elgar Publishing, 43-62.

Carlson, C. P., D. Butraw, M. Cropper and K. Palmer (2000), "SO_2 Control by Electric Utilities: What are the Gains from Trade," *Journal of Political Economy*, 108:6, 1292-1326.

Chan, G., R. Stavins, R. Stowe and R. Sweeney (2012), "The SO_2 Allowance-trading System and the Clean Air Act Ammendments of 1990: Reflections on 20 Years of Policy Innovation," *National Tax Journal*, 65:2, 419-452.

Chan, H. S., H. Fell, I. Lange and S. Li (2013), "Efficiency and Environmental Impacts of Electricity Restructuring on Coal-fired Power Plants," mimeo, Univ of Maryland.

Chu, L. Y. and D. E. M. Sappington (2007), "Simple Cost-Sharing Contracts," *American Economic Review,* 97:1, 419-428.

Cicala, S. (2014), "When Does Regulation Distort Costs? Lessons from Fuel Procurement in U.S. Electricity Generation, " *American Economic Review.*, forthcoming.

Cramton, P. and S. Kerr (2002), "Tradeable Carbon Permit Auctions: How and why to auction not grandfather," *Energy Policy*, 30, 333-345.

Davis, L. W. and C. D. Wolfram (2012), "Deregulation, Consolidation, and Efficiency: Evidence from U. S. Nuclear Power, " *American Economic Journal: Applied Economics*, 4:4, 194-225.

Ellerman, A. D., P. L. Joskow, R. Schmalensee, J.-P. Montero and E. Bailey (2000), *Markets for Clean Air: The U.S. Acid Rain Program*, Cambridge, UK: Cambridge University Press.

Ellerman, A. D. and J.-P. Montero (2007), "The efficiency and robustness of allowance banking in the US Acid Rain Program," *The Energy Journal,* 28, 47-61.

Environmental Protection Agency (2001), *Acid Rain Program: 2001 Progress Report*, EPA-430-R-02-009.

Environmental Protection Agency (2014, July 7) *Overview of Greenhouse Gases*, Retrived from http://www.epa.gov/climatechange/ghgemissions/gases/co2.html.

Evdokimov, K. and H. White (2012), "Some Extensions of a Lemma of Kotlarski," *Econometric Theory*, 28:4, 925-932.

Fabrizio, K., N. Rose and C. Wolfram (2007), "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency," *American Economic Review*, 97:4, 1250-1277.

Fowlie, M. (2010), "Emissions Trading, Electricity Industry Restructuring, and Investment in Pollution Control," *American Economic Review*, 100:3, 837-869.

Fowlie, M. and N. Muller (2013), "Market-based emissions regulation when damages vary across sources: What are the gains from differentiation?" mimeo, UC Berkeley.

Fowlie, M., M. Reguant and S. P. Ryan (2014), "Market-based emissions regulation and industry dynamics, " *Journal of Political Economy*, forthcoming.

Gagnepain, P. and M. Ivaldi (2002), "Incentive Regulatory Policies: The Case of Public Transit Systems in France," *RAND Journal of Economics*, 33:4, 605-629.

Gagnepain, P., M. Ivaldi and D. Martimort (2013), "The Cost of Contract Renegotiation: Evidence from the Local Public Sector," *American Economic Review*, 103:6, 2352-2383.

Goulder, L. H., I W. H. Parry and D. Butraw (1997), "Revenue-raising versus Other Approaches to Environmental Protection: The Critical Significance of Preexisting Tax Distortions," *RAND Journal of Economics*, 28:4, 708-731.

Hausdorff, F. (1923), "Momentprobleme für ein endliches Intervall", *Mathematische Zeitschrift*, 16, 220-246

Joskow, P. L. and R. Schmalensee (1986), "Incentive Regulation for Electric Utilities," *Yale Journal of Regulation*, 4, 1-49.

Joskow, P. L. and R. Schmalensee (1998), "The Political Economy of Market-Based Environmental Policy: The U.S. Acid Rain Program," *Journal of Law and Economics*, 41:1, 37-83.

Kleit, A. N. and D. Terrell (2001), "Measuring Potential Efficiency Gains from Deregulation of Electricity Generation: A Bayesian Approach," *Review of Economics and Statistics,* 83:3, 523-530.

Kotlarski, I. I. (1967), "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69-76.

Krasnokutskaya, E. (2011), "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *Review of Economic Studies*, 78:1, 293-327.

Laffont, J.-J. (1994), "Regulation of Pollution with Asymmetric Information," in C. Dosi and T. Tomasi (eds.), *Nonpoint Source Pollution Regulation: Issues and Analysis*, Kluwer Academic Publishers, 39-66.

Laffont, J.-J. (2005), *Regulation and Development*, Cambridge, UK: Cambridge University Press.

Laffont, J.-J. and J. Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614-641.

Laffont, J.-J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.

Lewis, T. R. (1996), "Protecting the Environment when Costs and Benefits are Privately Known," *RAND Journal of Economics*, 27:4: 819:847.

Li, T. and Q. Vuong (1998), "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139-165.

Lim, C. S. H. and A. Yurukoglu (2014), "Dynamic Natural Monopoly Regulation: Time Inconsistency, Asymmetric Information, and Political Environments," mimeo, Cornell and Stanford.

Muller, N. and R. Mendelsohn (2009), "Efficient Pollution Regulation: Getting the Prices Right," *American Economic Review*, 99:5, 1714-1739.

Nelson, R. A. (1985), "Returns to Scale from Variable and Total Cost Functions," *Economic Letters*, 18, 271-276.

Pint, E. (1992), "Price-Cap Versus Rate-of-Return Regulation in a Stochastic-Cost Model," *Rand Journal of Economics*, 23:4, 564-578.

Perrigne, I. and Q. H. Vuong (2011), "Nonparametric Identification of a Contract Model With Adverse Selection and Moral Hazard," *Econometrica*, 79:5, 1499-1539.

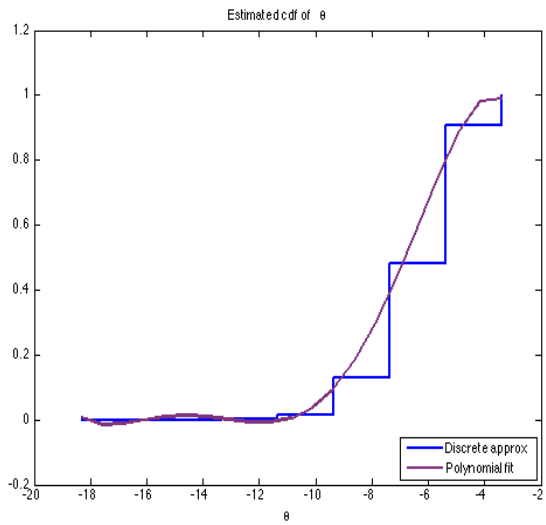
Rao, B. L. S. Prakasa (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions*, San Diego: Academic Press.

Schmalensee, R. and R. N. Stavins (2012), "The SO_2 Allowance Trading System: The Ironic History of a Grand Policy Experiment," Harvard Kennedy School Faculty Working Paper Series, RWP12-030.

Spulber, D. F. (1998), "Optimal Environmental Regulation under Asymmetric Information," *Journal of Public Economics*, 35:2, 163-181.

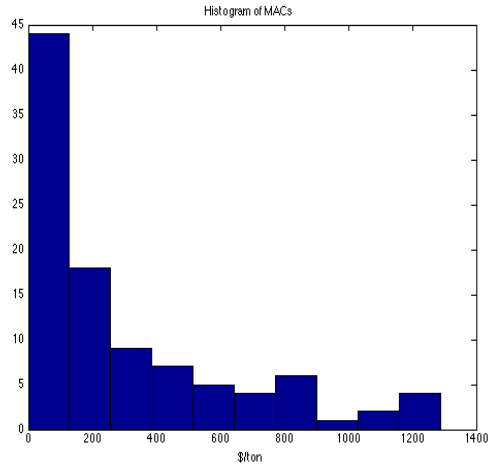
Wolak, F. A. (1994), "An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction," *Annales d'Economie et de Statistiques*, 34, 13-69.

Figure 1: Estimated cdf of θ



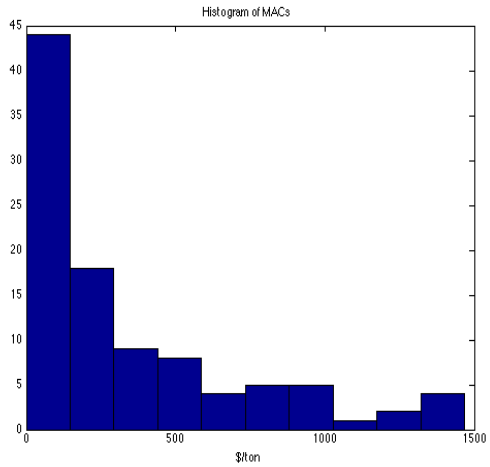
Notes: Stepwise cdf is the estimate using a modification of Hausdorff's (1923) discrete approximation and the Beran and Hall (1992) algorithm. The curve is a 5th order polynomial fit to this cdf.

Figure 2: Histogram of marginal abatement costs in \$ per ton of SO_2 emissions under optimal effort



Notes: The figure contains the histogram of marginal abatement costs (MAC) for the random sample I drew from the estimated type distribution. MACs are evaluated at an emission rate of 2.5 lbs/MMBtu and expressed in 1995\$ per ton. I assume firms exert optimal effort in generating this distribution.

Figure 3: Histogram of marginal abatement costs in \$ per ton of SO_2 emissions under no effort



Notes: The figure contains the histogram of marginal abatement costs (MAC) for the random sample I drew from the estimated type distribution. MACs are evaluated at an emission rate of 2.5 lbs/MMBtu and expressed in 1995\$ per ton. I assume firms exert zero effort in generating this distribution.

Figure 4: Difference in mean emission rates: Optimal mechanism vs Full-information

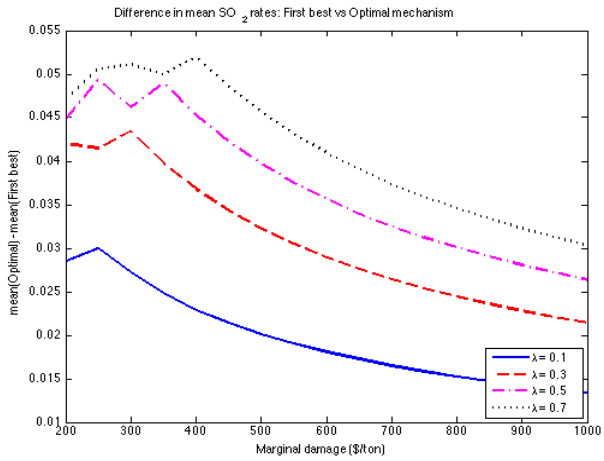


Figure 5: Emissions profile: Optimal mechanism vs Full-information

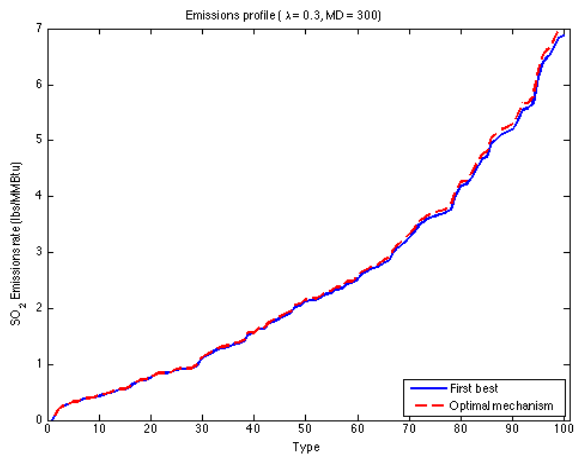


Figure 6: Density of emissions: Optimal mechanism vs Full-information

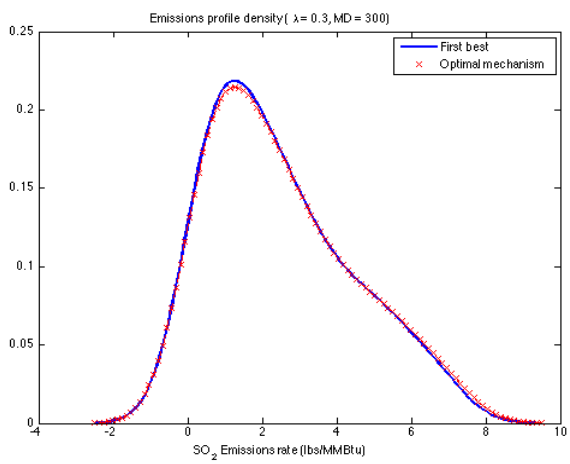


Figure 7: Information rents (unweighted by λ): Least cost

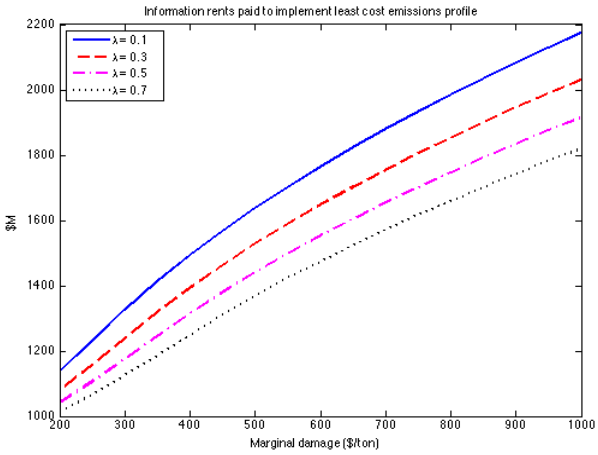


Figure 8: Welfare difference: Status quo vs Least cost

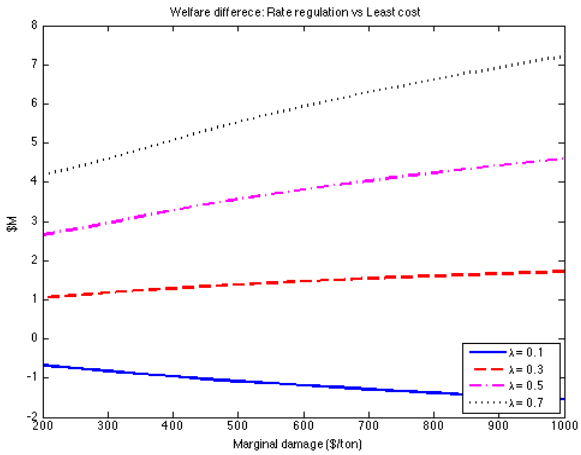
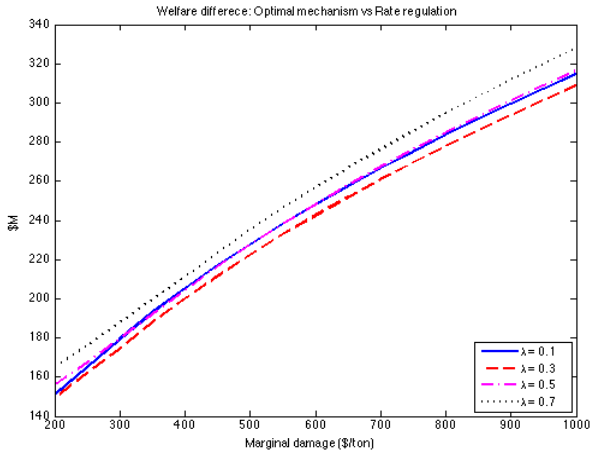


Figure 9: Welfare difference: Optimal mechanism vs Status quo



Appendix

Data construction

I construct a list of generating units affected by Phase I of the Acid Rain Program using compliance data from EPA's Air Markets Program. The compliance data includes all generating units that were part of Phase I. For each unit in this list, I get unit-level data on net electricity generation and nameplate capacity from the Energy Information Administration's (EIA) Form 767 for the period 1988-1999. I aggregate the data to the plant-level and get data on emissions, fuel consumption (coal, oil and natural gas), and on whether the plant has a flue gas desulfurization (FGD) unit installed. I then aggregate these measures at the utility-level so that I can match these to the regulatory and rate case data. Utility-level fuel prices are constructed from the Federal Energy Regulatory Commissions' (FERC) Form 423 by taking a weighted-average of delivered prices across a utility's plants for each fuel type. Finally I match these utilities to the regulatory database of SNL Financial and extract data on fuel expense and non-fuel operations and maintenance expense related to electricity generation, excluding expenses from nuclear plants. I also get average monthly salaries of full-time employees involved in electricity generation. This comprises the operations and cost data.

The rate case data comes from Regulatory Research Associates (RRA), a research and consulting company owned by SNL Financial. These contain SNL utility codes that I use to match the rate case data with operations and cost data. I get data on the year the rate case was proposed, the year it was authorized, the test year, proposed and authorized rate base, and the proposed and authorized rate of return (ROR). From these data I can construct the proposed and authorized return on the rate base (RRB).

I am able to identify 84 utilities that own at least one Phase I plant by matching the EPA data with the EIA data. Of these I can match 69 utility codes to SNL's regulatory data. My primary variables are net generation from coal, oil and natural gas plants; SO_2emission rate; a dummy for whether the utility has at least one plant with an FGD; total nameplate capacity; and average prices for coal, oil and gas. The number of utilities with nonmissing data goes down to 41.

Estimating the distribution of θ

I modify the algorithm described in Beran and Hall (1992) which adapts the discrete approximation of Hausdorff (1923). The algorithm approximates the distribution of $\theta_{it_{\tau-1}}$ by a discrete distribution constructed from estimated moments of $\theta_{it_{\tau-1}}$. The algorithm is implemented as follows:

1. I estimate the first $m = 15$ moments of $\theta_{it_{\tau-1}}$ using data on

$$\left(\frac{\ln \widetilde{C}_{it_\tau} - \ln C_{it_\tau}(\beta)}{\rho}, \ln \widetilde{C}_{it_{\tau-1}} - \ln C_{it_{\tau-1}}(\beta) \right).$$

   I only use data for the first two rate cases for each firm.

2. Define the k-th moment of $\theta_{it_{\tau-1}}$ as μ_k. I modify the algorithm proposed by Beran and Hall (1992) to allow for an arbitrary compact support, i.e. I assume the distribution of $\theta_{it_{\tau-1}}$, i.e. $\mathcal{F}_\theta$, is supported on the compact interval $[a, b]$.[28] Define the transformed moment

$$\widetilde{\mu}_k = \sum_{j=0}^{k} \begin{pmatrix} k \\ j \end{pmatrix} (b-a)^{-j} \left(\frac{a-b}{a} \right)^{-(k-j)} \mu_k$$

   for $k = 0, 1, 2, ..., m$ where $\mu_0 = 1$. This transformation maps the moments of $\mathcal{F}_\theta$ to the moments of a distribution supported on $[0, 1]$.

3. Construct the discrete distribution over $\{0, 1/m, 2/m, ..., 1\}$ with

$$\Pr\left(\frac{j}{m}\right) = \begin{pmatrix} m \\ j \end{pmatrix} \Delta^{m-j} \widetilde{\mu}_k$$

   for $j = 0, 1, 2, ..., m$, where Δ^r is the r-th order difference operator defined as

$$\Delta^r \widetilde{\mu}_k = \sum_{i=0}^{r} \begin{pmatrix} r \\ i \end{pmatrix} (-1)^i \widetilde{\mu}_{k+i}$$

   and $\widetilde{\mu}_k$'s are sample moments of $\theta_{it_{\tau-1}}$. Hausdorff (1923) shows that this discrete distribution converges to $\mathcal{F}_\theta$ (Shohat and Tamarkin, 1943, p. 93-94).

Characterization of optimal pollution regulation

The regulator maximizes welfare $\widetilde{W}$ subject to individual rationality (IR) and incentive compatibility (IC) constraints. Although the original type space is two-dimensional, there is no instrument to screen R-types. Thus all firms will pool at the highest possible R. I solve the problem as

[28] Beran and Hall (1992) assume a distribution supported on $[-c, c]$ where $c = 5\sqrt{\mu_2}$. This forces the algorithm to attempt to put some mass (although very small) on discrete points in $[0, c]$, which then would imply an average (across the sample) marginal abatement costs well over \$90,000 under the implicit emission standard of 2.5 lbs/MMBtu in Phase 1 of the Acid Rain Program. Instead, I restrict the distribution such that average MACs are between \$0.001 to \$3,000, which reflect reasonable bounds for marginal abatement costs. For example, the penalty for non-compliance with the Acid Rain Program amounts to \$2,000 for every ton of emission in excess of allowances held at the end of the compliance year. Allowance prices never went over \$1,000 so this penalty was always non-binding.

a one-dimensional screening problem since R does not affect welfare comparisons (except for the full-information regime). The distribution of types is discrete so I adapt standard methods for continuous types (e.g. Laffont and Tirole, 1993; Laffont, 1994) to my setting. The first step is to reduce the set of IC constraints into upward local ICs. I solve the problem in terms of firms' profits instead of transfers. For any type θ_i and θ_j, IC requires

$$\begin{aligned}\Pi_i &\geq \Pi_j + \left[\psi\left(\theta_j - \ln\frac{\tilde{C}_j}{\Psi s_j^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_j}{\Psi s_j^{\beta_s}}\right)\right] \\ \Pi_j &\geq \Pi_i - \left[\psi\left(\theta_j - \ln\frac{\tilde{C}_i}{\Psi s_i^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_i}{\Psi s_i^{\beta_s}}\right)\right].\end{aligned}$$

Combining these and using the assumption that ψ is increasing, we get

$$\psi\left(\theta_j - \ln\frac{\tilde{C}_i}{\Psi s_i^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_i}{\Psi s_i^{\beta_s}}\right) \geq \psi\left(\theta_j - \ln\frac{\tilde{C}_j}{\Psi s_j^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_j}{\Psi s_j^{\beta_s}}\right).$$

As long as $\left(s, \tilde{C}\right)$'s satisfy this inequality, we can focus on upward local ICs. I solve the reduced problem and check this inequality ex-post.

By standard arguments, the IR of the most inefficient type will be binding while the ICs of the rest of the types will be binding. Thus $\Pi_N = 0$ and for $i = 1, 2, ..., N-1$,

$$\Pi_i = \Pi_{i+1} + \left[\psi\left(\theta_{i+1} - \ln\frac{\tilde{C}_{i+1}}{\Psi s_{i+1}^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_{i+1}}{\Psi s_{i+1}^{\beta_s}}\right)\right].$$

Thus the social planner's (relaxed) problem is given by

$$\max_{(s_i,\tilde{C}_i)_{i=1}^n} \widetilde{W} = \frac{1}{N}\sum_{i=1}^{N}\left\{-Ds_i - (1+\lambda)\left[\tilde{C}_i + \psi\left(\theta_i - \ln\frac{\tilde{C}_i}{\Psi s_i^{\beta_s}}\right)\right] - \lambda\Pi_i\right\}$$

subject to

$$\begin{aligned}\Pi_N &= 0 \\ \Pi_i &= \Pi_{i+1} + \left[\psi\left(\theta_{i+1} - \ln\frac{\tilde{C}_{i+1}}{\Psi s_{i+1}^{\beta_s}}\right) - \psi\left(\theta_i - \ln\frac{\tilde{C}_{i+1}}{\Psi s_{i+1}^{\beta_s}}\right)\right],\end{aligned}$$

where $D > 0$ is the marginal damage from an increase in the emission rate.