Cost Efficiency and Environmental Policy in US Electricity Generation:

The Case of the Tradable SO_2 Allowance Scheme

Lindsay Tuthill

June 2008

EV 41

The contents of this paper are the author's sole responsibility. They do not necessarily represent the views of the Oxford Institute for Energy Studies or any of its Members.

ISBN
978-1-901795-74-8

Contents

Tables and Figures

Abstract

This paper addresses the effects of emissions policy on the ability of electricity generating firms to operate at minimum cost. More specifically, we analyze the impact of the tradable SO2 permit scheme established by the Clean Air Act Amendments of 1990 in the US on the cost efficiency of 37 electricity generating firms over the years 1990-2004. We estimate a stochastic cost frontier and model the cost inefficiency component as a function of several firm- and policy-specific variables in order to analyze their impact on firm cost efficiency estimates. We find aggregate mean cost efficiency was lower post-regulation and that increases in the capacity factor and coal's percentage of total capacity were associated with higher firm cost efficiency, while the opposite was true of increases in scrubbed capacity. We find also that inclusion in the SO2 scheme was associated with a reduction in cost efficiency while increased permit prices were linked to cost efficiency improvements. Thus the SO2 scheme appears to have led to a reduction in firms' ability to minimize costs, though cost inefficiencies were recovered at higher permit prices, presumably because of capital constraints and the improved cost effectiveness of scrubbers at higher emissions prices.

Introduction

This paper analyzes cost efficiency and its determinants in the US electricity generating industry, accounting for firm-specific characteristics and focusing on the effects of the tradable SO_2 allowance scheme created by the Clean Air Act Amendments (CAAA) of 1990. The goal is to enhance the understanding of cost-minimizing behaviour under such cap-and-trade programs, given conflicting regulatory objectives related to environmental quality, output prices and supply security (see, e.g. Lee (2002) or Helm, Hepburn and Mash (2003)). Much of the literature on producer performance in the electricity generating industry focuses on technical efficiency or total factor productivity (TFP) measurement at the plant level, and has neglected the question of the attainment of the behavioural objective of production at least cost, particularly as affected by environmental policy. We seek to address the issue of the firm-level cost efficiency effects of the tradable SO_2 allowance scheme in the US, and our work is, to our knowledge, the first to do so. Our objectives are to calculate firm- and time-specific cost inefficiency estimates and to determine the effects of the SO_2 scheme and other firm characteristics on these estimates in order to inform both the design of and expected response to future environmental policy.

We use a stochastic frontier analysis (SFA) model, in which the error term of an otherwise standard neoclassical translog cost function is assumed to have two components: one stochastic portion representing statistical noise, and a second one-sided component meant to capture deviations from optimal cost-minimizing behaviour. The cost function is thus converted into a minimum cost frontier with observations on the frontier deemed fully cost efficient. Deviations from the frontier are then interpreted as cost inefficiency reflecting a failure to properly minimize costs. We seek to explain these deviations with several firm- and policy-specific characteristics in a model based on the cost frontier dual of Battese and Coelli's (1995) SFA production frontier. Thus, we construct a variable cost frontier representing "best-practice" over the period, and calculate firm- and time-specific relative variable cost inefficiency as an observation's distance the cost frontier.

The dataset is described in Section 6, and we note that it spans the period prior to the implementation of the tradable SO_2 scheme (1990-1994), the Phase I period during which only Table A plants were regulated (1995-1999), and Phase II (2000 onward) during which all remaining plants in the sample came under regulation. We are therefore able to perform tests on the equality of inefficiency estimates across both time and policy-determined firm subsets to determine the impact of the allowance scheme on firm-level cost efficiency. Using data on firm characteristics (e.g. size, percentage of capacity that is coal-fired, and percentage of coal-fired capacity that has scrubbers installed) and policy features (scheme inclusion and permit prices), we estimate the effects of firm- and policy-specific qualities on cost inefficiency estimates.

Our results indicate that the effects of inclusion in the CAAA 1990's tradable allowance scheme on firms' cost-minimizing ability was insignificant and that cost efficiency improved with higher permit prices. We note that this was presumably due to a combination of reasons associated with the fixed capital stock and SO_2-specific scrubbing technology, shifts in the minimum cost frontier itself, as well as changes in managerial behaviour and a reduction in x-inefficiencies. We also find, as expected, that small firms were less cost efficient than their larger counterparts, and that a firm's percentage of coal capacity shared a positive relationship with cost efficiency.

We provide some background on the CAAA 1990 and the tradable SO_2 allowance scheme in the next section, followed by a literature review of related work in the fields of producer performance and environmental policy analysis in Section 3. In Section 4 we define and discuss cost efficiency and the SFA methodology, after which we present our model in Section 5. We describe the data in Section 6 and present our results in Section 7. Conclusions are made in Section 8.

2 The Clean Air Act: A Brief History

The Clean Air Act (CAA), originally passed in the United States in 1963, focused primarily on the reduction of sulfur dioxide (SO_2) and nitrous oxides (NO_x) emissions because of their responsibility for the acid deposition problem, particularly in the northeastern United States. Several authors have studied various aspects of the CAA and its amendments (see Joskow et al. (1998), Stavins (2003), Rezek and Blair (2005), Schmalensee et al. (1998) and others mentioned in the following section), though few have addressed firm-level responses to the programme, and none have, to our knowledge, addressed the issue of cost inefficiency impacts specifically. The purpose of this section is to describe the history and implementation of the CAA in order to motivate the modeling and testing assumptions made in later sections.

The Clean Air Act marked the first federal regulation of SO_2 and NO_x emissions as well as other pollutants in the USA. At the outset, and up until the onset of the allowance trading scheme created by the Clean Air Act Amendments of 1990, the emissions regulations of the CAA were all of the command-and-control type. In 1970, the Environmental Protection Agency (EPA) was granted the power to establish enforceable air quality standards, and in 1971, the Agency established New Source Performance Standards (NSPS). The NSPS permitted coal-fired utility boilers built after 17 August 1971 to emit no more than 1.2 pounds of SO_2 per million Btu of heat input. NO_x requirements were less rigid, as they allowed from 0.2 to 0.9 pounds per million Btu, depending on the fuel input and the combustion technology employed.

The CAA was amended in 1977 to require that States establish limits on existing pollution sources in regions not attaining the goals of the original Act. In 1979, the EPA established the Revised New Source Performance Standards (RNSPS), which retained the provisions of the NSPS, but required that all new or modified boilers reduce SO_2 emissions by 90%, unless such a reduction would bring emissions below 0.6 pounds per million Btu, when a 70-90% reduction is acceptable. The CAA was again amended in 1990 to create America's first national, long-term environmental programme based around the trading of emissions permits. It is this amendment and its cap-and-trade scheme that serve as the focus of our research and, as such, the remainder of this section.

2.1 Title IV: Clean Air Act Amendments 1990

Title IV of the CAAA 1990 prescribed a 10 million ton reduction in SO_2 emissions and a 2 million ton reduction in NO_x emissions by electric utilities from 1980 levels by the year 2010 (Clean Air Act). The amendment established an emissions cap and prescribed a two-phase implementation. Phase I covered the years 1995 through 1999 and capped the total emissions of 261 "Table A" units[1]. These Table A units, the dirtiest generating units in the US, accounted for 17 percent of the nation's capacity in 1990 (Schmalensee et. al., 1998) and were required to reduce SO_2 emissions to a rate of 2.5 lbs of SO_2 per mmBtu of heat input times 1985 baseline heat input (see Ellerman and Montero (2005)). Phase II, which began in 2000, brought every fossil fuel generating unit with capacity of 25 MW or greater under regulation, regardless of past emissions rates. The emissions cap was tightened during Phase II such that all units were required to reduce emissions to a rate of 1.2 lbs of SO_2 per mmBtu of thermal heat input times 1985 baseline (see Ellerman and Montero (2005) and Clean Air Act.) To encourage firms to adopt their own least-cost compliance methods[2], Title IV established a market-based tradable permit system that deviated quite radically from the more prevalent command-and-control alternatives.

The permits (or allowances) are and have been allocated at no cost to the owners of the units regulated under Phase I and Phase II based on 1985 emissions levels. Each allowance represents the right to emit one ton of SO_2, and each has a specific year, called its vintage, during which it can first be used as an emissions right. Once a firm has been allocated its permits, it is free to sell them, use them, or bank them for future use or sale. In order to encourage the development of a permit market, Title IV implemented annual allowance auctions[3] to be managed by the EPA. The goal of Title IV, then, was to maximize economic efficiency subject to the chosen environmental standards through the use of a free market for emissions permits. For more details on the trading and auctioning of SO_2 permits, see Joskow et. al. (1998), Burtraw (1996), EPA (July, 2000) or Chicago Climate Exchange (2004).

Before allowance trading began, there was a wide range of estimates as to what the market price of the allowances would actually be, with some figures as high as $1,500 (Insley, 2003). In 1990, the EPA provided $750 as their best estimate of what the permits would cost (Bohi and Burtraw, 1997). After the first inter-utility trades were recorded in 1992 between $250 and $300, the Electric Power Research Institute predicted in 1993 that under favorable trading conditions, the equilibrium allowance price would be approximately $273 (Ellerman and Montero, 1998). This, in fact, was not the case, and permit prices did not approach predicted levels until well into Phase II of the programme. The oft-cited reason for the divergence in predicted and observed marginal abatement costs is the deregulation of the rail industry which occurred in the early 1990s and brought down the price of lower-sulfur western coal.

Tradable permit schemes are appealing from an economic perspective (particularly as compared to more standard command-and-control policies) as they promote emissions reduction at least cost through the establishment of a free allowance market. Here, the firms with the lowest marginal abatement costs are capable of choosing to reduce their emissions in favor of either banking or selling their permits, while those firms with higher marginal abatement costs are able to purchase the permits they require for their emissions on the allowance market. The ability of firms to select individual abatement methods has been important in the electricity generation industry, as deregulation has caused increases in

[1] A "unit" refers to the combustion unit. A firm may own several plants, each composed of several generating units. Table A units were all coal-fired.
[2] The most prevalent compliance methods have been fuel-switching, particularly high- to low-sulfur coal switching, scrubbers and permit use (EIA (1997)).
[3] These auctions account for 2.8 percent of the year's total allowance allocation. (Clean Air Act)

competition over the past twenty five years, and cost minimization has become increasingly important.

The political success, academic justification, and economic efficiency achievements of the allowance trading scheme of Title IV of the Clean Air Act (see, e.g., Ellerman and Montero (2005), Stavins (1998) and Schmalensee et al. (1998)) provided motivation for the design of the European Union's CO_2 emissions trading scheme (EU ETS) and the proposed Kyoto trading scheme for beyond 2008. The necessity for emissions regulation is well understood (see, e.g. The Stern Review) and the theoretical superiority of market-based instruments for this purpose has been well documented. It is our objective to add to the knowledge on the firm-level cost inefficiency impacts of the longest-running market-based emissions scheme. We provide an overview of the related literature on producer performance and the impact of environmental policy in the following section before discussing stochastic frontier analysis and our model specifically.

3 Related Work

The related literature can be broken into two categories: those papers dealing with producer performance in the electricity generating industry generally, and those related to the Clean Air Act Amendments and their more specific impacts on producer performance. A brief review of each is provided below.

3.1 General Producer Performance

In an early example of performance analysis, Joskow and Schmalensee (1987) used the fixed effects method[4] of Hausman and Taylor (1981) to estimate the thermal efficiency and unit reliability of coal-burning units in the US using data for 1960-1980. They did not analyze the effects of environmental policy and many of their variables are relevant only at the unit level, but they are mentioned here for their efforts at understanding the performance-determining characteristics of the industry.

Most of the more recent literature on producer performance in electricity generation[5] has focused on the issues of regulation and ownership. Michael Pollit (1995), for example, assessed the productive efficiency of a cross section of 95 electric utilities operating in 1985 in order to determine whether private firms operated at higher levels of productive efficiency than public firms. He used two methods (one parametric based on Atkinson and Halvorsen (1986) and the nonparametric DEA methodology based on Färe et al. (1985)), and found no significant difference in productive efficiency by ownership type once scale effects are accounted for.

[4] This method, based on least squared dummy variables (LSDV) was a precursor to the SFA methodology, and is similar in practice, though LSDV does not require the individual effects to be one-sided, and thus does not estimate a proper lower or upper bound for the data. Stochastic frontiers can, in fact, be estimated via LSDV, requiring only a normalization of the estimated intercept parameter, which can then be used to find strictly positive inefficiency estimates. See Kumbhakar and Lovell (2000) for further discussion.

[5] While not the focus of this section, we note that there is a rich literature on producer performance within electricity distribution employing SFA and DEA methods to identify and explain cost and technical efficiency (see e.g. Estache et al. (2004), Burns and Weyman-Jones (1996), Jamasb and Pollit (2000)).

In two separate chapters of his 1995 book, Pollitt used four different methods (stochastic frontier analysis (SFA), data envelopment analysis (DEA), parametric programming, and a deterministic statistical frontier) to assess the technical and overall productive efficiency of an international cross-section of 768 electricity generating plants operating in 1989. His objective was to determine whether privately owned power plants are operated 1) with a higher level of technical efficiency and/or 2) with greater cost or allocative efficiency. He found little significant difference in technical inefficiency estimates when the data was estimated in subsets determined by load factors and technology and country-specific factors were accounted for. He found slightly lower cost and greater allocative efficiency in private baseload power plants, however, indicating more efficient technology investment decisions in privately-owned utilities, though he noted the cost difference varied with estimation methodology, and the most efficient privately-owned plants operated with cost efficiency levels similar to the most efficient publicly-owned plants.

Knittel (2002) used the SFA methodology to assess the effect of various state-level regulatory programs on the technical inefficiency of coal and natural gas generating units in the US electricity generating industry. He used the Battese and Coelli (1995) model on an annual plant-level panel dataset covering the years 1981-1996, using dummies for various regulatory programs as explanatory variables for the inefficiency estimates. He found modified fuel cost pass-through programs were associated with higher technical efficiency than traditional cost pass-through methods, and that schemes directly incentivizing thermal efficiency and unit availability are also tied to higher efficiency estimates.

Rungsuriyawiboon and Coelli (2004) also analyzed the impact of regulatory reform on the performance of US electricity generating firms, but focused on its effects on total factor productivity (TFP). They used three methods (Tornqvist index numbers, a stochastic frontier and a distance function approach) to assess the impacts of incentive regulation on TFP change on a panel dataset of 61 firms for the years 1986 to 1998. They rejected the SFA approach, citing violations of cost-minimizing behaviour, and found productivity growth to be negatively related to the introduction of incentive regulation, thus questioning whether it was the incentive methods themselves or their implementation that had failed over the period.

Two papers that use the SFA approach to asses cost efficiency specifically (as opposed to technical efficiency or TFP growth as above) within electricity generation are mentioned here briefly: Kumbhakar (1996) and Hiebert (2002). Kumbhakar (1996) proposed a modified version of Baltagi and Griffin's (1988) generalized error component model for panel data to allow for heteroskedasticity[6] in the inefficiency error component. The model, which includes the cost share equations, was estimated in several steps on a panel dataset consisting of 10 Texan utilities for the years 1980-1996. Fuel inputs are not disaggregated, and environmental regulations are not considered. Kumbhakar (1996) estimated the overall cost efficiency for firms in his sample to be 76.34%, and found small increasing returns to scale as well as "very little" technical change.

The final general producer performance paper we mention is that of Hiebert (2002), who modelled the cost efficiency of generating plants from 1988-1997 using the Battese and Coelli (1995) model in a manner similar to ours. He found plant efficiencies to be associated with ownership form, capacity utilization and the number of plants owned by a utility, but did

[6] It should be noted that our model does not incorporate heteroskedasticity, though does attempt to account for firm size and other firm-specific variables (as inefficiency is often found to vary with these characteristics) in the modelling of the inefficiency component.

not address the issue of environmental policy. His conclusions note that regulators should take into account the effect of multiple-plant ownership on operating efficiency, providing motivation for both our firm-level aggregation and inclusion of coal's percentage of total capacity as an explanatory variable for our cost inefficiency estimates.

3.2 Environmental Policy and Firm Performance

The impact of environmental policy on the electricity industry has been of interest for decades. In one of the earliest empirical papers, Gollop and Roberts (1983) studied the effect of one of the command-and-control precursors to the SO_2 trading scheme, namely the CAAA 1970, on productivity growth in US electric power generation. The authors estimated an industry cost function with data for 56 utilities for the years 1973-1979 and broke productivity growth into three components: scale economies, environmental regulations, and technical change. They found the CAAA 1970 regulations led to "significantly higher" generating costs, while also rejecting both the economies of scale hypothesis (implying output growth is important for productivity growth in the electricity industry), and the hypothesis of no technical change.

Färe et al. (1989b) looked at a sample of 23 US electric power plants in two time periods (1969 and 1977) in an attempt to identify a decline in technical efficiency attributable to environmental regulation. Each of their plants installed precipitators between 1969 and 1977, leading the authors to attribute changes in technical efficiency (measured via a DEA precursor method) to the installation of pollution control equipment. They find a small decline in technical efficiency, but the change is not significant. As noted by Pollitt (1995), however, this result is not particularly useful, given that retrofitting power plants will obviously increase costs. Our study is therefore different in two ways: 1) we address the impact of the cap-and-trade SO_2 scheme which allowed firms to choose their own least-cost compliance methods and 2) we assess variable fuel costs, abstracting from the question of sunk capital costs and focusing on generation.

More recently, Lee (2002) analyzed the effects of the pre-CAAA 1990 command-and-control SO_2 emissions regulations on electric utilities. He used a generalized cost function analysis that incorporated input shadow prices on a panel of mostly coal-burning electric utilities from 1975-1990. His results imply that firms do, in fact, fail to fully minimize costs, which provides motivation for our investigation of cost inefficiency and its determinants. He also provided estimates of marginal abatement costs and percentage change in input demand and productivity growth. Unlike the authors below, Lee (2002) analyzed a dataset aggregated to the firm-level, rather than the plant level.

3.2.1 SO_2 Trading Scheme and Producer Performance

Perhaps more relevant to our research is the literature related to the CAAA 1990's SO_2 scheme. Because it was the first practical application of the previously textbook-only model of market-based environmental policy, much of the early literature on the tradable sulphur scheme focused on the price of allowances and the permit market itself. Ellerman and Montero (1998) provided a review of reasons explaining lower-than-expected allowance prices (rail deregulation and access to lower-sulphur western coal being the most prominent), and Joskow et al. (1998) considered several operational aspects of the allowance market (e.g.

bid-ask spreads, purchase quantities and prices, etc), concluding that the market does, in fact, operate efficiently. Further discussion of the cap-and-trade scheme, marginal abatement cost estimates, the role of rail deregulation and availability of cleaner-burning western coal at lower costs, as well as several potential caveats for the design of future policy can be found in Schmalensee et al. (1998) and Carlson et al. (2000). More recent analyses of various aspects of the allowance market's efficiency can be found in Ellerman and Montero (2005), Albrecht et al. (2006), and Helfand et al. (2006).

As far as producer performance in response to the SO_2 allowance scheme has been concerned, the vast majority of the literature has focused on the distance function modelling approach and an alternative measure of technical efficiency that incorporates environmental outputs. Distance functions (which can be used in either SFA or DEA context) allow the modelling of joint outputs and can have either an input- or output-orientation. They are dual in the former case to the cost function (see Färe and Grosskopf (1990)) and in the latter to the revenue function (see Färe et al. (1992)) and yield deflated shadow prices of all inputs or outputs, respectively, through a dual Shephard's lemma (see Färe and Primont (1995) and Färe and Grosskopf (2004)). The distance function itself was originally introduced by Shephard (1953, 1970), but applied and extended more recently, particularly in the context of electricity generation and in the presence of at least one undesirable output, by Färe (1988) and Färe et al. (1989a, 1992, 2005).

This distance function approach followed from efforts to include environmental changes into productivity measures. Pittman (1983), for example, extended the work of Caves, Christensen and Diewert (1982) to include undesirable outputs, calculating shadow prices[7] for "bad" outputs to be used in the TFP measure, while Färe et al. (1989a) proposed a nonparametric method of productivity measurement reflecting the lack of "free disposability[8]" of bads in the presence of environmental regulation. This latter approach does not require the calculation of shadow prices, though from Färe et al.'s 1992 paper onwards, much of the literature has employed shadow price calculation methods for determining either the marginal cost of abatement or the marginal costs of emissions-related inputs. We briefly discuss some of the papers related to the analysis of the SO_2 allowance scheme below. See Atkinson and Dorfman's (2005) introduction for further references.

One of the early applications of the dual-output method to SO_2 regulation and electric utility performance was that of Yaisawarng and Klein (1994), who used the DEA method to assess the effects of SO_2 controls on productivity in US electricity generating plants. Their dataset covered the years 1985-1989, and they looked at the effects of command-and-control SO_2 regulation on the TFP of coal-burning plants in the U.S. using a Malmquist index which they decomposed into plant efficiency, scale efficiency and technology change effects. They argued that expenditure on emissions reduction and resulting environmental quality improvement should be accounted for in productivity change, and that the conventional measure of TFP which would state that higher costs and lower "good" output implies a reduction in productivity.

Coggins and Swinton (1996) and Swinton (1998) applied Färe and Grosskopf's (1990) methodology on plant-level data to calculate the shadow price of SO_2 emissions, providing an estimate of marginal abatement costs and thus a prediction for SO_2 prices. Rezek and Blair

[7] In most of this literature, shadow prices are considered to be the opportunity cost of emissions abatement in terms of foregone electricity output.
[8] Free disposability would imply that "bad" outputs could be reduced without a reduction in "good" outputs.

(2005) also used an output distance function approach to estimate emissions shadow prices and to assess the evolution of marginal cost heterogeneity across the 91 Table A plants and 5 years (1995-1999) of their panel dataset. They found a reduction in the variance of plant-level shadow prices upon the implementation of Phase 1, with a decline in their convergence later on, implying that cost-savings from the allowance scheme were realized relatively early on. Vardanyan and Noh (2006) provided estimates of the SO_2 shadow price for a panel of 187 generating units for the years 1997-1999 obtained via different functional form and mapping vector assumptions. They found, as has the literature generally, a wide range of shadow price estimates. Vardanyan and Noh (2006) explain this divergence in shadow price estimates by their "extreme sensitivity" to functional form and mapping regime, suggesting caution when interpreting estimates for policy recommendation purposes.

In addition to marginal abatement cost estimates, this literature provides estimation techniques for an alternative definition of technical efficiency. This definition incorporates changes in both good and bad outputs, such that expenditure on emissions reduction is not considered wasted resources as in conventional technical efficiency measures. Papers measuring technical efficiency in this context generally rely on the directional output distance approach of Färe et al. (1994), where technical efficiency is a radial measure incorporating both environmental and productive efficiency, implying that inefficient firms have a positive-valued directional output distance function reflecting the increase in good output or reduction in bad output that the firm could achieve, were it on the efficient frontier.

Pasurka (2006) focused on changes in bad output production (specifically SO_2 and NO_x) in a multiple output setting, and used directional output distance functions to decompose changes in a manner similar to growth accounting studies. He divided changes in emissions output into several components (technical change, technical efficiency, changes in input quantities, and changes in the emissions' proportion of total output), and found within his panel of 92 coal-fired power plants from 1987-1995 that SO_2 reductions were generally associated with changes in the output mix. He also found technical change to be associated with increased SO_2 and NO_x emissions, and changes in technical efficiency to be related to small decreases in bad output production. Färe et al. (2005) also use a directional output distance function approach to analyze the technical efficiency of 209 electric utilities before (1993) and after (1997) the implementation of Phase 1 of the allowance trading scheme. They estimated the shadow price of SO_2 and the output elasticity of substitution between the good (electricity) and bad (SO_2) outputs, and found increasing inefficiency in production, and increasing difficulty in substituting reductions in electricity production for reductions in SO_2 from 1993 to 1997.

Another recent paper by Lee (2005) used the input distance function approach to calculate the SO_2 shadow price and substitution elasticity between sulphur and capital. He defined the shadow price of sulphur emissions slightly differently as the opportunity cost of SO_2 reduction in terms of foregone capital, rather than foregone electricity output. Lee (2005) noted that previous SO_2 shadow price estimate failed to account for the substitution possibility between sulphur and productive capital in attaining sulphur regulation limits. He argued that sulphur reduction can be achieved through capital investments (e.g. boiler upgrades, shifting to other production processes), implying a lower cost of regulation when sulphur-capital substitutability is high, a situation which he noted might also provide support for the Porter Hypothesis[9] (Porter (1991) and Porter and van der Linde (1995)) in the long

[9] The Porter Hypothesis suggests that increased stringency of environmental regulation my lead to private firm benefits via innovation and re-organization in the long run.

run. Lee's (2005) results, using data for 51 coal-fired units over the year 1977-1986, suggest a "relatively high" degree of substitutability between capital and sulphur.

3.3 Summary

The literature on producer performance and its response to environmental policy in the US is thus wide and varied. The main themes, however, appear to be differential firm response to incentive regulation, the possibility that ownership may play a role in performance, and the technical efficiency impact, particularly at the plant level, of environmental regulation. Our contribution is to estimate and assess impacts of the SO_2 allowance scheme and other firm characteristics on generating firm's cost-minimizing behaviour through the analysis of firm-level cost inefficiency estimates calculated via a stochastic cost frontier model.

Unlike most papers mentioned above, we have aggregated our data to the firm level (the level at which input and operating decisions are made), we assume a single good output (electricity), and we incorporate both the sulphur allowance price and a dummy to delineate between periods of regulation and non-regulation, allowing us to assess the impact of both policy inclusion and stringency on cost inefficiency. We are able to test whether the CAAA 1990's allowance program had an effect on relative variable cost efficiency estimates, and to determine which firm and policy characteristics have the greatest impacts on cost efficiency estimates. While not explicitly accounted for in our model, we bear in mind the potential impact of state-level regulation and ownership type in the interpretation of our results as indicated by the literature above. We turn now to a discussion of cost efficiency and the stochastic frontier analysis (SFA) methodology.

4 Cost Efficiency and its Estimation

Producer performance has interested economists for decades, though the literature on the theory and application of the measurement of productive efficiency has grown immensely in recent years[10]. In general, the term productive efficiency refers to the varying degrees of success a firm sees in obtaining its objective through the allocation of inputs and outputs. As Kumbhakar and Lovell (2000) note, the firm's objective could be simply to maximize output for a given set of inputs (or, similarly, to minimize input use in the production of given outputs), in which case productive efficiency is known as *technical efficiency*, or the firm could have economic objectives such as cost minimization, revenue maximization or profit maximization, in which case productive efficiency refers to *economic (cost, revenue or profit) efficiency*.

We focus on economic efficiency, and specifically cost efficiency, as we are interested in the effects of the CAAA 1990's tradable permit scheme and various firm characteristics on producers' cost-minimizing behaviour, particularly given regulatory objectives related to energy prices and questions relating to competition within the industry. Technical efficiency measurement uses input and output quantity data in methods involving a production function

[10] See, for example, the introductory chapter of Kumbhakar and Lovell (2000) and citations therein, as well as Coelli et al. (2003a) for a survey.

to ascertain efficiency measures; it does not make use of input or output price data, nor does it make any assumptions about the firms' objectives. Economic efficiency, however, requires an assumption regarding firms' economic objectives, and here we assume that electricity generating firms are cost-minimisers. Because of the non-storability of electricity as an output, and because of the requirement that the grid to be balanced at all times, we assume electricity generating firms are faced with a fixed, exogenous output requirement, leaving input quantities as their main decision variables in the short run[11]. Thus, output quantity selection for revenue maximization is ruled out, leaving firms with the objective of profit maximization subject to a fixed output quantity, which is equivalent to cost-minimization.

The method we use for our efficiency estimation is known as stochastic frontier analysis (SFA), and we use the data described in Section 6 to estimate time- and firm-specific cost inefficiency. We then assess the effects of several policy and firm characteristics on these inefficiency estimates. First, however, we provide a brief background on cost efficiency, followed by an explanation of SFA methods, their development and the alternatives. The explanations below reflect our assumption of multiple inputs (coal, oil and gas) and a single output (electricity measured in kWh).

4.1 Cost Efficiency[12]

Cost efficiency is an input-oriented measure of producer performance that uses a cost frontier as a benchmark. The cost frontier defines the minimum achievable cost of producing a given output, *y*, with an input vector $x = (x_1, ..., x_N) \in R_+^N$, and a strictly positive input price vector $w = (w_1, ..., w_N) \in R_{++}^N$. We follow Kumbhakar and Lovell (2000) in defining the cost frontier as

$$\text{c(y,w)} = \min_{\text{x}} \left\{ \sum_n w_n x_n : x \in L(y) \right\} = \min_{\text{x}} \left\{ \sum_n w_n x_n : y \le f(x) \right\} \qquad \textbf{Eq. 1}$$

where *L(y)* denotes the feasible input sets for producing output *y*. *f(x)* is the production frontier defined as

$$f(x) = \max \{ y : y \in P(x) \} = \max \{ y : x \in L(y) \}, \qquad \textbf{Eq.1}$$

where *P(x)* denotes the output possibilities for each set of inputs *x*. The cost frontier thereby defines the fully-efficient minimum cost production technology, and so, by definition, bounds observed producer expenditure from below and satisfies the following properties (Kumbhakar and Lovell (2000), p. 34):

c1: $c(0,w)=0$ and $c(y,w)>0$ for $y \ge 0$.
c2: $c(y,\lambda w)=\lambda c(y,w)$ for $\lambda>0$.
c3: $c(y,w') \ge c(y,w)$ for $w' \ge w$.

[11] In reality, each firm has some degree of control over output quantities, as firms submit price/quantity bids to the ISO for service provision. This allows for increased competition in the industry, however, to some extent reinforcing the exogeneity of each firm's output quantity.
[12] This section is based on material in Kumbhakar and Lovell (2000) and Chambers (1988).

c4: $c(y,w)$ is concave in w.
c5: $c(y,w)$ is continuous in w.
c6: $c(\lambda y,w) \leq c(y,w)$ for $0 \leq \lambda \leq 1$. (weak monotonicity)
c7: $c(y,w)$ is lower semicontinuous in y.

Cost efficiency can be measured as the ratio of minimum cost (as given by the cost frontier) to actual cost, i.e.

$$CE(y,x,w) = \frac{c(y,w)}{\sum_n w_n x_n}. \qquad \textbf{Eq. 2}$$

Kumbhakar and Lovell (2000) note that the measure of cost efficiency[13] has the following properties:

$$CE1: 0 < CE(y,x,w) \leq 1, \textit{with}\, CE(y,x,w) = 1 \Leftrightarrow x = x(y,w) \text{ so that } w^T x = c(y,w)$$

i.e. CE is positive and less than one unless the firm uses the cost-minimizing input vector, in which case CE=1.

$$CE2: CE(y,\lambda x,w) = \lambda^{-1} CE(y,x,w) \textit{ for } \lambda > 0$$

i.e. CE is homogenous of degree -1 in inputs.

$$CE3: CE(\lambda y,x,w) \geq CE(y,x,w) \textit{ for } \lambda \geq 0$$

i.e. CE is nondecreasing in outputs for a given set of inputs.

$$CE4: CE(y,x,\lambda w) = CE(y,x,w) \textit{ for } \lambda > 0$$

i.e. CE is homogenous of degree zero in input prices.

Cost efficiency can be broken into two components- technical and allocative efficiency, which share a relationship as follows:

$$CE(y,x,w) = TE_I(y,x) * AE_I(y,x,w) \qquad \textbf{Eq. 3}$$

where TE refers to technical efficiency, AE refers to allocative efficiency, and the subscript I denotes input orientation. TE is therefore necessary, but not sufficient for cost efficiency. The breakdown of cost efficiency empirically into its technical and allocative components requires the estimation of the cost share equations associated with the cost frontier, and is left in our case to future research.

Because our analysis is short run in nature, capital is included as a quasi-fixed factor, and we focus on cost efficiency as measured relative to a *variable* cost frontier. Here

$$vc(y,w,z) = \min_x \left\{ \sum_n w_n x_n : (y,x,z) \in GR \right\}, \qquad \textbf{Eq. 4}$$

[13] Note that the Frontier 4.1 software that we use in our analysis provides an estimate of cost *inefficiency*, which is just the inverse (1/CE(y,x,w)) of the cost efficiency measure discussed here.

where z is a vector of fixed (or quasi-fixed) inputs[14] and GR describes the set of technologically feasible input-output combinations. It satisfies the same properties as a total cost frontier. A measure of variable cost efficiency relative to the variable cost function is given by the ratio of minimum variable cost to actual variable costs, i.e.

$$VCE(y,w,z) = \frac{vc(y,w,z)}{\sum_n w_n x_n}, \quad \textbf{Eq. 5}$$

which shares the same properties as CE.

The above should serve to clarify cost inefficiency as a residual concept based upon the behavioural assumption of cost-minimization. The source of this inefficiency could be understood as the manifestation of unmeasured (or improperly measured) inputs or outputs (as by Stigler (1976) and de Alessi (1983)) or "x-inefficiency" (e.g. Simon (1955, 1957) and Leibenstein (1966)), though its interpretation is not the focus of this paper. Rather, we assume that firms may, in fact, deviate from optimal cost-minimizing operation at the cost frontier[15], and that these deviations reflect measurable inefficiencies that can be explained by firm- and policy-specific characteristics. We seek, then, to quantify these cost inefficiencies and to determine how or whether they were affected by the SO_2 trading scheme and various firm characteristics.

It should be noted that the measures we present are measures of *relative* cost inefficiency. We follow Farrell (1957) who argues in favour of constructing the cost frontier, c(y,w) from observed data rather than using a theoretical function based on assumed maximum economic efficiency. As such, the cost frontier in what follows is defined by the most efficient firm observed, implying that full efficiency does not preclude the possibility of improved performance, but rather that an observation with zero inefficiency is the best performing in the sample. In order to compare the optimizing success across time and between firms, we use stochastic frontier analysis (SFA) to calculate the distance from firms' observed cost positions to the minimum cost frontier. Before discussing SFA in greater detail, however, we briefly consider some of the alternative techniques for efficiency measurement.

4.2 Methods of Frontier Efficiency Estimation

Efficiency can be measured with several methods depending on data availability (cross section versus panel data, as well as price versus quantity data), frontier assumptions, and whether a parametric or non-parametric technique is preferred. We discuss the most prevalent methods below, starting with those using deterministic production frontiers[16,17] (i.e. corrected OLS (COLS) and modified OLS (MOLS)) before discussing the nonparametric

[14] In our case, capital, as measured by generating capacity.

[15] Lee (2002), for example, rejects the hypothesis of cost-minimization for the electric utilities in his sample.

[16] Deterministic frontiers ignore random shocks and assume all variation in producer behaviour can be attributed to cost inefficiency.

[17] We discuss production frontiers rather than cost frontiers in what follows because much of the original work dealt with production frontiers and the measurement of technical inefficiency. The extension to a cost frontier setting is straightforward.

data envelopment analysis (DEA) technique[18]. The discussion of stochastic frontier analysis (SFA) is granted greater detail below.

4.2.1 Deterministic Frontiers: Corrected OLS and Modified OLS

Corrected OLS (COLS) is a method dating back to Winsten (1957), who proposed a two-step estimation procedure for the deterministic production frontier model

$$\ln y_i = \beta_o + \sum_n \beta_n \ln x_{ni} - u_i .$$

In the first stage, OLS provides consistent estimates of all parameters and unbiased estimates of the slope parameters, β_n, using the second and third moments of the OLS residuals (see Ruggiero (1999)). The biased OLS estimate of the intercept parameter, β_o, is shifted up in the second stage, leaving the COLS intercept which is estimated by

$$\hat{\beta}_o * = \hat{\beta}_o + \max_i \{\hat{u}_i\}$$

where $\hat{u}_i$ are the OLS residuals from the first stage. The OLS residuals are also corrected such that

$$-\hat{u}_i * = \hat{u}_i - \max_i \{\hat{u}_i\} .$$

Thus, the COLS production frontier bounds the observed data from above, and the COLS residuals, $\hat{u}_i *$, are nonnegative values that provide consistent estimates of technical inefficiency of each producer. The COLS frontier is just a parallel shifted version of the OLS estimate, however, meaning that COLS assumes the frontier (i.e. "best practice") production function has the same shape as that of the central-tendency function[19]. This, combined with its lack of a stochastic component make COLS a somewhat less desirable procedure, despite its ease of implementation..

Modified OLS (MOLS) was proposed by Afriat (1972) and Richmond (1974) as an alternative to COLS. With this method, the error term of the deterministic production function, u_i, is assumed to follow a one-sided distribution, and then the frontier is estimated in a two-step procedure similar to that for COLS. Here, OLS is used in the first stage to estimate the slope parameters, β_n, efficiently and consistently. The OLS β_o is then modified as follows:

$$\hat{\beta}_o ** = \hat{\beta}_o + E(\hat{u}_i)$$

[18] Linear and quadratic programming methods of the sort introduced by Aigner and Chu (1968) also provide estimates of technical or cost efficiency from deterministic frontiers, but these methods calculate (rather than estimate) the frontier's parameters, making statistical inference difficult. For this reason, we do not discuss them here.

[19] This implies that COLS does not bound the data as tightly as possible.

and the first-stage OLS residuals are corrected in the opposite direction as in COLS, but here the corrected u_i term is

$$-\hat{u}_i{**} = \hat{u}_i - E(\hat{u}_i).$$

The MOLS frontier is still parallel to the OLS production technology, and there is still a chance that the MOLS frontier will not bound the data from above if there are producers with sufficiently large OLS residuals. The frontier is also still deterministic in nature. Because of the possibility of statistical noise, we believe it unwise to assign all deviation from the frontier to firm inefficiency, and instead focus our research on stochastic frontiers.

4.2.2 Non-Parametric Methods: DEA[20]

Data envelopment analysis (DEA) is a linear programming method influenced by Farrell's seminal 1957 article, and introduced by Charnes, Cooper and Rhodes (1978). The DEA frontier is constructed by joining extreme data points to ensure that observed producer behaviour is bounded from above in the case of the production frontier and below in the case of the cost frontier. Producer-specific efficiency is then evaluated relative to this frontier. The advantage of the DEA method is that neither a functional form nor a distribution need be assumed for the frontier or the efficiency term, respectively. Because DEA is an extreme point method, however, it is very sensitive to outliers and noise. Other disadvantages include the lack of testable hypotheses because of DEA's nonparametric nature, and the frontier's nonstochastic quality.

We reject the COLS and MOLS methods for their deterministic frontier specification, arguing that statistical noise should be incorporated into the cost frontier and the associated efficiency estimates. DEA and SFA are both prevalent in the applied efficiency analysis literature, and several authors (see, e.g. Carrington, Coelli and Groom (2002), Hollingsworth (2003) and Hattori et al. (2005)) have compared the results of the two methods, finding generally that their efficiency estimates are comparable but sensitive to data quality and model specification. Coelli et al. (2003a) note that neither method is necessarily preferable to the other and that differences in estimates can not be predicted beforehand.

Thus neither method is decidedly "better", though different estimates imply that estimated inefficiency rankings should perhaps receive greater attention than the inefficiency estimates themselves. DEA has been more prominent in applied studies of regulation, possibly for reasons of software availability and ease of calculation (see Coelli et al. 2003a). While SFA has its drawbacks (namely that efficiency estimation as a portion of a composed error term may be affected by the distributional assumptions regarding the error components), our objective of the estimation and testing of the effects of the SO_2 scheme and firm characteristics on firms' cost inefficiency is most readily achieved with the stochastic frontier method. We therefore employ the SFA technique and now turn to a discussion of its method and development.

4.2.3 Stochastic Frontier Analysis: History and Background[21]

[20] See Seiford and Thrall (1990) and Thanassoulis (2001) for further discussion of DEA methods.

SFA finds its roots in the literature on technical efficiency beginning in the 1950s with Koopmans (1951), Debreu (1951) and Shephard (1951). Koopmans formally defined technical efficiency, while Debreu and Shephard respectively introduced distance functions as a means of radially measuring output- and input-oriented distance from a frontier. Farell (1957) was the first to estimate productive efficiency, and formally defined cost efficiency and a means for decomposing it into technical and allocative subcomponents, though his method was one of linear programming, which eventually influenced Charnes, Cooper and Rhodes (1978) in their exposition of data envelopment analysis.

More importantly for SFA development, Farrell's (1957) work influenced that of Aigner and Chu (1968), Seitz (1971), Timmer (1971), Afriat (1972) and Richmond (1974). These authors effectively estimated a deterministic (i.e. noiseless) production frontier either via linear programming or adaptations of OLS as discussed above. Because of the frontier's deterministic nature, all deviations from it were interpreted as technical inefficiency. SFA itself which, as the name implies, allows for the presence of statistical noise via stochastic frontiers, originated in 1977 with two separate papers: one by Meeusen and van den Broeck (1977) and a second by Aigner, Lovell and Schmidt (1977). A third paper, Battese and Corra (1977), appeared shortly after these two, but, as noted by Kumbhakar and Lovell (2000), was probably influenced by Aigner, Lovell and Schmidt (1977).

These papers moved beyond deterministic frontier measures by introducing a two-component error term to the production function such that

$Y = f(x;\beta) * \exp\{v-u\}$. **Eq. 6**

Here v is the familiar noise term where $v \sim N(0,\sigma_v^2)$, and u is meant to capture technical inefficiency. Because firms are assumed to operate at or below their stochastic production frontier, $u \geq 0$, and u is therefore be given a one-sided distribution. Meeusen and van den Broeck (1977) chose an exponential distribution for u, while Battesse and Corra (1977) gave it a half normal distribution, and Aigner, Lovell and Schmidt (1977) considered both of these distributions. Kumbhakar and Lovell (2000) note that while mean inefficiency estimates may be sensitive to the choice of distribution, the ranking of efficiencies is largely unaffected. These original authors represented estimates of sample mean technical inefficiency by $E(-u) = E(v-u) = -\left(2/\pi\right)^{1/2}\sigma_u$ when v was assigned the half-normal distribution, and by $E(-u) = E(v-u) = -\sigma_u$ when v was assumed to be distributed exponentially.

In the case of cost efficiency, the estimation techniques are exactly the same, though firms are assumed to operate at or *above* the minimum cost frontier, so that the nonnegative efficiency term, u, is now added to the noise term, v, and the frontier is defined by

$VC = f(y,w;\beta) * \exp\{v+u\}$. **Eq. 7**

[21] Much of the information in this section comes from Kumbhakar and Lovell (2000), who provide an excellent and more detailed background on the origins and development of SFA.

The parameters to be estimated are then β, σ_v^2, and σ_u^2 with a distributional assumption for u that implies $(v+u)$ has a positive skew. SFA models are estimated by maximum likelihood techniques for reasons of statistical efficiency[22].

While increasingly flexible distributions for the one-sided inefficiency error component have been proposed (see e.g. Greene (1980a, b), Stevenson (1980) and Lee (1983)), the original exponential and half-normal distributions are the most common in applied work. In response to a paper by Førsund, Lovell and Schmidt (1980) which noted the inability of the SFA method to determine firm-specific inefficiencies, Jondrow et al. (1982) proposed producer-specific (rather than sample mean) measures of technical efficiency given a production frontier. They proposed the mean or the mode of the conditional distribution $[u_i|v_i\text{-}u_i]$ as the firm-specific measure of technical inefficiency relative to a production frontier.

More recent contributions to the SFA literature relate to time-varying inefficiency in a panel data setting (see, e.g. Cornwell, Schmidt and Sickles (1990), Kumbhakar (1990) and Battese and Coelli (1992)) and to methods for explaining efficiency differences. These latter methods are at the core of our research, and have fallen into two categories: those with two stages, where inefficiency is estimated in the first stage and then regressed against explanatory variables, and those with only one. where variables explaining inefficiency are incorporated into the estimation of the mean or variance of the inefficiency error component, u. These methods are discussed briefly in our exposition of the model in Section 5.

4.3 Stochastic Frontier Analysis: Method

Models in the literature vary in many ways, most notably in their ability to handle unbalanced panel data , the inclusion of quasi-fixed inputs, assumptions about the distribution of the one-sided efficiency error term, single versus simultaneous equations (i.e. cost frontier or cost frontier plus input share equations derived through Shephard's Lemma), whether or not efficiency is treated as time invariant or time-varying, and whether or not efficiency estimates are explained by a set of explanatory variables. Given our software[23], data and objectives, we focus on a single-equation cost frontier model for unbalanced panel data, and incorporate several explanatory variables for cost inefficiency. As noted, the frontier lies at the heart of SFA methods, and for input-oriented cost efficiency estimation, the cost frontier provides the standard against which firm-specific relative efficiency is measured. We first discuss the frontier, then the maximum likelihood method for the estimation of the relevant parameters, and finally the estimation of firm-specific cost efficiency.

For ease of notation, the following exposition assumes a balanced panel of I firms observed over T time periods, though our panel is in fact unbalanced. We begin with the variable cost frontier itself

$$\ln E_{it} \geq vc(y_{it}, w_{it}, z_{it}; \beta) + v_{it} \qquad \textbf{Eq. 8}$$

where $E_{it} = w_{it}^T x_{it} = \sum_n w_{nit} x_{nit} \geq vc(y_{it}, w_{it}, z_{it}; \beta)$ is the expenditure of firm i in period t and $vc(y_{iy}, w_{it}, z_{it}; \beta)$ is the deterministic kernel of the cost frontier common to all producers across

[22] See Kumbhakar and Lovell (2000), as well as Battese and Coelli (1992, 1993, 1995).
[23] Frontier 4.1 cannot accommodate simultaneous equations.

all periods[24], and v_{it} is a firm- and time-specific random noise component. Actual expenditure is written as

$$\ln E_{it} = vc(y_{it}, w_{it}, z_{it}; \beta) + v_{it} + u_{it} \qquad \textbf{Eq. 9}$$

where $u_i \geq 0$ represents firm-specific cost inefficiency. Note that our model, discussed in Section 5, follows the Battese and Coelli (1995) specification for panel data, and allows for firm- and time-specific inefficiency, u_{it}, which can also be expressed as a function of explanatory variables including time. Here, however, we write u_i simply as a firm-specific measure of cost efficiency in order to simplify notation in the exposition of the SFA method. Thus, the composed error term in what follows is defined as $\varepsilon_{it} = v_{it} + u_i$.

The frontier is then estimated via maximum likelihood estimation (MLE), which requires distributional assumptions for the error components, v_{it} and u_i. We illustrate with a normal-half normal model, though, as noted, several other options exist for the distribution of the efficiency error component. The assumptions are then

(i) $v_{it} \sim iid\ N(0, \sigma_v^2)$
(ii) $u_i \sim iid\ N^+(0, \sigma_u^2)$ (i.e. u_i is distributed as a non-negative half-normal random variable)
(iii) u_i and v_{it} are distributed independently of each other and of the regressors.

SFA via MLE relies on the maximization of a log likelihood function based on the marginal density function of the composed error term ε_{it} to provide estimates of the cost function parameters, β, and of σ_v^2 and σ_u^2. Estimates of u_i to be used in $CE_{it} = \exp\{\hat{u}_i\}$ are obtained from the conditional distribution of u_i given ε_i as noted by Jondrow et al. (1982) in what Kumbhakar and Lovell (2000) refer to as the JLMS decomposition. We briefly describe the MLE estimation procedure for SFA. For details and further discussion, see Kumbhakar and Lovell (2000) or the original papers cited here and therein.

Given assumption (i) and following Kumbhakar and Lovell (2000) in defining $\mathbf{v}=(v_1 \ldots, v_T)'$, the density function of $\mathbf{v}$ is

$$f(\mathbf{v}) = \frac{1}{(2\pi)^{T/2}\sigma_v^T} \exp\left\{-\frac{\mathbf{v}'\mathbf{v}}{2\sigma_v^2}\right\}. \qquad \textbf{Eq. 10}$$

The density function of u is

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \qquad \textbf{Eq. 11}$$

and given the assumption of independence of v_{it} and u_i, the joint density function of $\mathbf{v}$ and u is the product of their individual densities:

[24] We later assume the flexible translog functional form.

$$f(u,\mathbf{v}) = f(u)f(\mathbf{v}) = \frac{2}{2\pi^{T+1/2}\sigma_u\sigma_v^T}\exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{\mathbf{v}'\mathbf{v}}{2\sigma_v^2}\right\}. \qquad \textbf{Eq. 12}$$

The joint density function of u_i and ε where $\varepsilon = (v_1 - u, \ldots, v_T - u)'$ is then

$$f(u,\varepsilon) = \frac{2}{2\pi^{T+1/2}\sigma_u\sigma_v^T}\exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2} - \frac{\varepsilon'\varepsilon}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2}\right\} \qquad \textbf{Eq. 13}$$

where

$$\mu_* = \frac{T\sigma_u^2\bar{\varepsilon}}{\sigma_v^2 + T\sigma_u^2}, \qquad \sigma_*^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}, \quad \text{and} \quad \bar{\varepsilon} = \frac{1}{T}\sum_t \varepsilon_{it}.$$

The marginal density function of ε can then be written as:

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(\varepsilon,u)du \\ &= \frac{2\left[1-\Phi\left(-\mu_*/\sigma_*\right)\right]}{(2\pi)^{T/2}\sigma_v^{T-1}(\sigma_v^2+T\sigma_u^2)^{1/2}}\exp\left\{-\frac{\varepsilon'\varepsilon}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2}\right\} \end{aligned} \qquad \textbf{Eq. 14}$$

where Φ(*) is the standard normal cumulative distribution function.

From $f(\varepsilon_{it})$, the log likelihood function for I producers and T time periods is

$$\begin{aligned} \log L = &\text{constant} - \frac{I(T-1)}{2}\ln\sigma_v^2 - \frac{I}{2}\ln(\sigma_u^2 + T\sigma_v^2) \\ &+ \sum_i \ln\left[1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right] - \left(\frac{\sum_i \varepsilon_i'\varepsilon_i}{2\sigma_v^2}\right) + \frac{1}{2}\sum_i\left(\frac{\mu_{*i}}{\sigma_*}\right)^2 \end{aligned} \qquad \textbf{Eq. 15}$$

where $\mu_{*i} = \dfrac{T\sigma_u^2\varepsilon}{\sigma_v^2 + T\sigma_u^2}$, and σ_*^2 is as above. The log likelihood function 16 is then maximized with respect to the parameters to provide estimates of β, σ_u^2 and σ_v^2.[25]

In order to obtain estimates of producer-specific cost efficiency, we need information on u_i, which is contained in the estimates of $\varepsilon_{it}=v_{it}+u_i$. When $u_i \sim N^+(0,\sigma_u^2)$, the conditional

[25] Note that Frontier 4.1 equivalently minimizes the negative of the log Likelihood function.

distribution of $u_i|\varepsilon_i$ was shown by Jondrow et al. (1982) to be the density function of a variable distributed as $N^+(\mu_{*i},\sigma_*^2)$:

$$f(u_i \mid \varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_*\left[1-\Phi\left(\frac{-\mu_{*i}}{\sigma_*}\right)\right]}\exp\left\{-\frac{(u_i-\mu_{*i})^2}{2\sigma_*^2}\right\}.$$

The mean of this distribution can then serve as an estimator for u_i, and is given by[26]

$$\hat{u}_i = E(u_i \mid \varepsilon_i) = \mu_{*i} + \sigma_*\left[\frac{\phi\left(\frac{-\mu_{*i}}{\sigma_*}\right)}{1-\Phi\left(\frac{-\mu_{*i}}{\sigma_*}\right)}\right] \qquad \textbf{Eq. 16}$$

where $\phi(*)$ is the standard normal density function and $\Phi(*)$ is the standard normal cumulative distribution function. The u_i estimates are then used to estimate cost efficiency for each firm using[27]

$$CE_i = \frac{c(y,w;\beta)\exp\{v_{it}+u_i\}}{c(y,w;\beta)\exp\{v_{it}\}} = \exp\{\hat{u}_i\}. \qquad \textbf{Eq. 17}$$

As noted, our model allows for the calculation of time- and firm-specific cost inefficiency, and describes inefficiency as a function of several firm- and policy-specific explanatory variables. The model is discussed in the following section, but relies on the underlying techniques described above.

5 Model

In order to assess the sources of cost inefficiency and the effects of the SO_2 scheme, our stochastic frontier model expresses cost inefficiency, u_{it}, as a function of a set of explanatory variables, z_{it}, and a parameter vector, δ. The model is based on the cost frontier dual to the production frontier model proposed by Battese and Coelli (1995) for unbalanced panel data in which the inefficiency parameters, δ, are estimated along with the β parameters of the cost frontier in a single stage MLE estimation procedure.

[26] The mode of $u_i|\varepsilon_i$ can also be used as an estimator for u_i and is given by $M(u_i \mid \varepsilon_i) = \begin{cases} u_{*i} & \text{if } \sum_t \beta_t \varepsilon_{it} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

[27] An alternative point estimator for cost efficiency adapted from the minimum mean squared error predictor of technical efficiency proposed by Battese and Coelli (1988) is also possible, but not employed by the software we use for our estimation.

Earlier two-stage models (see e.g. Pitt and Lee (1981) and Kalirajan (1981)) attempted to explain technical inefficiency effects through a second-stage regression that used the predicted inefficiencies from the first stage as the dependent variable. Because the second stage is a regression of the *predicted* inefficiency effects on explanatory variables, the assumption of identically distributed inefficiency effects in the first-stage frontier estimation is violated and these two-stage methods lead to estimates that are less efficient than single-stage estimates (see Coelli (1996) and Battese and Coelli (1995)). In addition to the efficiency issue, Wang and Schmidt (2002) note a concern over bias in two-stage efficiency estimates caused by correlation between the z and x variables and by the neglect of z's effects on inefficiency in the first stage. Kumbhakar, Ghosh and McGuckin (1991), Reifschneider and Stevenson (1991) and Huang and Liu (1994) thus proposed single-stage technical inefficiency effects models that allowed for the simultaneous estimation of the parameters in both the stochastic production frontier and the inefficiency model. Battese and Coelli (1995) extended the cross sectional technical inefficiency effects model of Huang and Liu (1994) to allow for panel data, and this serves as the basis for our dual variable cost inefficiency effects model.

We implicitly assume the firms in our dataset operate with a cost-minimizing objective, an assumption that some claim (e.g. Rungsuriyawiboon and Coelli (2004)) is not reasonable in the US electricity generating industry. We argue, however, that assuming a cost-minimizing objective is acceptable given recent deregulatory activity in the industry and the rise in the number of independent power producers[28], as well as the largely exogenous nature of firms' output which precludes the assumption of revenue maximization. As noted, firms do not always achieve this minimum cost, either for reasons of randomness (i.e. because of a stochastic relationship between cost function variables and observed expenditure), or for reasons of inefficiency- either technical or allocative. (See e.g. Stigler (1976), Kumbhakar and Lovell (2000) and Lee (2002).) The model below focuses on variable fuel costs[29] and the related fuel-cost efficiency of electricity generating firms. We define an industry minimum variable cost frontier against which each firm's actual expenditure is compared in each period, and we analyze the effects of several policy and firm characteristics on inefficiency estimates.

The general variable cost frontier is specified as:

$$\ln VC_{it} = x_{it}\beta + (v_{it} + u_{it}) \qquad \text{for } i = 1,...,N \text{ and } t = 1,...,T \qquad \textbf{Eq. 18}$$

Where:

- $lnVC_{it}$ is the natural log of the variable cost of firm i in period t;
- x_{it} is a $kx1$ vector of explanatory variables (notably fuel prices, output quantity and capacity) for firm i in period t;
- β is a vector of parameters to be estimated;
- v_{it} represents the noise portion of the compound error term $\varepsilon_{it} = v_{it} + u_{it}$. It is assumed to capture any exogenous shocks in the production process, can be either positive or negative, and is assumed to be distributed $NID(0,\sigma_v^2)$ and independent of

[28] Total net generation from independent power producers (IPPS) rose from 58,222,000 MWh in 1995 to 1,118,870,000 MWh in 2004, while net generation from electric utilities fell from 2,994,529,000 MWh in 1995 to 2,505,231,000 MWh in 2004. (EIA, 2006)

[29] We assume the cost function is separable in labour, capital and fuel inputs.

- u_{it} is the non-negative portion of the error component meant to indicate the amount by which firm *i* in period *t* exceeds the minimum cost defined by the stochastic frontier. The u_{it}*'s* are assumed to be iid truncations at zero of $N(\mu_{it},\sigma_u^2)$, and
  - $\mu_{it} = z_{it}\delta$, where z_{it} is a (1 x *m*) vector of variables explaining cost inefficiency and δ is a an (*m* x 1) vector of parameters to be estimated. As long as inefficiency effects are stochastic, as defined below, z_{it} and x_{it} may share some explanatory variables.[30]

Given the one-sided nature of the inefficiency error component, u_{it}, we follow Battese and Coelli (1995) and write the cost inefficiency equation as follows:

$$u_{it} = z_{it}\delta + w_{it} \qquad \textbf{Eq. 19}$$

where the random variable w_{it} is the truncation at $-z_{it}\delta$ of a $N(0,\sigma^2)$ distribution such that $w_{it} \geq -z_{it}\delta$, making it possible for $u_{it} \sim N(z_{it},\delta)$.

Equations 19 and 20 are estimated simultaneously via maximum likelihood estimation (MLE) with the Frontier 4.1 software written by Coelli (1996). The error component density functions and corresponding likelihood function as derived by Piacenza (2002) are presented in Appendix 1. Given the parameter estimates, cost inefficiency for firm *i* in period *t* is defined as the ratio of observed cost to minimum stochastic frontier cost (where $u_{it}=0$), such that:

$$\begin{aligned} CE_{it} &= \frac{\ln VC(y_{it}, p_{it}, x_{it};\beta)\exp\{v_{it}+u_{it}\}}{\ln VC(y_{it}, p_{it}, x_{it};\beta)\exp\{v_{it}\}} \qquad \textbf{Eq. 20} \\ &= \exp\{u_{it}\} = \exp\{z_{it}\delta + w_{it}\} \end{aligned}$$

and $1 \leq CE_{it} < \infty$. CE_{it} will equal one when $u_{it}=0$ (i.e. when the firm is operating on the stochastic frontier) and increases as $u_{it} \to \infty$. Cost inefficiency predictions are based on the conditional expectation of $(u_{it} \mid \varepsilon_{it})$ and are provided in Appendix 1. They were derived by Piacenza (2002) as a modification of Battese and Coelli's (1993) technical efficiency estimator and a generalization of Jondrow et al. (1982) and Battese and Coelli (1988)'s estimators.

5.1 Cost Frontier Specification

Because of its flexibility and generality (a translog cost function is a second order Taylor approximation of an arbitrary underlying production technology), we chose the translog functional form for the deterministic kernel of the stochastic variable cost frontier[31]. The cost frontier portion of the model is then written as

[30] See Battese and Coelli (1993, 1995) and Piacenza (2006).

[31] The translog functional form was introduced by Christensen, Jorgensen and Lau (1971, 1973) and has been used extensively in empirical research.

$$\ln\frac{VC_{it}}{P_{g,it}} = \beta_0 + \beta_Y \ln Y_{it} + \frac{1}{2}\beta_{YY}(\ln Y_{it})^2 + \beta_K \ln K + \frac{1}{2}\beta_{KK}(\ln K_{it})^2 + \beta_{KY}\ln K_{it}\ln Y_{it}$$

$$+\sum_{f=1}^{2}\beta_f \ln\left(\frac{P_{f,it}}{P_{g,it}}\right) + \frac{1}{2}\sum_{f=1}^{2}\sum_{j=1}^{2}\beta_{fj}\ln\left(\frac{P_{f,it}}{P_{g,it}}\right)\ln\left(\frac{P_{j,it}}{P_{g,it}}\right) + \sum_{f=1}^{2}\beta_{fY}\ln\left(\frac{P_{f,it}}{P_{g,it}}\right)\ln Y_{it} \qquad \textbf{Eq. 212}$$

$$+\sum_{f=1}^{2}\beta_f \ln\left(\frac{P_{f,it}}{P_{g,it}}\right)\ln K_{it} + \beta_T T_t + \frac{1}{2}\beta_{TT}T_t^2 + \beta_{TY}T_t\ln Y_{it} + \sum_{f=1}^{2}\beta_f \ln\left(\frac{P_{f,it}}{P_{g,it}}\right)T_t + v_{it} + u_{it}$$

where the subscripts *i* and *t* denote firm- and time-specificity, VC_{it} denotes observed variable fuel costs, Y_{it} is total output measured as 1000s of MWhs of electricity, K_{it} denotes generating capacity measured in kW, $P_{f,it}$ is the total price (in cents per mmBtu) of each fuel, *f*, and *T* is a trend variable meant to account for technical change.[32] v_{it} is the standard stochastic error component, and u_{it} is the inefficiency component of the error, accounting for firm *i*'s cost inefficiency in period *t* . u_{it} is discussed in greater detail below. The standard restrictions for linear homogeneity[33] in input prices are imposed by normalizing VC_{it}, $P_{c,it}$, and $P_{o,it}$ by $P_{g,it}$, where *c, o* and *g* denote coal, oil, and gas, respectively. Symmetry is also imposed such that $\beta_{jh} = \beta_{hj}$, where *h* and *j* denote fuel inputs.

5.2 Cost Inefficiency Specification

Because our objective is the identification and analysis of the cost inefficiency of the electricity generating firms in our sample under the tradable SO_2 permit scheme, we opt for a model that allows for the parameterization and explanation of inefficiency. This type of model finds its roots in the time-varying inefficiency model of Kumbhakar (1990), while the methodology was expanded by Kumbhakar, Ghosh and McGuckin (1991) and Huang and Liu (1994) to allow the mean of the inefficiency error component's distribution to depend on a set of explanatory variables. Battese and Coelli's (1995) panel-data generalized Huang and Liu's (1992) technical inefficiency model panel data estimation. It is this inefficiency-mean parameterizing method (a method given the author-based acronym KGMHCBC by Wang (2002)) that we use here.

Variables included in the inefficiency equation (Eq. 20) should be those that might explain the firm's deviation from the minimum cost frontier. In our case, these variables include both policy and firm-specific characteristics. The policy variables we include are 1) a dummy variable denoting a firm's inclusion in the SO_2 scheme for a given period, and 2) the natural log of the SO_2 permit price. We use the former simply to delineate the effects of scheme inclusion, and the latter as a proxy for policy stringency, assuming that greater stringency implies fewer permits, and therefore a higher permit price. The firm-specific characteristics we incorporate include the percentage of coal capacity covered by scrubbers, the percentage of total capacity that is coal-fired, a dummy variable for "small" firms (those having less than 1,000,000 kW total capacity), a dummy variable for "large" firms (those with greater than

---

[32] Note that the interaction term between time and capital have been left out because of our short run focus on disembodied technical change.

[33] $\sum_{f=1}^{3}\beta_f = 1;\ \sum_{f=1}^{3}\beta_{fj} = 0\ \ \forall j;\sum_{f=1}^{3}\beta_{Yf} = 0;\ \ \sum_{f=1}^{3}\beta_{Kf} = 0;\ \ \sum_{f=1}^{3}\beta_{Tf} = 0$. See Chambers (1988) and Greene (1997).

3,000,000 kW total capacity),[34] and capacity factor, a measure of how capacity-constrained a firm is in each period. Finally, we include interaction terms for some of the variables to account for interdependence, e.g. we interact ln(Permit Price) with the Inclusion Dummy, assuming that regulated firms will be more affected by the permit price than unregulated firms. We assign the first element of the z_{it} vector the value of one in order to include an intercept parameter.[35] Thus, Eq. 20 is specified here as:

$$\begin{aligned} u_{it} = {} & \delta_0 + \delta_1 \ln PermitP_{it} + \delta_2 \%CoalScrubbed_{it} + \delta_3 \%CapacityCoal \\ & + \delta_4 SmallDummy_{it} + \delta_5 LgDummy_{it} + \delta_6 InclusionDummy_{it} \\ & + \delta_7 \%Coal_{it} * \ln PermitP_{it} + \delta_8 \%CoalScrubbed * \ln PermitP_{it} \\ & + \delta_9 \ln PermitP * InclusionDummy + \delta_{10} CapacityFactor_{it} + w_{it} \end{aligned} \qquad \textbf{Eq. 23}$$

where w_{it} is the truncated random variable as explained for Eq. 20. The dataset is discussed in Section 6.

Because permit prices and scheme inclusion are incorporated into the total price of each fuel (see Section 6), these variables appear in both equations 22 and 23, meaning that their effects are felt in both the frontier and in an observation's distance from it. As long as the inefficiency effects are stochastic, the inefficiency model (Eq. 23) may contain variables present in the frontier itself (see Battese and Coelli (1993, 1995) and Piacenza (2006)). If the variance of the inefficiency parameter, σ_u^2, is zero, however, the model collapses to a standard cost function model with the variables in Eq. 23 not already appearing in Eq. 22 being included in the cost function. In that case, δ_0 would equal zero, as Eq. 22 already includes an intercept, and the standard errors for the estimates of δ_1, δ_6, and δ_9 would be inflated by multicollinearity issues, but the estimates themselves would remain unbiased (see, e.g., Greene (2003)). When $\sigma_u^2 > 0$, however, δ_0, δ_1, δ_6, and δ_9 can be identified separately in Eq. 23.

5.3 Estimation and Software

As noted, Equations 22 and 23 are estimated simultaneously through the method of Maximum Likelihood. The likelihood function (see Appendix 1 and Coelli (1996)) uses Battese and Corra's (1977) parameterization whereby $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and the proportion of the composed error term's variance attributable to the inefficiency component is written $\gamma = \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$ such that $0 \le \gamma \le 1$. As $\gamma \to 0$, the random noise component, v_{it}, dominates the composed error, while as $\gamma \to 1$, the inefficiency component, u_{it}, dominates. The parameters to be estimated, then, are β, δ, σ^2, and γ. These parameters are then employed in the estimation of firm- and time- specific cost inefficiencies as described by Eq. 18.

[34] Note that there are three firm-size categories: small, medium and large. The medium dummy variable is excluded to avoid multicollinearity issues, meaning medium firms are the benchmark against which the coefficients for 'large' and 'small' firms must be compared (see Greene (1997)). The 'small' category contains 2,447 observations, 'medium' contains 1,227 observations, and 'large' contains 1,244 observations. See Table A4.2 in Appendix 4.

[35] Because of the inclusion of dummy variables in the inefficiency model, not including an intercept parameter results in biased estimates of the δ parameters, and an unnecessarily-restricted distribution of u_{it}, just as in the standard OLS model. See Piacenza (2006).

The Frontier 4.1 software (Coelli (1996)) performs the maximum likelihood estimation in three steps: first, OLS estimates of the β parameters from the cost frontier are obtained[36]. Second, a two-phase grid search is performed to find γ with the β parameters (except the intercept, β_0) set to the OLS estimates, and β_0 and σ^2 adjusted by the corrected OLS method outlined in Coelli (1995); δ and μ are set to zero in this stage. In the third stage, the parameter values from the grid search are used as starting values for the David-Fletcher-Powell Quasi-Newton method which seeks a local minimum for the negative log likelihood function by iteratively updating the parameter vector until convergence is achieved.

6 Data

The dataset is an unbalanced panel consisting of 5128 monthly firm-level[37] observations for 37 US electric utilities over the period 1990-2004[38], representing approximately 7% of total annual generation from US electric utilities[39], with electric utilities accounting for a declining proportion[40] of total annual US generation: 89% in 1995 to 63% in 2004. Each firm has the ability to generate with coal, oil *and* natural gas, but lacks hydro and nuclear capacity[41], and each has a minimum of 84 time series observations[42]. The dataset is therefore an unbalanced panel with 1532 observations missing for reasons associated with data collection processes (e.g. sporadic data collection or clerical errors) and not market entry or exit. All nominal price data have been converted to real December 2004 values.

6.1 Sources[43]

We took monthly firm-level data on electricity generation, prime mover and fuel type from the form EIA-906 (formerly Form EIA-759) (Form EIA-759/906)[44]. Fuel-specific data (i.e. fuel type, quantity, sulfur and thermal contents, and delivered costs) came from the FERC Form 423, 'Monthly Report of Cost and Quality of Fuels for Electric Plants' (EIA, FERC Form 423). Form 423's plant-level data is reported monthly by electric utilities for each plant having at least 50MW steam or combined-cycle capacity.

Capacity data came from the Form EIA-860 (and EIA-860A), 'Annual Electric Generator Report' (Form EIA-860). Form EIA-860 is filed annually for all existing and proposed (to be completed within 5 years) plants of 1 or more MW and contains data for both utilities and

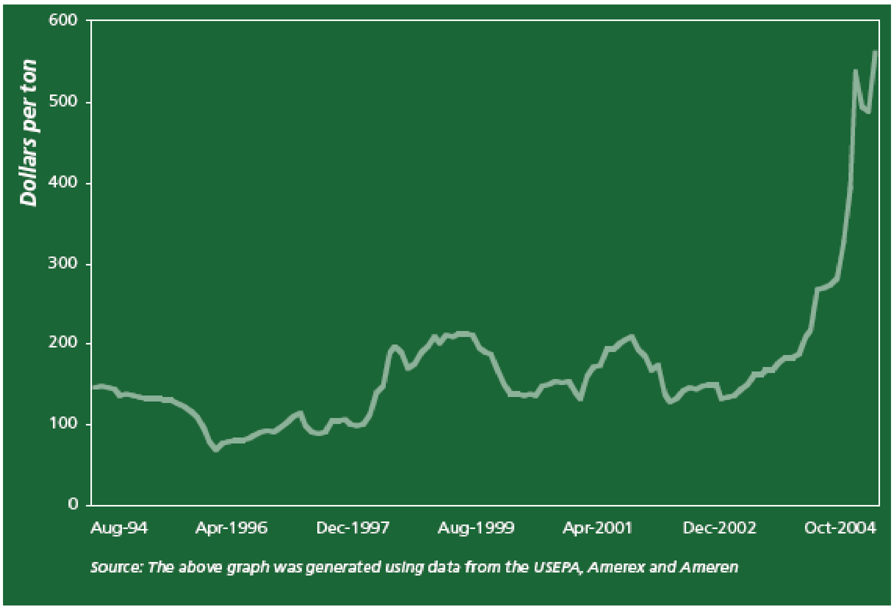
[36] All OLS β estimates will be unbiased except for the intercept.
[37] As previously noted, we chose to aggregate to the firm-level (rather than the somewhat more common plant level of aggregation) because it is at the firm level where operating decisions affecting costs are made.
[38] Because the Frontier 4.1 software can not cope with missing values, all observations with a missing value for any variable are omitted.
[39] Hydro and nuclear alone account for approximately 30% of total annual utility generation over the period.
[40] This serves as defense for our cost-minimizing assumption.
[41] This is to ensure 1) a degree of homogeneity of production technology and 2) the ability to switch between fuels in response to the scheme.
[42] We eliminate firms with less than 84 observations because of the testing discussed later related to changes over the 5-year time subsets of the data. Each time subset represents 60 (monthly) time periods and in order to ensure that a "reasonable" amount of data was present across at least two time periods, we require at least 84 observations per firm.
[43] All raw data was imported, sorted, organized and aggregated with a program written in SAS 9.1 for Windows.
[44] Form EIA-759, 'Monthly Power Plant Report' was discontinued in 2000 and replaced by the EIA-906, 'Power Plant Report'.

non-utilities. Prior to 2001, utility and nonutility data was filed separately with forms EIA-860A and EIA-860B, respectively.

We took monthly plant-level SO_2 emissions data from the EPA's "Clean Air Markets- Data and Maps" website. (See 'EPA Clean Air Markets Data and Maps'.) When necessary, missing SO_2 emissions were calculated as described below using FERC 423 fuel data and emissions factors published in the EPA's AP-42 (EPA, 2000). We also used the EPA's 'Clean Air Markets- Data and Maps' to find unit-level data regarding scrubber installation and inclusion in Phase 1 of the Acid Rain Program.

Cantor Fitzgerald is the main brokerage for sulfur permits, but their data is available only to paid registered users.[45] Thus, our historical allowance data comes from the EPA ("Clean Air Markets") for the years 1994-2000 and from Evolution Markets, Inc.[46] for 2002-2004. For the intervening period, we estimated monthly permit prices from a chart by the Chicago Climate Exchange based on data from the EPA, Amerex and Ameren (Chicago Climate Futures Exchange, 2004), and presented as Figure 1 below. Note the rise in allowance prices in 2004 to the levels predicted prior to the scheme's commencement. This provides evidence of "low hanging fruit" (mostly in the form of high-to-low sulfur coal switching made feasible by rail deregulation in the early 1990s) in the industry prior to the cap-and-trade scheme that was not exhausted until halfway through Phase II. This presumably had an effect on cost efficiency performance, and we investigate this further in the following section.

Figure 1: SO_2 Allowance Prices 1995-2004



6.2 *Variables and Problems*

Variables requiring explanation are discussed in this section.

[45] See http://www.emissionstrading.com/MarketData/
[46] See http://www.evomarkets.com/resources/index.php?xp1=1&type=mmu&msub=arc&mk=1

Fuel Price[47]

The main fossil fuel categories (coal, oil and gas) are aggregates of more specific sub-types, each of which have differing thermal and environmental characteristics. In our data, the coal subtypes are BIT (bituminous coal), LIG (lignite), ANT (anthracite) and SUB (subbituminous coal). The oil subtypes for FO2, FO4 and FO6 are generalized to subtype DFO (distillate fuel oil) following Forms EIA-906 and 860. Natural gas subtypes did not exist.

We calculated fuel prices in cents/mmBtu of heat input using the delivered cost data from the FERC Form 423. We calculated the monthly state average price for each fuel and used it as a proxy for the observed fuel price when a firm reported a zero purchase quantity for a particular fuel but had capacity to generate with that fuel. For example, if a firm was able to generate with coal but reported no coal purchases in a given month, the average coal price for the firm's state was substituted for the missing price value.

SO_2 Quantity

To find the SO_2 price for each fuel, we needed monthly emissions estimates. We calculated these values using fuel-specific conversion factors (EPA, AP-42) along with sulfur and quantity data from the FERC Form 423. For example, missing coal SO_2 quantities were calculated as follows for each subtype:

Quantity of SO_2 from Coal Combustion = (C*Sulfur*Quantity)/2000

where Sulfur and Quantity refer to the sulfur content and quantity of the fuel subtype, and 'C' represents the subtype-specific emissions factor.

SO_2 Prices

We note here the distinction between the SO_2 allowance price and the sulfur price associated with the combustion of a specific fuel. The allowance price represents the average monthly market price associated with the emission of one ton of SO_2, whereas the fuel-specific emissions price depends on the firm's technology and the specific fuel combusted as well as the allowance price and is measured in cents per mmBtu. We calculated this latter value using the actual permit price, fuel-specific emissions values as above, and fuel-specific BTU and quantity data from the FERC Form 423. We aggregated this data to the firm level, and missing values were replaced with a fuel-specific state average sulfur price. Thus, for example:

SO_2 Price of Coal= (**100***(Permit Price/PPI)*Emissions from Coal)/Q of Coal Combusted,

and if this value was missing, it was assigned the average SO_2 price for coal from the state in which the firm operates.

Total Fuel Price

The fuel price that appears in the cost frontier/functions is the *total* fuel price, i.e. the sum of the SO_2 and Btu price components for each fuel aggregate. For oil and gas, this is just the

[47] NB: All prices (fuel, sulfur and permit prices) are real values measured in December 2004 US$.

sum of the delivered fuel price and fuel's SO_2 price. Some plants, however, have installed scrubbers as a means of CAAA compliance, reducing required permit quantities associated with coal combustion. We account for this in the total coal price calculation by including the percentage of coal capacity covered by scrubbers as follows:

Total Coal Price= BTU price of coal+ (1-% Capacity Scrubbed)* SO_2 Price of Coal.

Capacity

Several boilers listed on the Form EIA-860 are dual-fired units capable of burning more than one fuel. In these cases, we counted the unit's capacity in both fuels' categories, but count it only once in the firm's total capacity. For example, if a unit was capable of burning either DFO or BIT, the unit's capacity was counted both as oil and as coal capacity, but the firm-level total reflects only the unit's nameplate capacity.

Capacity Factor

The capacity factor is the ratio of the energy produced in a period to the amount that would have been possible if the facility had operated at *full capacity* for the entire period. Söderholm (1998) notes the multi-dimensional nature of the output of electricity generating firms (both power (in wattage) and energy (measured in watt-hours)), and promotes the inclusion of the load factor[48] as an independent variable to account for this. We use the capacity factor variable instead because of the role of capacity constraints in the power industry, and because the capacity factor is slightly more indicative of capacity flexibility than the load factor.

6.3 Descriptive Statistics

Descriptive statistics for the relevant final variables are presented in the tables below.

All firms, all years:

Table 1: Cost Frontier Variables: All Firms, All Years

Variable	Units	Mean	Standard Deviation	Min	Max
Total Coal Price	¢/mmBtu	181.70	44.22	61.19	385.15
Total Gas Price	¢/mmBtu	415	195.96	40.72	2527.93
Total Oil Price	¢/mmBtu	520.93	177.67	133.42	1343.60
Monthly Generation	1000's MWh	580.58	625.76	1.02	3426.04
Total Capacity	kW	1,768,266	1,669,586	74,599.97	6,506,999
Total Variable Costs	Millions Dec. 2004 $	11.8287	13.3545	0.00005	223.1613

48 The load factor is the ratio of average electricity produced in a period to the amount that would have been produced if a firm/plant/unit had operated at its peak output for the entire period. Thus, the capacity factor provides information about spare capacity while the load factor provides more information about electricity demand.

Table 2: Inefficiency Model Continuous Variables: All Firms, All Years					
Variable	**Units**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Permit Price	$/ton	131.36	118.72	0[49]	705.95
% Capacity Coal	Percent	39.81	15.79	0	1
% Coal with Scrubber	Percent	9.57	24.84	0	1

Table 3: Summary of Dummy Variables: All Firms, All Years			
	Value	**Frequency**	**Percent of Sample**
Inclusion Dummy	0	3,052	59.52
	1	2,076	40.48
Small Dummy	0	2,681	52.28
	1	2,447	47.72
Large Dummy	0	3,884	75.74
	1	1,244	24.26
Table A Dummy	0	3,535	68.94
	1	1,593	31.06

By Table A Status, All Years:

Table 4: Cost Frontier Variables: Table A and Non Table A Firms, All Years					
Variable	**Units**	**Mean**	**Standard Deviation**	**Min**	**Max**
Table A Firms*					
Total Coal Price	¢/mmBtu	167.64	44.06	77.33	321.21
Total Gas Price	¢/mmBtu	413.74	202.58	140.01	2456.96
Total Oil Price	¢/mmBtu	530.66	165.22	174.99	1151.58
Monthly Generation	1000's MWh	860.90	700.42	21.11	3426.04
Total Capacity	kW	2,649,384	1,886,784	636,889.7	65,056,999
Total Variable Costs	Millions Dec. 2004 $	15.0660	11.9995	0.0023	81.8732
Non Table A Firms*					
Total Coal Price	¢/mmBtu	188.04	42.81	61.19	385.15
Total Gas Price	¢/mmBtu	415.79	192.94	40.72	2527.93
Total Oil Price	¢/mmBtu	516.55	182.85	133.42	1343.60
Monthly Generation	1000's MWh	454.26	543.78	1.02	2534.87
Total Capacity	kW	1,371,201	1,390,206	74,599.97	4,786,041
Total Variable Costs	Millions Dec. 2004 $	10.3699	13.6846	0.0005	223.1613
* NB: 11 of the 37 firms had units marked with Table A status. There are 3535 observations for Non-Table A firms and 1593 for Table A firms.					

[49] Note that zero values have been converted to $1.0e^{-10}$ because of logarithmic nature of the variable's appearance in the inefficiency model.

Table 5: Inefficiency Model Continuous Variables: Table A and Non Table A Firms, All Years					
Variable	**Units**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Table A Firms					
% Capacity Coal	Percent	0.4092	0.1073	0.0013	1
% Coal with Scrubber	Percent	0.0702	0.1331	0	0.4463
Non Table A Firms					
% Capacity Coal	Percent	0.3931	0.1758	0	1
% Coal with Scrubber	Percent	0.1072	0.2848	0	1

Table 6: Summary of Dummy Variables: Table A and Non Table A Firms, All Years			
	Value	**Frequency**	**Percent of Sample**
Table A Firms			
Inclusion Dummy	0	518	32.52
	1	1,075	67.48
Small Dummy	0	1,159	72.76
	1	434	27.24
Large Dummy	0	970	60.89
	1	623	39.11
Non Table A Firms			
Inclusion Dummy	0	2,534	71.68
	1	1,001	28.32
Small Dummy	0	1,522	43.06
	1	2,013	56.94
Large Dummy	0	2,914	82.43
	1	621	17.57

7 Results

The maximum likelihood estimates of the cost frontier parameters, β, and inefficiency model parameters, δ, are presented in Tables 7 and 8 below. Because the variable costs and input prices were normalized by the price of natural gas to ensure the frontier's linear homogeneity in input prices, $ln\ P_g$ and its interactions did not appear explicitly in the cost frontier equation. As such, the parameters in Table 7 associated with P_g have been calculated from the homogeneity restrictions and are, therefore, presented without standard errors.

Our focus in this paper is on the cost efficiency effects of the SO_2 trading scheme and firm-specific characteristics, and as such we do not discuss the β parameters explicitly. We note only that 1) most estimated parameters are significant at the 5% level, 2) the estimates of β_T, β_{TT} and β_{TY} indicate cost diminution across the period, and, as such, regressive technical change and 3) the signs of all parameters are as expected except for except for those of β_{OO}

and β_K, which indicate a positive own-price elasticity for oil[50] in the former case and that variable fuel costs per kWh increased with capacity,[51] the proxy for the quasi-fixed capital variable, in the latter. The δ estimates in Table 8 are discussed in subsections 7.2 and 7.4.

7.1 Model Specification

We begin with a series of likelihood ratio (LR) tests to assess various assumptions related to the model's specification. All of these tests take the form[52]

$$\lambda = 2*(\log L_u - \log L_r) \qquad \textbf{Eq. 22}$$

where $log\ L_u$ is the value of the log likelihood for the unrestricted model, and $log\ L_r$ is the value of the log likelihood for the restricted model (i.e. the model incorporating the parameter restrictions imposed by the null hypothesis, H_0). The LR test value, λ, has a χ^2 distribution with degrees of freedom equal to the number of restrictions imposed by H_0. Test results are presented in Table 9.

We first check that the translog functional form provides a more reasonable representation of the true variable cost structure than the more restrictive Cobb-Douglas (C-D) functional form. We do so by testing the parametric restrictions required to obtain the Cobb-Douglas form from the translog such that $logL_r$ corresponded to the C-D specification with H_0[53]: $\beta_{YY}= \beta_{KK}= \beta_{KY}= \beta_{ff}= \beta_{fY}= \beta_{fK}= \beta_{TT}= \beta_{Tf}= 0$. We then perform a second technology check by testing the restrictions required for a homothetic function, H_0: $\beta_{fY}= \beta_{fK}= 0$. The results indicate that the translog specification is, indeed, preferential to the more restrictive C-D technology and that the assumption of a homothetic technology implying a linear expansion path (i.e. constant input ratios) and independence of the elasticity of scale from input quantities (see Chambers (1988)) can be rejected.

[50] This suggests a violation of the concavity requirement for the cost frontier. Such a violation is not uncommon in the literature, particularly with the translog functional form (see, e.g. Arnberg and Bjørner (2007), and may be due to imperfections in the oil market, or oil's small share of total fuel costs (see Tuthill (2008)). See also Diewert and Wales (1987).

[51] A variable cost function should be non-increasing in capital stock (see Cornes (1992, p. 106)). This positive K parameter has been found in several other variable cost studies involving both the electricity industry and other industries, and employing both standard cost function and cost frontier analysis. See, e.g., Piacenza's (2006) study of Italian public transport, which cites also Fillipini (1996) and Caves et al. (1985). The former uses multicollinearity caused by a correlation between the capital measure and the dependent variable to explain the positive K parameter, and the latter argues that it is an indication that the industry in question over uses capital and does not minimize costs in the long run.

[52] The restricted and unrestricted log Likelihood values have the place they do in Equation 1 because Frontier 4.1 minimizes the negative of the log likelihood function (rather than equivalently maximizing its positive value).

[53] In this null hypothesis and the following, *Y* represents output, *K* capital, *T* time, and *f* fuel-type.

Table 7: Cost Frontier Parameter Estimates			
Parameter	**Frontier Variable**	**ML Estimate**	**Standard Error**
β_0	Constant	-59.2260**	1.9261
β_c	$\ln P_c$	5.8174**	0.6977
β_o	$\ln P_o$	2.8914**	0.7544
β_g	$\ln P_g$	-7.7088	n/a
β_K	$\ln K$	6.8576**	0.3698
β_{KK}	$\ln K^2$	-0.4934**	0.0362
β_Y	$\ln Y$	0.6371*	0. 2826
β_{YY}	$\ln Y^2$	-0.0977**	0.0258
β_{KY}	$\ln K \ln Y$	0.0729**	0.0271
β_{cc}	$\ln P_c \ln P_c$	-0.6166**	0. 1182
β_{oo}	$\ln P_o \ln P_o$	0.2019*	0.1302
β_{gg}	$\ln P_g \ln P_g$	-0.3479	n/a
β_{co}	$\ln P_c \ln P_o$	0.0334	0.0926
β_{cg}	$\ln P_c \ln P_g$	0.5832	n/a
β_{og}	$\ln P_o \ln P_g$	-0.2353	n/a
β_{cY}	$\ln P_c \ln Y$	-0.1465**	0.0526
β_{oY}	$\ln P_o \ln Y$	0.2285**	0.0545
β_{gY}	$\ln P_g \ln Y$	-0.0280	n/a
β_{cK}	$\ln P_c \ln K$	-0.2668**	0.0693
β_{oK}	$\ln P_o \ln K$	-0.3099**	0.0732
β_{gK}	$\ln P_g \ln K$	0.5767	n/a
β_T	Time	-0.0169**	0.0014
β_{TT}	Time^2	0.00002**	0.00001
β_{TY}	Time $\ln Y$	-0.0031**	0.0002
β_{cT}	Time $\ln P_c$	-0.0037**	0.0008
β_{oT}	Time $\ln P_o$	0.0005	0.0007
β_{gT}	Time $\ln P_g$	0.0032	n/a
*** Denotes significance at the 5% level. * Denotes significant at the 10% level.* *__indicates value calculated from homogeneity restrictions.*			

Table 8: Cost Inefficiency Model Parameter Estimates			
Parameter	**Variable**	**ML Estimate**	**Std Error**
δ_0	Constant	-8.0135**	0.3490
δ_1	ln(Permit Price)	-0.0411**	0.0047
δ_2	Percent of Coal Capacity with Scrubber	1.2490**	0.0899
δ_3	Coal's Percentage of Total Capacity	-2.9179**	0.3558
δ_4	Small Dummy	10.7267**	0.2599
δ_5	Large Dummy	1.2084**	0.1726
δ_6	Scheme Inclusion Dummy	0.0276	0.8228
δ_7	% Coal * ln(Permit Price)	0.0909**	0.0120
δ_8	% Coal Scrubbed * ln(Permit Price)	0.0819**	0.0069
δ_9	ln(Permit Price)*Inclusion Dummy	-0.1088	0.1573
δ_{10}	Capacity Factor	-0.0328**	0.0042
** Indicates significance at the 5% level. * Indicates significance at the 10% level.			

The last three of the LR tests presented in Table 9 relate to the inefficiency specification. As noted, we use the parameterization proposed by Battese and Corra (1977) where

$\sigma^2 \equiv \sigma_v^2 + \sigma_u^2$ and $\gamma \equiv \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$. The parameter γ (0≤γ≤1) therefore provides a measure of the relative contributions of each of the error components to the overall error term, ε_{it}. As $\gamma \to 0$, the noise component, v_{it}, dominates the compound error term, while as $\gamma \to 1$, the one-sided inefficiency term dominates. In the former case, the inefficiency component is not stochastic and standard cost function analysis with only random noise in the error term is sufficient. In the latter case, a deterministic cost frontier with no noise component is most appropriate.[54]

The third LR test in Table 9 thus tests H_0: γ=0 (or, equivalently $\sigma_u^2 = 0$) against the alternative that the inefficiency component is, in fact, stochastic. This is an important test for two reasons: 1) it serves to justify the SFA method over a standard cost function, and 2) it tests the ability of the model to separately identify technical change in the cost frontier and changes in inefficiency over time. In the absence of stochastic inefficiency (i.e. when $\gamma = \sigma_u^2 = 0$), parameters δ_0 and δ_1 can not be identified, as noted in the previous section. Note that the LR test statistic for H_0: γ=0 will not have a χ^2 distribution, but rather the mixed χ^2 distribution[55], as this restriction lies on the boundary of the parameter space for γ. The critical values for the LR test of γ are thus obtained from Table 1 in Kodde and Palm (1986). The null hypothesis that the inefficiency effects are not stochastic is strongly rejected, providing justification for the SFA approach and allowing us to delineate between time's effects on the cost frontier through disembodied technical change and its effect on estimates of cost inefficiency.

The final two LR tests are specified with null hypotheses claiming 1) the absence of inefficiency effects in the data and 2) that the inefficiency effects are not a linear function of the explanatory variables, respectively[56]. The results indicate that both nulls may be rejected, implying that the model does contain inefficiency effects, and that the joint effects of our explanatory variables on cost inefficiency are significant.

Table 9: Model Specification Likelihood Ratio Tests

Test	λ	Critical value[57]	Decision
H_0 (C-D): $\beta_{YY} = \beta_{KK} = \beta_{KY} = \beta_{ff} = \beta_{fY} = \beta_{fK} = \beta_{TT} = \beta_{TY} = \beta_{Tf} = 0$	1583.3568	23.685	Reject null
H_0 (homothetic): $\beta_{fY} = \beta_{fK} = 0$	461.6796	9.488	Reject null
H_0: $\gamma = \delta_0 = 0$	1158.6665	5.138#	Reject null
H_0: $\gamma = \delta_0 = \delta_1 = \ldots = \delta_{10} = 0$	1160.4594	19.045#	Reject null
H_0: $\delta_1 = \delta_2 = \ldots = \delta_{10} = 0$	1159.5094	18.307	Reject null

 # LR critical value with critical mixed χ^2 value from Table 1 in Kodde and Palm (1986)

[54] See Kumbhakar and Lovell (2000) for a brief discussion of deterministic cost frontiers and Piacenza (2006) and Coelli (1996) for more on the Battese and Corra (1977) parameterization.
[55] See Coelli (1993, 1994 and 1996) and Lee (1993).
[56] See Battese and Coelli (1993, 1995) for a further discussion.
[57] All critical values are reported at the five percent level.

Having determined that the model specified by Equations 22 and 23 in Section 5 is reasonable, Table 10 presents the estimated log likelihood value, the number of observations in the dataset, *n*, and estimates of both σ^2 and γ. Of particular interest is the estimate for γ, indicating that approximately 90 percent of the composed error term's variance is attributable to the cost inefficiency component, implying, as noted, that a straightforward cost frontier may not be appropriate. Piacenza (2006) and Coelli et al. (1998) note that relatively high estimates for γ are common in these types of models[58]. This could be due to the fact that our sample includes public utilities operating in a less than perfectly competitive market or to the fact that capital in this industry is, in many cases, fuel-specific, leaving firms less able to freely adjust their non-labour inputs than in other less capital-intensive industries. The fact that our fuel price and quantity data is for fuel purchased (as opposed to combusted) in each month may also increase the contribution of inefficiency to the composed error term. Piacenza (2006) notes a third possible explanation which may also pertain to the present model: the absence from the frontier of firm heterogeneity (e.g. via fixed effects) may cause a larger variance for the composed error term, and particularly for the inefficiency component. Thus, while caution should be used in interpreting the estimate of the γ parameter, our estimate is consistent with those found in the other applications of the Battese and Coelli (1995)-type models.

Table 10 : Model Characteristics and Error Component Variance Estimates

Parameter	Explanation	Estimate	Std Error
n	number of observations	5128	
logL	value of logL function	-5700.3921	
σ^2	$\sigma^2 = \sigma_v^2 + \sigma_u^2$	1.9988**	0.0439
γ	$\gamma = \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$	0.9020**	0.0052

** Indicates significance at the 5% level.*
 # LR test value with critical mixed χ^2 value in parentheses.

7.2 *Firm and Policy Effects on Cost Inefficiency*

The discussion in this subsection focuses on the δ estimates from the cost inefficiency model presented in Table 8. Having tested above for the joint significance of the *z* variables, we begin here by testing the individual significance of the impact of each *z* variable on cost inefficiency estimates. The LR test results presented in Table 11 below indicate that the null hypothesis of no effect can be rejected in all cases. Note that the χ^2 distribution of the LR test has one degree of freedom for all tests besides the size dummy test. The size dummy test has two degrees of freedom because both the "small" and "large" dummy must be excluded to test for the lack of a size impact on cost inefficiency, implying the imposition of two restrictions upon the full model (i.e. $\delta_4 = \delta_5 = 0$).

[58] Hiebert (2002), for example, found γ=0.9729 and γ= 0.9873 for coal and gas plants, respectively, in his study of plant-level technical efficiency in the US electricity generating industry.

Given both the joint and individual significance of the explanatory z variables, the ML estimates of the δ parameters in Table 8 above provide an indication of the direction of the effect of the z variables on cost inefficiency. Because the inefficiency effects are estimated as the conditional expectation of the inefficiency error component, u_{it}, given the observed value of the composed error, ε_{it},(i.e. because $\hat{CE}_{it} = E(\exp\{u_{it} \mid \varepsilon_{it}\})$), the estimated δ parameters are not directly interpretable as marginal effects on inefficiency estimates. We briefly discuss the information available from the sign of the estimates provided in Table 8 and consider the marginal effects in more detail below.

Table 11: Significance Tests for Inefficiency Parameters

Null Hypothesis	Variable	log Likelihood	λ	Decision
H_0: δ_1=0	ln(Permit Price)	-5628.8048	69.8418	Reject H_0
H_0: δ_2=0	Percent of Coal Capacity with Scrubber	-5640.0910	92.4142	Reject H_0
H_0: δ_3=0	Coal's Percentage of Total Capacity	-5612.3226	36.8774	Reject H_0
H_0: δ_4= δ_5=0	Small & Large Dummy	-6047.2337	906.6996	Reject H_0
H_0: δ_6=0	Scheme Inclusion Dummy	-5599.0533	10.3388	Reject H_0
H_0: δ_7=0	% Coal * ln(Permit Price)	-5652.7300	117.6922	Reject H_0
H_0: δ_8=0	% Coal Scrubbed * ln(Permit Price)	-5619.5352	51.3026	Reject H_0
H_0: δ_9=0	ln(Permit Price)*Inclusion Dummy	-5608.1746	28.5814	Reject H_0
H_0: δ_{10}=0	Capacity Factor	-6129.7524	1071.737	Reject H_0

As expected, the parameter for the percentage of coal capacity covered by scrubbers, δ_2, is positive (δ_2 = 1.2490), indicating increased cost inefficiency for increasing levels of scrubber coverage. This is likely due to the lower technical inefficiency of plants using scrubbers, as they require as much as three percent of the plant's electricity output for operation.[59] The other firm-specific continuous-variable (dummies are discussed in greater detail below), coal's percentage of total capacity, shares a positive relationship with cost efficiency, as indicated by its negative coefficient (δ_3 = -2.9179). This support's Hiebert's (2002) findings of a "fleet effect", evidence that utilities' mean technical efficiency increases with ownership of additional plants of the same fuel. We note also that many state and Federal environmental regulations prior to the Clean Air Act Amendments of 1990 pertained mostly to coal-fired generating units (see, e.g. EIA (1996) and Edison Electric Institute (EEI) (2001)). Thus, increased cost efficiency as the percentage of a firm's coal-fired capacity rises may also reflect efficiencies gained through experience with past environmental regulation.

The parameters for the two policy-related variables (scheme inclusion and permit price) have opposite signs, indicating opposing effects on cost inefficiency. The coefficient for the scheme inclusion dummy (δ_7 = 0.0276) is positive, but not significant, while the negative

[59] See, e.g., Bellas and Lange (2008). See also Yaisawarng and Klein (1994) or Markiewicz, Rose and Wolfram (2004); the former finds increased technical inefficiency for electricity generating plants operating with a scrubber, and the latter finds a positive relationship between a scrubber dummy (=1 in the presence of a scrubber) and the plant-specific usage of labour, non-fuel and fuel inputs.

coefficient on the natural log of the permit price (δ_1 = -0.0411) indicates a negative relationship between cost inefficiency and permit prices. The magnitude of the coefficient on the interaction term between scheme inclusion and ln(Permit Price) (δ_9 = -0.1088) relative to that of the δ_1 alone indicates that the efficiency-improving effects of higher permit prices were stronger for observations included in the scheme.[60] The parameter estimates associated with scheme inclusion and the interaction of scheme inclusion and permit price were insignificant, but their exclusion was rejected, as seen in Table 11. Thus, while the parameter estimates associated with scheme inclusion do not appear to be significant, the results in Table 11 indicate that there effects were nonnegligible.

The signs of these parameters are likely explained by the constraints that capital stock places on the firm's ability to minimize costs.[61] The sign of the scheme inclusion parameters indicates that the capital stock (and range of available fuels) chosen optimally prior to the onset of the SO_2 scheme was no longer optimal in the face of SO_2 allowance trading regardless of the SO_2 price. The fact that this fixed capital effect is not greater and that the inclusion parameter is not significant is presumably due to the interfuel substitution that occurred over the period, even with fixed capital (see Tuthill (2008)). This argument is supported by the counterintuitive increase in inefficiency with time as discussed above. The improvement in cost inefficiency with higher SO_2 prices may also be explained by capital, though this time by scrubbers rather than combustion capacity. As previously noted, scrubbers are costly in terms of output consumption. As permit prices rise and the minimum cost frontier shifts, however, scrubbed capacity becomes more cost effective, allowing for generation at costs closer to the frontier. Thus, it appears that the imposition of the SO_2 scheme may have caused an insignificant increase in cost efficiency, but that given regulation, minimum cost production was more readily attained at higher allowance prices.[62] We assess the SO_2 scheme's impact on cost inefficiency further in the following subsection.

Table 12: Interpretation of Dummy Variable Effects

	Small Firms	Medium Firms	Large Firms
Included in Scheme	2.7399	-7.9859	-6.7775
Not Included in Scheme	2.7132	-8.0135*	-6.8051

* Indicates reference case.
NB: Values are calculated from the constant and dummy parameters reported in Table 8 above.

Parameters for the firm size dummy variables are not directly interpretable as there are three firm-size categories as well as a scheme inclusion dummy. We excluded the "medium firms" dummy from our model to avoid issues of multicollinearity, leaving excluded mediums firm as the reference case. Interpreted dummy variable effects are presented in Table 12 above.

[60] Note that the efficiency-improving quality of higher permit prices holds for both included and excluded firms. This could be due to excluded firms' expectations about future inclusion. Note also that the efficiency-reducing effects of scheme inclusion hold regardless of permit price. This, again, is presumably due to expectations regarding future policy and to capital constraint issues noted below.
[61] Many thanks to Malcolm Keay for a useful discussion on this point.
[62] This latter point is consistent with the claims of the Porter Hypothesis (see Porter (1991) and Porter and van der Linde (1995)), which notes the potential for private benefits and increased firm competitiveness as regulated firms respond dynamically (through innovation and improved organization) to environmental regulation. While efficiency gains from increased environmental stringency appear to have existed in the case of SO_2 regulation, it is not obvious this would be the case for other pollutants (e.g. CO_2), as the result appears tied to SO_2-specific capital (i.e. scrubbers).

The results indicate that firm size is, indeed, related to cost efficiency in the manner expected,[63] i.e. that small firms are less cost efficient than medium firms, while large firms are more cost efficient than medium ones. For all three firm sizes, moving from exclusion to inclusion in the SO_2 scheme (inclusion dummy = 0 to inclusion dummy = 1) is associated with reduced cost efficiency as suggested above. In the case of small firms, inclusion increases cost inefficiency estimates outright, while for medium and larger firms, scheme inclusion is associated with a smaller reduction in cost inefficiency.

7.3 Inefficiency Estimates

Mean firm cost inefficiencies over the entire period are presented in Figure 2 with the associated firm number displayed on the horizontal axis. The inefficiency estimates for the firm on the far right (Firm 3, City of Cedar Falls, EIA Company Code 3203) are eliminated from the descriptions and analyses of the inefficiency estimates in this subsection because of their distorting effects[64,65]. Mean annual inefficiency estimates for all firms as well as for Table A and non-Table A subsets are presented in Figure 2 below. Because the scale of inefficiencies is so apparently different for Table A and non Table A firms, relative annual changes in inefficiency for Table A firms are not easily visible in Figure 3, and we therefore present them on their own in Figure 4.

Note that the measure reported by Frontier 4.1 and those reported in the Figures and Tables throughout this section are, unless otherwise noted, a measure of cost *inefficiency* and range from one to infinity with one indicating full cost efficiency (i.e. operation on the frontier). The inverse (ranging from zero to one) of these inefficiency measures, then, represents a measure of cost *efficiency*, which can be used to calculate cost distortions due to cost inefficiency.

It is apparent from Figure 3 cost inefficiency for Non-Table A firms generally declined over the period,[66] and Figures 3 and 4 together illustrate a rise in inefficiencies in the year 1993 for both Table A and non-Table A firms, followed by a period of inefficiency recovery, and a spike again in 2000 and 2001. The 1993 spike coincides with the first EPA auction for SO_2 allowances, providing firms their first hint of what actual market emissions prices would be. It is plausible that uncertainty over market emissions prices, with pre-scheme marginal abatement cost estimates as high as \$1,500 per ton and actual 1993 EPA auction price of \$122 per ton led to inefficient behaviour, most probably due to allocative inefficiency as firms selected fuels based on SO_2 price expectations that did not match their realizations. In addition, 15 February 1993 was the date compliance plans for Phase I were due to the EPA, prior to the first allowance auction (held in March 1993), and before rules regarding allowance accounting and cost recovery had been announced (Burtraw, 1996). It is also possible that rail deregulation, which came to a conclusion in 1993 and brought a reduction in

[63] Hiebert (2002) shows that electricity generating plants' technical inefficiency declined with capacity. See also Atkinson and Halvosen (1984) for a discussion of the returns to scale in electricity generation.
[64] It may be that the assumption of cost minimization was inappropriate for this firm.
[65] Note that these distorting effects are related only to the descriptive statistics for mean inefficiencies in each period. All of the following analyses were performed both with and without Firm 3, though only those without Firm 3 are reported, as the significance and implications of the mean-equality and normality tests were not changed by the inclusion or exclusion of Firm 3's inefficiency estimates.
[66] This could be the result of increasing competitive pressures in the industry.

the delivered price of cleaner-burning western coal, played a role in the rise of cost inefficiency in 1993 as firms changed their coal-purchase and emissions reduction plans.

Figure 2: Mean Inefficiency Estimates by Firm

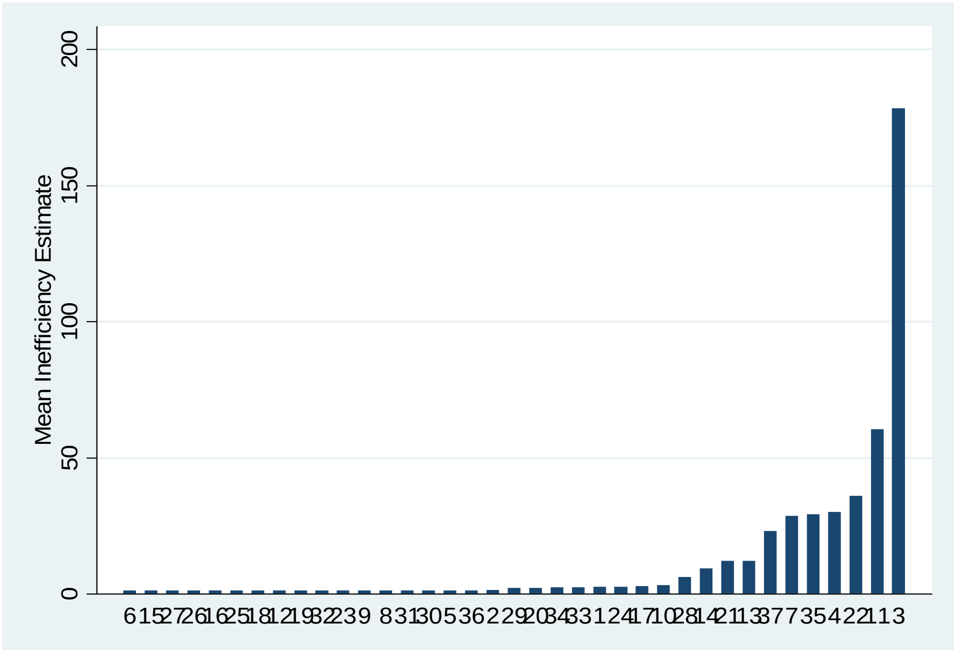
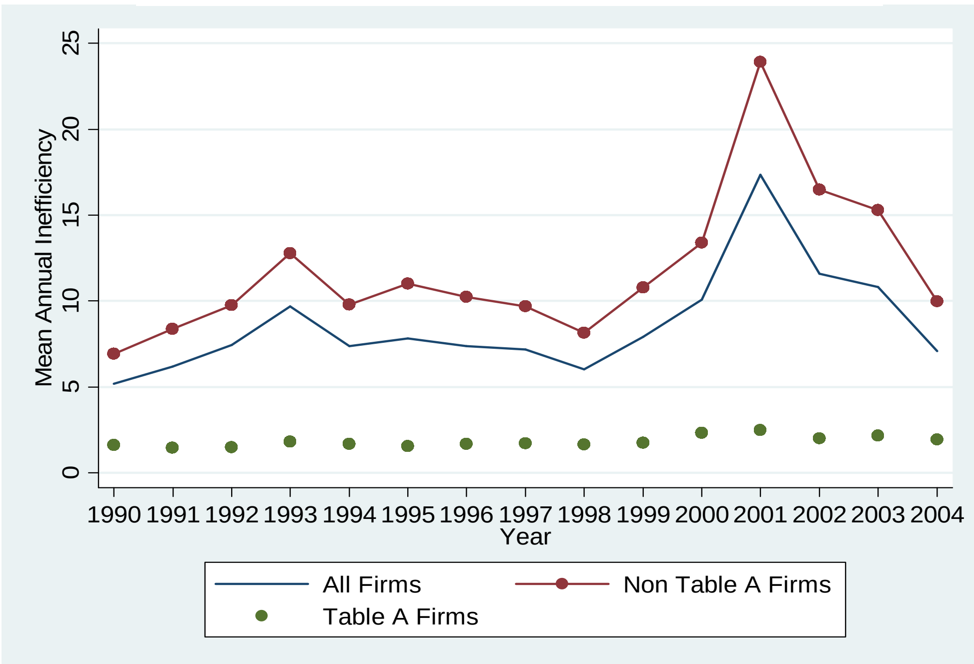


Figure 3: Mean Annual Inefficiencies by Firm



Cost inefficiency rose again for all firms in 2000 and 2001. This coincides with the onset of Phase II of the SO_2 scheme, when almost all fossil-fuel burning generating units came under regulation, as well as the dip in the capacity margin (unused capacity available for generation) reported in the year 2001 as the result of a reduction in capacity resources (EIA, 2007), which may have contributed to the inefficiency spike. There was also a (presumably endogenous) rise in natural gas prices in 2000 and 2001 (EIA (2001)), which, along with expectations about SO_2 prices as the emissions cap fell and more units came under regulation, also presumably contributed to the increase in cost inefficiency for both classes of firms. The California (2000) and New York (2001) electricity crises also occurred over this period, causing uncertainty over the future of the industry's deregulation, potentially contributing to the reduction in capacity and the rise in cost inefficiency in these years. Finally, the nation experienced a general decline in economic activity in 2001, with a business cycle peak identified in March and a trough in November of that year (NBER).

Figure 4: Mean Annual Cost Inefficiencies: Table A Firms

Thus, several factors converged to cause the inefficiency spike in 2000-2001, of which the SO_2 scheme was just one. We note that the mean annual inefficiency estimates then decline, even in the face of the higher allowance prices in 2004. This may reflect only that newly regulated firms took time to discover appropriate cost-minimizing techniques under the SO_2 scheme with fixed capital.

Descriptive statistics for the inefficiency estimates are presented in Table 13. The estimates and subsequent analyses are organized by firm-type (all firms, as well as Table A and Non-Table A subsets) and by time period (all years, pre-Phase I (1990-1994), Phase I (1995-1999) and Phase II (2000-2004)) in order to focus on SO_2 policy effects. The high mean cost inefficiency for all firms and for non-Table A firms during the 1990-1994 period are driven by the large 1993 estimates seen in Figure 3. The variance of the inefficiency estimates for Non-Table A firms indicates a wide ranging degree of minimum-cost attainment.

Table 13: Descriptive Statistics for Inefficiency Estimates by Firm Type and Period							
Time	**Firm Type**	**n**	**Mean**	**Median**	**Variance**	**Skewness**	**Kurtosis**
All Periods: 1990-2004	All Firms	5021	7.7255	1.3744	1236.744	32.7457	1633.244
	Non-Table A	3428	10.5571	1.6857	1785.653	27.4496	1140.362
	Table A	1593	1.6323	1.2639	1.5035	11.4021	216.2381
1990-1994	All Firms	1722	11.2154	1.3954	2984.231	25.5758	858.306
	Non-Table A	1204	15.3175	1.7118	4083.145	21.6462	611.4961
	Table A	518	1.6806	1.2742	3.0864	10.7893	152.8808
Phase I: 1995-1999	All Firms	1870	5.8526	1.3306	342.1623	12.3442	230.6874
	Non-Table A	1261	7.9365	1.5305	493.9969	10.2780	160.2967
	Table A	609	1.5375	1.2614	0.4098	3.2490	17.6180
Phase II: 2000-2004	All Firms	1429	5.9719	1.4323	389.3922	9.6358	125.3412
	Non-Table A	963	8.0367	1.8216	564.3577	7.9406	85.7961
	Table A	466	1.7024	1.2501	1.1609	4.2111	28.906

In order to further assess the impacts of the SO_2 scheme on firms' cost-minimizing ability, we test for inefficiency differences between periods prior to and periods post regulation for both Table A and non-Table A firms. The t-test for mean equality is the standard test used for this purpose, but is only applicable when the data being tested has been drawn from a population with a normal distribution (Greene, 2003). Thus we perform three tests for normality of the inefficiency distributions (see Appendix 3) and present the results in Table 14 below.[67]

Table 14: Normality test for Inefficiency Estimates				
Time	**Firm Type**	**Jarque-Bera**	**Shapiro-Wilk W**	**Shapiro-Francia W'**
All Periods: 1990-2004	All Firms	556,909,366*	0.1446* (0.0000)	0.1429* (0.0000)
	Non-Table A	185,198,591*	0.1785* (0.0000)	0.1763* (0.0000)
	Table A	3,052,621*	0.3427* (0.0000)	0.3393* (0.0000)
1990-1994	All Firms	52,676,327*	0.1438* (0.0000)	0.1406* (0.0000)
	Non-Table A	18,669,110*	0.1750* (0.0000)	0.1708* (0.0000)
	Table A	494,903*	0.2511* (0.0000)	0.2450* (0.0000)
Phase I: 1995-1999	All Firms	4,086,812*	0.2421* (0.0000)	0.2397* (0.0000)
	Non-Table A	1,322,201*	0.3003* (0.0000)	0.2973* (0.0000)
	Table A	6,494*	0.5963* (0.0000)	0.5936* (0.0000)
Phase II: 2000-2004	All Firms	913,296*	0.2335* (0.0000)	0.2311* (0.0000)
	Non-Table A	285,185*	0.2897* (0.0000)	0.2867* (0.0000)
	Table A	1,408*	0.5272* (0.0000)	0.5229* (0.0000)
* Indicates significance at the five percent level. P-values for corresponding z statistics reported in parentheses. NB: The χ^2 critical value for the Jarque-Bera test is 5.99146.				

[67] Note also that the one-sided distribution for w_{it} in Eq. 20 should ensure that the inefficiency estimates are not distributed normally.

The results in Table 14 show that normality of the inefficiency samples is strongly rejected by all three tests. The two sample t-test and one factor analysis of variance (ANOVA) tests (both of which assume normally distributed samples) are therefore inadequate in our case, and we present the results of the Wilcoxon Rank-Sum[68] and Kruskal-Wallis mean-equality tests in our analysis of inefficiency differences. Our objective is to first test for mean inefficiency equality between Table A and non-Table A firms in order to further justify our separation of the estimates into subsets.[69] We then test for mean inefficiency differences between pre- and post-regulation periods for Table A and non-Table A firms. The results are presented in Table 15.

The four tests at the top of Table 15 imply that the inefficiency estimates for Table A and non-Table A firms were indeed drawn from different distributions. Note that the sign of the Wilcoxon test's *z* value provides information about which group has a larger mean; a negative *z* value implies that observations from sample 1, the former in the test statement, have a larger rank-sum and are thus larger on average than the observations from sample 2. A positive *z* implies the opposite. The positive *z* values for the first four tests in Table 15 therefore confirm the implications of Figure 2, indicating that the inefficiency estimates for Non-Table A firms were significantly higher than those for Table A firms.

The four Wilcoxon rank-sum tests at the bottom of Table 15 focus on testing the inefficiency differences before and after regulation for Table A and non-Table A firms. The first two use data from all years (five years prior to regulation and ten years post-regulation for Table A firms and ten years prior to regulation and five years post regulation for non-Table A firms), while the latter two use data only for the five years before and after regulation for each firm type in order to account for the effect of time on inefficiency and to keep the pre- and post-regulation time periods balanced. The sign of the *z* statistics imply larger mean inefficiency in the post-regulation period for all but the last test, though the p-values indicate that the differences in the inefficiency rank-sums before and after regulation were not significant. The negative *z* and corresponding p-value for the final Wilcoxon test in Table 15 indicate that the rank sum of inefficiencies in the period after regulation as smaller for non-Table A firms when only data from the five years before and after regulation were assessed.

The second test presented in Table 15 is the Kruskal-Wallis test for equality of population means (Kruskal and Wallis (1952)).[70] The *K* values presented in the final column of Table 15 support the findings of the Wilcoxon test in that they show that Table A and non-Table A inefficiencies appear to be drawn from different distributions, while pre- and post-regulation inefficiencies do not. The Kruskal-Wallis test, however, is unable to provide an indication of which group's mean is larger.

For more detailed evidence of cost efficiency differences before and after inclusion in the SO_2 scheme, we analysed firm-specific inefficiency means for the periods before and after regulation. Of the 11 Table A firms, only two saw a rise in mean inefficiency estimates for the five years following emissions regulation (1995-1999), while eight had smaller mean inefficiency estimates post-regulation. One firm had no observations from the period prior regulation. When the time period was extended to encompass all observations (i.e. five years prior to regulation and ten years post-regulation), the number of firms Table A who saw a

[68] See Appendix 4.
[69] Figure 2 provides the first visual evidence.
[70] See Appendix 5.

decline in inefficiency post-regulation fell to six, and four saw larger mean inefficiency estimates post regulation.

Table 15: Tests for Differences in Distribution of Inefficiency Estimates			
		Wilcoxon-Rank-Sum Test	**Kruskal-Wallis**
Null Hypothesis	**Data**	**z (p-value)**	**K (p-value)**
Table A Inefficiency= Non-Table A Inefficiency	All Firms, All Years	20.603** (0.0000)	424.464** (0.0001)
Table A Inefficiency= Non-Table A Inefficiency	All Firms, 1990-1994	11.901** (0.0000)	141.635** (0.0001)
Table A Inefficiency= Non-Table A Inefficiency	All Firms, 1995-1999	12.366** (0.0000)	152.913** (0.0001)
Table A Inefficiency= Non-Table A Inefficiency	All Firms, 2000-2004	11.285** (0.0000)	127.355** (0.0001)
Pre-Phase1 Inefficiency = Post-Phase1 Inefficiency	Table A Firms, All Years	0.429 (0.6680)	0.184 (0.6681)
Pre-Phase II Inefficiency = Post-Phase II Inefficiency	Non-Table A Firms, All Years	0.308 (0.7579)	0.095 (0. 7576)
Pre-Phase I Inefficiency = Phase I Inefficiency	Table A Firms, 1990-1999	0.362 (0.7173)	0.131 (0. 7171)
Pre-Phase II Inefficiency = Phase II Inefficiency	Non-Table A Firms, 1995-2004	-1.017** (0.3093)	1.034 (0.3093)
** Indicates significance at the 5% level. * Indicates significance at the 10% level.			

As for the 26 non-Table A firms, ten had higher mean inefficiency after regulation when only the five years before (1995-1999) and after (2000-2004) Phase II were considered, while 15 had lower mean inefficiency estimates in the post regulation period, with three of these mean values being approximately unchanged (two in the category of larger pre-regulation inefficiency, and one in the category of larger post-regulation). One firm had no observations available for the post-regulation period. When all time periods were included, the number of firms with lower mean inefficiency post-regulation rose to 18, while seven saw and increase in mean post regulation inefficiency. Having looked at the individual firms and their characteristics, there did not appear to be a quality shared by firms with either higher or lower post-regulation mean inefficiencies.

Overall, these results appear to support those related to scheme inclusion and the SO_2 price from the previous subsection. There, we saw an insignificant increase in cost inefficiency upon inclusion in the SO_2 scheme, a result confirmed by the insignificantly larger rank-sum of post-regulation inefficiencies for the first three of the last four tests in Table 15. The fact that the last Wilcoxon test in Table 15 indicates an (insignificant) decline in cost inefficiency in the period after regulation when the periods before and after regulation are balanced may be due to the inefficiency reducing effect of higher permit prices, as these rose to the highest levels in 2004. The analysis of individual firm mean inefficiency changes before and after regulation conforms with these hypotheses.

7.4 Marginal Effects

Table 16 presents the marginal effects on relative variable fuel cost inefficiency estimates of several of the inefficiency variables. Note that while the *signs* of the coefficients of the δ parameters presented in Table 8 are informative, the inefficiency component of the composed error, u_{it}, is distributed as a truncated normal random variable with mean $z_{it}\delta$, implying that the δ parameters can not be interpreted directly as marginal effects on cost inefficiency estimates. As noted, cost inefficiency is predicted as the conditional expectation of u_{it} *given* the observed composed error, ε_{it} (see Frame and Coelli (2001) and Piacenza (2002)) such that:

$$C\hat{E}_{it} = E\left(u_{it} \mid \varepsilon_{it} = e_{it}\right) = \left\{\exp(\mu_* + 0.5\sigma_*^2\right\}\left\{ \Phi\left(\left(\mu_* / \sigma_*\right) + \sigma_*\right) \Big/ \Phi\left(\mu_* / \sigma_*\right) \right\} \qquad \textbf{Eq. 23}$$

where $\mu_* = \dfrac{\sigma_v^2 z_{it}\delta + \sigma_u^2 \varepsilon_{it}}{\sigma_u^2 + \sigma_v^2}$, $\sigma_*^2 = \dfrac{\sigma_u^2 \sigma_v^2}{\sigma_u^2 + \sigma_v^2}$, and $\Phi(.)$ refers to the standard normal cumulative distribution function.

Following Wang (2002, 2003)[71], the *unconditional* mean and variance of u_{it} are then

$$E\left(u_{it}\right) = \sigma_u \left[\Lambda + \frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] \qquad \textbf{Eq. 24}$$

and

$$V(u_{it}) = \sigma_u^2 \left[1 - \Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] - \left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right] \qquad \textbf{Eq. 25}$$

where $\Lambda = z_{it}\delta \big/ \sigma_u^2$ and $\varphi(.)$ represents the standard normal density function, implying that the marginal effect of the k[th] element of the **z** vector, z_k, can be written:

$$\frac{\partial E(u_{it})}{\partial z_k} = \frac{\partial z_{it}\delta}{\partial z_k}\left[1 - \Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] - \left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right]. \qquad \textbf{Eq. 26}$$

We evaluate Eq. 28 for all observations for the relevant z variables and report the means in Table 16 below.

---

[71] Wang (2002, 2003) considers a model where the variance of the inefficiency component is parameterized and can vary with time and observation such that $\sigma_u^2 = \sigma_{it}^2$. Relevant changes to the mean and variance definitions have been made, and the portion of the marginal effect pertaining to the parameterized inefficiency variance has been omitted, as Wang's (2002, 2003) $\gamma[k]=0$ in our case.

In addition to information regarding the marginal effects of the various inefficiency variables, the results in Table 16 illuminate the role that firm size and scheme inclusion had on these marginal effects. It appears that firm and policy variables did not affect all firms equally, which should serve as support for the market-based environmental policies which allow firms to choose their own least-cost emissions reduction methods. As implied by the size dummies (δ_4 and δ_5) in Table 8 and their interpretation in Table 12, the marginal effects in Table 16 indicate that the cost inefficiency impacts of changes in the *z* variables are strongest for small firms and weakest for medium firms. This is possibly due to the fact that smaller firms have lower fuel switching capacity, leaving them more exposed to allocative inefficiency with a fixed capital stock and changing factor prices. It is also possible that the largest firms may have exceeded their efficient scale, leaving them slightly less cost efficient than their medium counterparts for reasons of technical efficiency and managerial behaviour.

The results in Table 16 also indicate that inclusion in the SO_2 scheme strengthened the cost-inefficiency-improving effects of permit prices, while the efficiency-reducing marginal effects of scrubbed capacity declined for firms included in the scheme. This supports our hypothesis relating to the increased cost effectiveness of scrubbers at higher SO_2 prices. We note again that these effects are strongest for small firms, who presumably have the least ability to switch away from scrubbed coal at lower sulphur prices. It is not surprising that scheme inclusion strengthened the marginal effects of coal's percentage of total capacity, as coal combustion creates the most SO_2, leaving firms with a large coal share more susceptible to the effects of SO_2 regulation. The impact of scheme inclusion on the efficiency-enhancing effects of increases in the capacity factor was small, but detectable.

Table 16: Mean Marginal Effects of z Variables on Inefficiency Estimates

Marginal Effects of ln(Permit Price)			
	Small Firms	Medium Firms	Large Firms
Not Included in Scheme	-0.0014	-0.00003	0.0002
Included in Scheme	-0.0625	-0.0025	-0.0028
Marginal Effects of Percentage of Percentage Coal Capacity			
	Small Firms	Medium Firms	Large Firms
Not Included in Scheme	-1.7171	-0.0663	-0.0765
Included in Scheme	-2.0667	-0.0843	-0.0962
Marginal Effects of Percentage of Capacity Covered by Scrubbers			
	Small Firms	Medium Firms	Large Firms
Not Included in Scheme	0.4662	0.0181	0.0243
Included in Scheme	0.3219	0.0137	0.0187
Marginal Effects of Capacity Factor			
	Small Firms	Medium Firms	Large Firms
Not Included in Scheme	-0.0170	-0.0007	-0.0008
Included in Scheme	-0.0184	-0.0008	-0.0009

The non-policy variables (coal's percentage of total capacity and capacity factor) had marginal effects with the negative same sign, indicating that they both shared a negative relationship with cost inefficiency. The effects of the permit price and percentage of scrubbed capacity are perhaps more interesting. We discuss the marginal effects of permit prices in greater detail below, but note that they generally shared a positive relationship with cost efficiency. For scrubbed capacity, the effect on cost efficiency was positive and still greatest for small firms. Upon inclusion, however, the increase in cost inefficiency with scrubbed capacity was less pronounced. These results may, again, be related to the capital stock, as small firms were presumably least able to alter their fuel mix/production processes in the face of changes in the other *z* variables. This would imply that they had the least ability to switch between scrubbed and non-scrubbed capacity even prior to the scheme, leading to their larger benefit from higher permit prices, which served to make scrubbers more cost effective.

We briefly discuss the effects of the permit price in greater detail. Increases in the natural log of the permit price appear to reduce inefficiency estimates most for small firms, followed by large, and then medium firms. For example, a one percent increase in the natural log of the permit price implies a 6.25% reduction in the relative cost inefficiency estimate for small included firms and a 0.28% reduction for large included firms. Excluded small firms also appear to see a small reduction in inefficiency estimates for increasing permit prices (a 0.14% decline for a one percent increase in ln(Permit Price)), while this inefficiency-improving effect declines to very nearly zero for medium excluded firms, and becomes positive for large excluded firms. It appears that increases in the permit price had a positive impact on cost efficiency, even when a firm was not included in the scheme. This could have been due to expectations surrounding future regulation, or to an outward shift in the minimum cost frontier resulting from higher SO_2 prices, an endogenous relationship between SO_2 prices and natural gas prices, or a combination of these effects.

Recall that the permit price affects both the firm's cost observation and the frontier itself, since the permit price is incorporated into the SO_2 price of each of the fuels as discussed in Section 6. This means that inefficiency changes related to SO_2 prices may be the result of changes in the frontier or in the position of cost observations relative to it. The scrubber cost explanation for the decline in relative cost inefficiency at higher permit prices appears plausible, as do the natural gas price in crease and SO_2 price effects, all of which would be associated with an increase in minimum cost defined by the frontier. It is, however, also possible that inefficiency improvements came from a reduction in the firm's variable costs of production (via managerial changes or reduced x-inefficiencies, say), or from a combination of expenditure reduction and a minimum-cost increase. Given that inclusion in the scheme appears to reduce cost efficiency, and given that scheme inclusion is incorporated into the frontier in a manner similar to the permit price, recovery in cost efficiency seen at higher permit prices is most probably not due solely to an increase in the minimum cost frontier, and presumably incorporates some degree of change in producer behaviour, particularly as Tuthill (2008) notes at least some potential for fuel switching in the face of relative fuel price changes. The analysis of the relative changes in the frontier and the expenditure observations are left for future research.

7.5 Summary

The results above provide several insights into the cost inefficiency effects of the CAAA 1990's SO_2 allowance scheme. First, we find evidence of a slight negative impact of SO_2 scheme inclusion on cost efficiency, but note that this effect does not appear to be significant. We note also that any increase in cost inefficiency upon regulation under the SO_2 scheme, significant or not, is presumably due to the capital stock, chosen optimally prior to the SO_2

scheme's onset, no longer being optimal in the face of SO_2 prices. Second, we find that Table A firms showed a greater level of cost efficiency than non-Table A firms across all years. Third, our results relating to the effects of firm size and coal's percentage of total capacity indicate that the effects of these are presumably due to scale economies and a "fleet effect" (Hiebert, 2002) realized by firms operating with a technology specialization. This supports Swinton's (1998) findings that coal-burning electric plans with the highest emissions rates also tend to have the lowest marginal abatement costs.

Fourth, we show that the increase in aggregate mean cost inefficiency estimates from periods prior to regulation to periods post-regulation were insignificant for both Table A and non-Table A firms (the former coming under regulation during Phase I and the latter during Phase II), confirming the results found from the SFA model's inefficiency equation estimate, and indicating that inclusion in the SO_2 scheme did not appear to have a significant effect on the ability of firms to achieve minimum-cost operation. Finally, we find that higher permit prices were associated with cost efficiency improvements. We note that this was probably due to a combination of the increased cost effectiveness of scrubbed capacity with higher permit prices (i.e. to shifts in the cost frontier) and changes managerial behaviour and reductions in x-inefficiencies as emissions prices rise (i.e. to changes in cost observations' location relative to the frontier). Thus, while it does appear that the imposition of the scheme did not significantly alter firms' ability to achieve minimum-cost, firms do appear to have experienced an improvement in cost efficiency with increased regulatory stringency (i.e. through lower emissions caps and higher permit prices). We believe that these results are tied to SO_2-related capital, and may not therefore generalize to the case of CO_2 regulation. This would, however, make interesting further research.

8 Conclusion

With regulators of the electricity industry balancing objectives relating to both environmental quality and energy supply at minimum cost, the effect of recently popular cap-and-trade emissions schemes on distortions from minimum cost operation in the power industry becomes an interesting question and provides the motivation for the analysis in this paper. Our analysis focused on the impacts of the American cap-and-trade SO_2 allowance scheme and other firm characteristics on the cost-minimizing behaviour of a sample of 37 US electricity generating firms over the years 1990-2004. We used a stochastic frontier analysis model based on the cost frontier dual of Battese and Coelli's (1995) model of technical inefficiency effects, which allows for the estimation of a second cost inefficiency equation simultaneously along with the cost frontier. Accordingly, we used Coelli's (1996) Frontier 4.1 software to estimate firm- and time-specific cost inefficiencies and to determine the effects of several firm- and policy-specific variables (namely firm size, coal's percentage of total capacity, percentage of capacity covered by scrubbers, scheme inclusion and permit price) on the cost inefficiency estimates. We then tested for differences in inefficiency estimates by firm type and across time, and particularly for differences before and after the onset of emissions regulation.

Our findings suggest that the small negative impact of the CAAA 1990's SO_2 scheme on aggregate mean cost efficiency was not significant, and the analysis of individual firm's mean cost inefficiency before and after regulation confirms this result. Any small negative effects of inclusion on cost efficiency are presumably due to capital-based combustion constraints, and we note that the insignificance of these inefficiency impacts are possibly explained by firms' fuel-switching ability as analysed in Tuthill (2008). We also found that coal's percentage of total capacity shared a positive relationship with cost efficiency, while the percentage of capacity covered by scrubbers was generally associated with a decline in cost efficiency. Our results pertaining to firm size indicate that the marginal effects of all of the explanatory variables in our inefficiency equation were largest for small firms. We also found that increases in the permit price appeared to be linked to cost efficiency recovery. We ascribed this last effect at least partly to the increased cost effectiveness of scrubbed capacity at higher SO_2 prices, and note that cost efficiency improvements with higher permit prices may not, therefore, generalize to the case of cap-and-trade CO_2 regulation.

Several interesting extensions are possible. The cost efficiency estimates could be decomposed into their technical and allocative inefficiency components (TE and AE, respectively) with the use of cost share equations, and TFP growth could be analysed and decomposed into contributions from scale economies, technical change as well as technical and allocative efficiency components, particularly if the TE/AE decomposition is undertaken. We could also add firms to the dataset (i.e. those with hydro or nuclear capacity, those with only coal and gas capacity, those with other renewable capacity), and state-level regulations could be incorporated. A similar analysis with data from the European Union (EU) under the EU Emissions Trading Scheme would also be interesting now that a reasonable amount of data is available. These extensions are saved for future research.

Appendix 1

Piacenza (2002) obtains the log-likelihood function for the cost frontier dual to Battese and Coelli's (1993, 1995) technical inefficiency effects model. Battese and Coelli (1993) provide the log-likelihood function for the production frontier specification where the inefficiency error component measures *technical* inefficiency and firms are assumed to operate *below* the production frontier. Here, firms are assumed to be minimizing cost (rather than maximizing output) and thus are assumed to operate *above* the frontier. As such, the composed error term $\varepsilon_{it} = v_{it} + u_{it}$ has the inefficiency component, u_{it}, *added*[72] to the random noise component, v_{it}, as seen in Equations 8 and 10 of Section 4.

We present Piacenza's (2002) derivation of the log-likelihood function and corresponding cost efficiency measure for the model represented by Equations 22 and 23 in Section 5. What follows is derived from the log-likelihood function presented in the appendix of Battese and Coelli (1993) by altering a few signs in order to accommodate the cost frontier's definition of the composed error term.

Begin with the stochastic cost frontier model from Section 5 defined here as:

$$vc_{it} = vc(x_{it};\beta) + \varepsilon_{it} \qquad \textbf{Eq. A27}$$

$$\varepsilon_{it} = v_{it} + u_{it} \qquad \textbf{Eq. A28}$$

and i= 1, …, I and t= 1, …, T_f such that an unbalanced panel is accommodated.

Here, $vc_{it}=\ln VC_{it}$ and x_{it} is a vector of explanatory variables from the cost frontier. As in Section 5, $v_{it} \sim NID(0,\sigma_v^2)$ and $u_{it} \sim N^+(z_{it}\delta,\sigma_u^2)$. We omit subscripts in what follows to avoid clutter.

The density functions for v_{it} and u_{it} are modified from the simple exposition in Section 4 to incorporate the cost inefficiency effects:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left\{-\frac{v^2}{2\sigma_v^2}\right\}, \quad -\infty < v < \infty \qquad \textbf{Eq. A29}$$

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u\Phi\left[\delta' z \big/ \sigma_u\right]}\exp\left\{-\frac{(u-\delta' z)^2}{2\sigma_u^2}\right\}, \quad u \geq 0, \qquad \textbf{Eq. A30}$$

where Φ[.] is the standard normal cumulative distribution.

The joint density for *u* and *v* is then just the product of their individual densities, given the assumption of independence, and the joint density function for *ε* and *u* (given *ε*=*v*+*u*) is

[72] As opposed to *subtracted* as in the case of Battese and Coelli's (1993, 1995) model of technical inefficiency effects.

$$f(\varepsilon,u)=\frac{\exp\left\{\frac{-(u-\mu_*)^2}{2\sigma_*^2}-\frac{(\varepsilon-\delta' z)^2}{\sigma_v^2+\sigma_u^2}\right\}}{2\pi\sigma_u\sigma_v\Phi\left[\delta' z/\sigma_u\right]},\quad u\geq 0, \qquad \textbf{Eq. A31}$$

where

$$\mu_*=\frac{\sigma_v^2\delta' z+\sigma_u^2\varepsilon}{\sigma_v^2+\sigma_u^2}\quad \textit{and}\quad \sigma_*^2=\frac{\sigma_v^2\sigma_u^2}{\sigma_v^2+\sigma_u^2}.$$

The marginal density function for $\varepsilon = v+u$ (which, recall, will be used to obtain estimates of u) is obtained by integrating u out of Equation A5:

$$f(\varepsilon)=\frac{\exp\left\{\frac{-(\varepsilon-\delta' z)^2}{2(\sigma_v^2+\sigma_u^2)}\right\}}{\sqrt{2\pi}(\sigma_v^2+\sigma_u^2)^{1/2}\left\{\Phi\left[\delta' z/\sigma_u\right]\Big/\Phi\left[\mu_*/\sigma_*\right]\right\}}. \qquad \textbf{Eq. A32}$$

Then, following Piacenza (2002) and using Eq A6, the density function for the cost observation from Eq A1 is

$$f(vc_{it})=\frac{\exp\left\{-\frac{1}{2}\left(\frac{(vc_{it}-vc(x_{it};\beta)-\delta' z)^2}{\sigma_v^2+\sigma_u^2}\right)\right\}}{\sqrt{2\pi}\left(\sigma_v^2+\sigma_v^2\right)^{1/2}\left\{\Phi[d_{it}]\Big/\Phi[d_{it}^*]\right\}}, \qquad \textbf{Eq. A33}$$

where $d_{it}=\delta' z/\sigma_u$, $d_{it}^*=\mu_{it}^*/\sigma_*$, and $\mu_{it}^*=\dfrac{\sigma_v^2\delta' z_{it}+\sigma_u^2(vc_{it}-vc(x_{it};\beta))}{\sigma_v^2+\sigma_v^2}$.

Because the panel is unbalanced, there are T_i observations for firm i, where $1\leq T_i\leq T$. Denote the vector of variable cost observations for firm i as $vc_i\equiv(vc_{i1},...,vc_{iT_i})'$. Then the log likelihood for the full sample of observations, $vc\equiv(vc_1',...,vc_I')'$ is:

$$\begin{aligned}L(\Theta;vc)=&-\frac{1}{2}\left(\sum_{i=1}^{I}T_i\right)\left\{\ln 2\pi+\ln\sigma^2\right\}\\&-\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T_i}\left\{\frac{(vc_{it}-vc(x_{it};\beta)-\delta' z_{it})^2}{\sigma^2}\right\}\\&-\sum_{i=1}^{I}\sum_{t=1}^{T_i}\left\{\ln\Phi[d_{it}]-\ln\Phi[d_{it}^*]\right\}\end{aligned}, \qquad \textbf{Eq. A34}$$

where Battese and Corra's (1977) parameterization is used as mentioned in Sections 4 and 5 such that $\sigma^2=\sigma_v^2+\sigma_v^2$ and $\gamma=\sigma_u^2\Big/\sigma_v^2+\sigma_u^2$. Also in Eq A8, $\Theta\equiv(\beta',\delta',\sigma^2,\gamma)$, and

$d_{it} = \frac{\delta' z_{it}}{\left(\gamma\sigma^2\right)^{1/2}}$, $d_{it}^* = \frac{\mu_{it}^*}{\left[\gamma(1-\gamma)\sigma^2\right]^{1/2}}$, $\mu_{it}^* = (1-\gamma)\delta' z_{it} + \gamma(vc_{it} - vc(x_{it};\beta))$ and $\sigma_* = \left[\gamma(1-\gamma)\sigma^2\right]^{1/2}$ as interpreted from the definitions provided after Equation A7.

Equation A8 is then maximized (or, in the case of Frontier 4.1, the negative of Eq A8 is then minimized) with respect to each element of Θ to obtain ML estimates of β, δ, σ^2, and γ as explained in Section 4.

Estimates of cost inefficiency component of the composed error term for each firm and each time period can then be derived via the conditional distribution of u_{it} given ε_{it} as suggested by Jondrow et al. (1982). Again omitting subscripts and using Equations A5 and A6 above, this conditional density function is

$$f(u \mid \varepsilon) = \frac{\exp\left\{-\frac{1}{2}\left({(u-\mu_*)^2}/{} \right)\right\}}{\sqrt{2\pi}\sigma_* \Phi\left[{\mu_*}/{\sigma_*}\right]}. \qquad \textbf{Eq. A35}$$

Cost *inefficiency* is then estimated in a modified version of Equation 17 of Section 4, as the ratio of observed cost to the frontier minimum cost where $u_{it}=0$ such that[73]

$$CE_{it} = \frac{c(y,w;\beta)\exp\{v_{it} + u_{it}\}}{c(y,w;\beta)\exp\{v_{it}\}} = \exp\{\hat{u}_{it}\}. \qquad \textbf{Eq. A36}$$

Estimates of CE_{it} will thus be bounded by 1 and infinity, where $CE_{it}=1$ implies full relative cost efficiency. The estimates for Equation A10 are obtained as the *inverse* of the following (see footnote 2), which is derived from the conditional density function given in Equation A9, and is a generalization of the results of Battese and Coelli (1993), Battese and Coelli (1988) and Jondrow et al. (1982):

$$C\hat{E}_{it} = E(\exp\{-u_{it}\} \mid \mathcal{E}_{it} = \varepsilon_{it}) = \left[\frac{\Phi\left[\left({\mu_{it}^*}/{\sigma^*}\right) - \sigma_*\right]}{\Phi\left[{\mu_{it}^*}/{\sigma_*}\right]}\right] \exp\left\{-\mu_{it}^* + \frac{1}{2}\sigma_*^2\right\}, \qquad \textbf{Eq. A37}$$

where E(.) denote the conditional expectation of CE_{it} given the observation of ε_{it}, and μ_{it}^* and σ_* are as defined above.

[73] Note that Piacenza's (2002) appendix illustrates the measure of cost *efficiency*.

Appendix 2

The three normality tests we performed on the inefficiency estimates were the Jarque-Bera, Shapiro-Wilk and Shapiro-Francia tests. The Jarque-Bera test (Jarque and Bera (1980, 1981)) tests the joint null hypothesis that the skewness of the distribution from which the data is drawn is equal to zero and that the kurtosis is equal to three[74]. The test statistic is, then:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4} \right) \qquad \textbf{Eq. A12}$$

where n= the number of observations, S is the skewness of the sample, and K is the sample's kurtosis[75]. The statistic has a χ^2 distribution with two degrees of freedom, and the results show a strong rejection of the null of a normally distributed inefficiency population.

The Shapiro-Wilk test[76] (Shapiro and Wilk (1965)) is the ratio of a "best estimator" of the sample variance to the standard sum-of-squares estimate that tests for normality by regressing observed ordered sample values on corresponding ordered statistics from a sample drawn from normal distribution. The statistic, W, requires a sample size $7 \le n \le 2{,}000$ and looks as follows:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)} \right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \textbf{Eq. A13}$$

where the $x_{(i)}$ are the ordered sample values ($x_{(1)}$ being the smallest) and

$$a_i = m'V^{-1}\left[m'V^{-1}V^{-1}m \right]^{-1/2} \qquad \textbf{Eq. A14}$$

where m'= (m_1, …,m_n) is a vector of expected values of standard normal order statistics from an equivalent sample size and V is the n x n covariance matrix. Small values of W indicate non-normality, and $0 \le W \le 1$. (See Park (2006) and Greene (2003) for further information.)

The final normality test we perform is the Shapiro-Francia test, proposed by Shapiro and Francia (1972) and Royston (1983) as a simplification of the Shapiro-Wilk test, and both have been shown to be consistent (Sarkadi (1975)). The W' statistic is calculated as the W in equation 7 above, except that a' is replaced by b':

$$b' = m'(m'm)^{-1/2}. \qquad \textbf{Eq. 38A15}$$

The Shapiro-Francia test can be performed by Stata on a sample size $5 \le n \le 5{,}000$, and $0 \le W' \le 1$ as above. The p-values for both the W and W' estimates are provided in parentheses in Table 14 in Section 7.3.

---

[74] Both of these hypotheses would hold if the data were normally distributed.

[75] $S = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$ and $K = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{2}}$ and are reported in Table 6 in Section 6.

[76] The Shapiro-Wilk and Shapiro-Francia test were both performed with STATA. The Jarque-Bera was calculated by hand from the statistics in Table 6.

Appendix 3

The Wilcoxson rank-sum test (a.k.a. the Mann Whitney two sample statistic) (Wilcoxon (1945) and Mann and Whitney (1947)) is a non-parametric alternative to the two-sample t-test that is valid for non-normal distributions and is less sensitive to outliers. It tests the null hypothesis that two samples are drawn from the same population and is based on the ordering of the observations from two samples where each observation is assigned a rank (smallest observation has rank=1, largest has rank=n, where n is the total number of observations). The ranks are then summed for each group such that $R_1 = \sum_{i=1}^{n_1} rank_i$ for group 1 and $R_2 = \sum_{i=1}^{n_2} rank_i$ for group 2. The test statistic, *U*, is calculated as:

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - (R_2 - R_1) \qquad \textbf{Eq. A16}$$

where n_1 and n_2 represent the number of observations in samples 1 and 2, respectively, and $n_2 > n_1$. Stata uses a normal approximation of the *U* test:

$$z = U - m_U \big/ \sigma_U \qquad \textbf{Eq. A17}$$

where m_u and σ_U are the mean and standard deviation of *U* when the null hypothesis of equivalent distributions is true. The *z* value, reported in Table 15, is then a standard normal deviate whose significance can be checked with the normal distribution[77]. The Wilcoxon rank-sum results in Table 15 of Section 7.3 indicate that the samples tested in all cases were, indeed, drawn from different distributions.

[77] See "What Statistical Analysis…" (UCLA) for the application of the Wilcoxon Rank-Sum and Kruskal-Wallis test (below) in Stata.

References

Afriat, S.N. (1972). "Efficiency Estimation of Production Functions," *International Economic Review*. 46(2): 5-23.

Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977), "Formulation and estimation of stochastic frontier production function models". *Journal of Econometrics.* 5: 21–38.

Aigner, D.J. (1968). "On Estimating the Industry Production Function," *The American Economic Review*. 58(4): 826-839.

Aiken, D.V. and C.A. Paskura, Jr. (2003). "Adjusting the Measurement of US Manufacturing Productivity for Air Pollution Emissions Control", *Resource and Energy Economics*. 25: 329-351.

Albrecht, J., T. Verbeke and M. De Clerq (2006). "Informational Efficiency of the US SO_2 Permit Market," *Environmental Modelling and Software*. 21(10): 1471-1478.

Atkinston, S. and J. Dorfman (2005). "Bayesian Measurement of Productivity and Efficiency in the Presence of Undesirable Outputs: Crediting Electric Utilities for Reducing Air Pollution," *Journal of Econometrics*. 126: 445-468.

Atkinson, S. and R. Halvorsen (1984). "Parametric Efficiency Tests, Economies of Scale and Input Demand in US Electric Power Generation," *International Economic Review*. 25(3): 647-662.

Baltagi, B. and J. Griffin (1988). "A General Index of Technical Change," *The Journal of Political Economy*. 96(1): 20-41.

Battese, G. E. and T. J. Coelli (1988). "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data", *Journal of Econometrics*. 338: 387-399.

Battese, G. E. and T. J. Coelli (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India", *The Journal of Productivity Analysis.* 3: 153-169.

Battese, G.E. and T.J. Coelli (1993), "A Stochastic Frontier Production Incorporating a Model for Technical Inefficiency Effects," Working Papers in Applied Statistics, No. 65, Department of Economics, University of New England, Amrmidale, NSW, Australia.

Battese, G.E. and T.J. Coelli (1995), " A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data". *Empirical Economics.* 20**:** 325–332.

Battese, G.E. and G.S. Corra (1977), "Estimation of a Production Frontier with Application to the Pastoral Zone of Eastern Australia," *Australian Journal of Agricultural Economics.* 21: 169–179.

Bauer, P.W. (1990). "Recent Developments in the Econometric Estimation of Frontiers", *Journal of Econometrics.* 46(1-2):39–56.

Bellas, A. and I. Lange (2008). "Impacts of Market-Based Environmental and Generation Policy on Scrubber Electricity Usage," *The Energy Journal*. 29(2): 151-164.

Bera, A.K. and C.M. Jarque (1980). "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals". *Economics Letters* 6 (3): 255–259.

Bera, A. K. and C. M. Jarque (1981). "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo evidence". *Economics Letters* 7 (4): 313–318.

Bohi, D.R. and D. Burtraw (1997). "SO2 Allowance Trading: How Experience and Expectations Measure Up," *Resources for the Future*: Discussion Paper 97-24.

Burns, P. and T.G. Weyman-Jones (1996). "Cost Functions and Cost Efficiency in Electricity Distribution: A Stochastic Frontier Approach", *Bulletin of Economic Research*. 48(1): 41-64.
Burtraw, D. (1996). "Cost Savings Sans Allowance Trades? Evaluating the SO2 Allowance Trading Program to Date," Resources for the Future, *Discussion Paper 95-30-REV.*
Bushnell, J.B. and C. Wolfram (2005). "Ownership Change, Incentives and Plant Efficiency: The Divestiture of US Electricity Generating Plants". CSEM, UCEI, University of California at Berkeley. CSEM Working Paper No. 140.
Cambini, C. and M. Filippini (2003). "Competitive Tendering and Optimal Size in the Regional Bus Transportation Industry: An Example from Italy," *Annals of Public and Cooperative Economics.* 74(1): 163-182.
Carlson, C. and D. Burtraw, M Cropper and KL Palmer (2000). "Sulfur Dioxide Control by Electric Utilities: What are the Gains from Trade?," *Journal of Political Economy*. 108(6): 1292-1326.
Carrington, R. and T.J. Coelli, and E. Groom (2002). "International Benchmarking for Monopoly Price Regulation: The Case of Australian Gas Distribution." *Journal of Regulatory Economics*. 21(2): 191-216.
Caves, D., L. Christensen and W.E. Diewert (1982). "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica*. 50(6): 1393-1414.
Chambers, R. (1988). *Applied Production Analysis: A Dual Approach.* New York: Cambridge University Press.
Charnes, A., W.W. Cooper, and E. Rhodes (1981). "Evaluating Program and Managerial Efficiency: An Application of Panel Data Envelopment Analysis to Program Follow Through," *Management Science*. 27(6): 668-697.
Charnes, A., W.W. Cooper, and E. Rhodes (1978). "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*. 2(6): 429-444.
Chicago Climate Futures Exchange (2004). "The Sulfur Dioxide Emission Allowance Trading Program: Market Architecture, Market Dynamics and Pricing." Chicago Climate Exchange, Inc. Available at: http://www.ccfe.com/education_ccfe/SO2_Background_Drivers_Pricing_PDF.pdf
Christopoulos, D.K., J. Loizides and E.G. Tsionas (2000). "Measuring Input-Specific Technical Inefficiency in European Railways: A Panel Data Approach", *International Journal of Transport Economics*. 27(2), 147-171.
Clean Air Act, US Senate Committee on Environment and Public Works, 24 February 2004. Available online at: http://epw.senate.gov/rnblsed/cleanair.pdf.
Coelli, T.J. (1996) " A guide to Frontier 4.1: a computer program for stochastic frontier production and cost estimation," Center for Efficiency and Productivity Analysis Working Paper 96/07, Department of Economics, University of New England, Amrmidale, NSW, Australia.
Coelli, T. J. (1995). "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis," *Journal of Productivity Analysis*. 6(3): 247-268.
Coelli, T., A. Estache, S. Perelman, L. Trujillo (2003a). "A Primer on Efficiency Measurement for Utilities and Transport Regulators." (Washington, DC: World Bank Institute Development Studies, World Bank.)
Coelli, T., S. Singh and E. Fleming (2003b). "An Input Distance Function Approach to the Measurement of Technical and Allocative Efficiency: With Application to Indian Dairy Processing Plants". Draft available online: http://gemini.econ.umd.edu/cgi-bin/conference/download.cgi?db_name=esam2003&paper_id=105

Coggins, J. S., and J. R. Swinton. 1996. "The Price of Pollution: A Dual Approach to Valuing Allowances." *Journal of Environmental Economics and Management* 30(1): 58-72.
Cornwell, C., P. Schmidt and R.C. Sickles (1990). "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels," *Journal of Econometrics*. 46(1/2): 185-200.
Datta, A. and S. Christofferson (2005). "Production Costs, Scale Economies and Technical Change in US Textile and Apparel Industries". *Atlantic Economic Journal*. 33(2): 201-213.
De Alessi, L. (1983) "Property Rights, Transaction Costs, and X-Efficiency: An Essay in Economic Theory," *American Economic Review*, 73(1): 64-81.
Debreu, G (1951). "The Coefficient of Resource Utilization," *Econometrica*. 19(3): 273-292.
Dorfman, J. H. and G. Koop (2005). "Current Developments in Productivity and Efficiency Measurement," *Journal of Econometrics*. 126: 233-240.
Edison Electric Institute (EEI) (2001), "New Source Review: A History". Available at http://www.eei.org/industry_issues/environment/air/New_Source_Review/NSR_history.pdf
Ellerman, A.D. and J-P Montero (2005). "The Efficiency and Robustness of Allowance Banking in the US Acid Rain Program," *Massachusetts Institute of Technology, Center for Energy and Environmental Policy Research*. Working Paper 0505.
Ellerman, A.D. and J-P Montero (1998). "The Declining Trend in Sulfur Dioxide Emissions: Implications for Allowance Prices," *Journal of Environmental Economics and Management,* 36(1): 26-45.
Energy Information Administration (EIA) (1996), "The Changing Structure of the Electric Power Industry: An Update," DOE/EIA-0562(96), Distribution Category UC-950.
Energy Information Administration (EIA) (2001), "US Natural Gas Markets: Recent Trends and Prospects for the Future," SR/OIAF 2001-2002. Available at: http://tonto.eia.doe.gov/FTPROOT/service/oiaf0102.pdf
Energy Information Administration (EIA), "Electric Power Industry". Available at: www.eia.doe.gov/cneaf/electricity/page/prim2/chapter1.html. Chapters 1-6.
EIA, "Electric Power Industry Restructuring Fact Sheet". Available at http://www.eia.doe.gov/cneaf/electricity/page/fact_sheets/restructuring.html.
EIA, "Energy Information Sheets". Available at: www.eia.doe.gov/neic/infosheets/preface.html.
EIA, FERC Form 423 Database, "Monthly Cost and Quality of Fuels for Electric Plants Data". Available at: http://www.eia.doe.gov/cneaf/electricity/page/ferc423.html
EIA, Form EIA 759/906, Energy Information Administration, 1990-2004. Available at: www.eia.doe.gov/cneaf/electricity/page/eia906u.html.
EIA, Form EIA 860, "Annual Electricity Generator Report", 1990-2004. Available at: www.eia.doe.gov/cneaf/electricity/page/eia860.html .
EIA, (1995), "Electricity Generation and Environmental Externalities: Case Studies," Report no. DOE-EIA 0598. Available at www.eia.doe.gov/cneaf/electricity/external/external.pdf.
EIA (1996), "The Changing Structure of the Electric Power Industry: An Update", DOE/EIA-0562(96), Distribution Category UC-950.
EIA (1997), "The Effects of Title IV of the Clean Air Act Amendments of 1990 on Electric Utilities: An Update," Report no: DOE/EIA-0582(97). Distribution Category: UC-950.

EIA (2001), “US Natural Gas Markets: Recent Trends and Prospects for the Future”, SR/OIAF 2001-2002. Available at: tonto.eia.doe.gov/FTPROOT/service/oiaf0102.pdf
EIA (2004a), *Annual Energy Outlook 2004 with Projections to 2025*. Report no. DOE/EIA-0383(2004). Available at: www.eia.doe.gov/oiaf/archive/aeo04/index.html
EIA (2004b), “Assumptions for the Annual Energy Outlook 2004 with Projections to 2025*”*, Report no. DOE/EIA-0554(2004). Available at http://www.eia.doe.gov/oiaf/archive/aeo04/assumption/index.html.
EIA (2006), *Electric Power Annual 2005*. Report no. DOE/EIA-0348(2005). Available at: tonto.eia.doe.gov/ftproot/electricity/034805.pdf
EIA (2006), “Annual Energy Review 2006”, Report no. DOE/EIA-0384-2006. Available at www.eia.doe.gov/emeu/aer/consump.html.
EIA (2007a), “Electric Power Annual 2006”, Report No: DOE/EIA 0348(2006). Available at: www.eia.doe.gov/cneaf/electricity/epa/epa.pdf.
EIA (2007b), “Annual Energy Outlook 2007 with Projections to 2030,”. Report no.: DOE/EIA-0383(2007). Available at www.eia.doe.gov/oiaf/aeo/index.html.
EIA and EPA (2000), “Carbon Dioxide Emissions from the Generation of Electric Power in the United States”. Available at tonto.eia.doe.gov/FTPROOT/environment/co2emiss00.pdf.
EPA, “Clean Air Markets- Allowance Trading-Allowance Prices (1995-2004)”. Available at www.epa.gov/airmarkets/progress/alprices.html
EPA, “Clean Air Markets- Data and Maps”. Last updated 10 January 2007. cfpub.epa.gov/gdm/index.cfm?fuseaction=emissions.wizard.
EPA, “Clean Air Markets- Programmes and Regulations – Acid Rain Programme,” Available at www.epa.gov/airmarkets/progsregs/arp/basic.html
EPA, “Clean Energy- How Does Energy Use Affect the Environment?”. Available at: www.epa.gov/solar/energy-and-you/affect/index.html.
EPA (1999), “Emissions Factors, Global Warming Potentials, Unit Conversions, Emissions, and Related Facts,” complied by ICF Consulting. Available at www.epa.gov/appdstar/pdf/brochure.pdf.
EPA (2000), AP 42, Fifth Edition**, “**Compilation of Air Pollutant Emission Factors, Volume 1: Stationary Point and Area Sources”, Available at: www.epa.gov/ttn/chief/ap42/index.html.
EPA, (2007), “The Plain English Guide to the Clean Air Act”, Publication Number EPA-456/K-07-001.
Estache, A., M. Rossi and C. Ruzzier (2004). “The Case for International Coordination of Electricity Regulation: Evidence from the Measurement of Efficiency in South America,” *Journal of Regulatory Economics*. 25(3): 271-295.
Fan, S. (2000). “Technological Change, Technical and Allocative Efficiency in Chinese Agriculture: The Case of Rice Production in Jiangsu”, *Journal of International Development*. 12(1): 1-12.
Färe, R., S. Grosskopf, and C. A. K. Lovell.(1994). *Production Frontiers*. Cambridge: Cambridge University Press.

Färe, R., S. Grosskopf, C.A.K. Lovell, and Carl Paskura (1989). “Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach”, *The Review of Economics and Statistics.*
Färe, R., S. Grosskopf and C.A.K. Lovell (1992). “Indirect Productivity Measurement,” *Journal of Productivity Analysis*. 2: 283-298.

Färe, R. and S. Grosskopf (1990). "A Distance Function Approach to Price Efficiency," *Journal of Public Economics*. 43(1): 123-126.
Färe, R. and S. Grosskopf (2004). *New Directions: Efficiency and Productivity*. Kluwer Academic Publishers.
Färe, R., S. Grosskopf, D-W Noh, W. Weber (2005). "Characteristics of a Polluting Technology: Theory and Practice," *Journal of Econometrics*. 126: 469-492.
Färe, R. and D. Primont (1995). *Mutli-Output Production and Duality: Theory and Applications.* Kluwer Academic Publishers.
Farrell, M.J. (1957), "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society,* 120(3): 253-290.
Førsund, F.R., C.A.D. Lovell, P. Schmidt (1980) A survey of frontier production functions and their relationship to efficiency measurement. Journal of Econometrics 13:5–25.
Frame, W.S. and T.J. Coelli (2001). "U.S. Financial Services Consolidation: The Case of Corporate Credit Unions," *Review of Industrial Organization*. 18(2): 229-241.
Fraquelli, G., M. Piacenza, and G. Abrate (2004). "Regulating Public Transit Networks: How do Urban-Intercity Diversification and Speed-up Measures Affect Firms" Cost Performance?"*Annals of Public and Cooperative Economics* 75 (2): 193–225.
Fried, Harold O., C. A. Knox Lovell, Shelton S. Schmidt (1993). "The Measurement of Productive Efficiency: Techniques and Applications". (Oxford: Oxford University Press.)
Gollop, F.M. and M.F. Roberts (1983). "Environmental Regulations and Productivity Growth: The Case of Fossil-Fueled Electric Power Generation," *The Journal of Political Economy*. 91(4): 654-674.
Greene, W. (2005). "Fixed and Random Effects in Stochastic Frontier Models", *Journal of Productivity Analysis*. 23: 7-32.
Greene, W. (2003). *Econometric Analysis.* 5th Edition. Prentice Hall: Upper Saddle River, NJ.
Greene, W. (1993), "The econometric approach to efficiency analysis". In: Fried H.O,. Lovell C.A.K., Schmidt S.S. (Eds) *The measurement of productive efficiency: Techniques and applications*. New York: Oxford University Press. Pp 68–119.
Greene, W. (1997). "Frontier Production Funcitons," in M.H. Pesaran and P. Schmidt eds, *Handbook of Applied Econometrics, Volume II: Microeconomics*. Oxford: Blackwell Publishers, Ltd.
Greene, W. (1980a). "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics*. 13(1): 27-56.
Greene, W. (1980b). "On the Estimation of a Flexible Frontier Production Model," *Journal of Econometrics*. 13(1): 101-115.
Grosskopf, S. and R. Fare (1998). "An Indirect Approach to the Evaluation of Producer Performance", *Journal of Public Economics*. 37: 71-89.
Hattori, T. T Jamash and M. Pollitt (2005). "Electricity Distribution in the UK and Japan: A Comparative Efficiency Analysis 1985-1998." *The Energy Journal*. 26(2): 23-47.
Hausman, J.A. and W.E. Taylor (1981). "Panel Data and Unobservable Individual Effects," *Econometrica*. 9(6):1377-1398.
Helfand, G.E., M.R. Moore, and Y. Liu (2006). "Testing for Dynamic Efficiency of the Sulfur Dioxide Allowance Market," Working Paper, University of Michigan.

Helm, D, C. Hepburn and R. Mash (2003). "Credible Carbon Policy," *Oxford Review of Economic Policy*. 19(3): 438-450.
Herrero, I. and S. Pascoe (2002). "Estimation of Technical Efficiency: A Review of Some of the Stochastic Frontier and DEA Software", *Computers in Higher Education Economics Review*. 15(1).
Hiebert, L.D. (2002) "The Determinants of Cist Efficiency of Electric Generating Plants: A Stochastic Frontier Approach", *Southern Economic Journal*. 68(4), 935-946.
Hollingsworth, B. (2003). "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care", *Health Care Management Science*. 6: 203-218.
Huang C.J. and J-T Liu (1994) "Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5: 171–180.
Insley, M.C. (2003). "On the Option to Invest in Pollution Control Under a Regime of Tradable Emissions Allowances," *Canadian Journal of Economics*. 36(4): 860-883.
Jarque, B. and A. Bera (1980). "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals," *Economics Letters.* 6(3): 255-259.
Jarque, B. and A. Bera (1981). "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence," *Economics Letters.* 7(4): 313-318.
Jondrow, J., C.A.K. Lovell, I.S. Materov and P. Schmidt (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics*. 19(2/3): 233-38.
Joskow, P.L., R. Schmalensee and E.M. Bailey (1998). "The Market for Sulfur Dioxide Emissions," *The American Economic Review*. 88(4): 669-685.
Joskow, P.L. and R. Schmalensee (1987). "The Performance of Coal-Burning Electric Generating Units in the United States: 1960-1980," *Journal of Applied Econometrics*. 2(2): 85-109.
Kalirajan, K. (1981), "An Econometric Analysis of Yield Variability in Paddy Production," *Canadian Journal of Agricultural Economics.,* 29: 283–294.
Knittel, Christopher R. (2002). "Alternative Regulatory Methods and Firm Efficiency: Stochastic Frontier Evidence From the U.S. Electricity Industry," *The Review of Economics and Statistics*. 84(3): 530-540.
Kodde, D. and F. Palm (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica.* 54: 1243–1248.
Koopmans, T.C. (1951). "An Analysis of Production as an Efficient Combination of Activities," in TC Koopmans, ed., *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph No. 13, New York: Wiley.
Kruskal, W.H. and W.A. Wallis (1952). "Use of ranks in one-criterion variance analysis." *Journal of the American Statistical Association* 46(260): 583–621.
Kumbhakar, S.C. (1990). "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiecny," *Journal of Econometrics*. 46(1/2): 201-212.
Kumbhakar, S.C. (1991). "The Measurement and Decomposition of Cost-Inefficiency: The Translog Cost System", *Oxford Economic Papers*. 43(4): 667-683.
Kumbhakar, S.C. (1996). "Estimation of Cost Efficiency with Heteroskedasticity: An Application to Electricity Utilities", *The Statistician*. 45(3), 319-335.
Kumbhakar, S.C. (1997). "Modeling Allocative Inefficiency in a Translog Cost Function and Cost Share Equations: An Exact Relationship", *Journal of Econometrics*. 76: 351-356.
Kummbhakar, S.C. and C.A.K. Lovell (2000). "Stochastic Frontier Analysis". (Cambridge: Cambridge University Press.)

Kumbhakar, S.C. (1988). “On the Estimation of Technical and Allocative Inefficiency Using Stochastic Frontier Functions: The Case of the US Class 1 Railroads,” *International Economic Review*. 29(4):727-743.
Kumbhakar S.C., S. Ghosh, and J.T. McGuckin (1991). “A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms,” *Journal of Business and Economic Statistics*. 9:279–286
Lange, I. and A.S. Bellas (2007). "The 1990 Clean Air Act and the Implicit Price of Sulfur in Coal," *The B.E. Journal of Economic Analysis & Policy*. 7(1): Article 41.
Lee, L.F. (1983). “A Test for Distributional Assumptions for the Stochastic Frontier Functions,” *Journal of Econometrics*. 22(3): 245-267.
Lee, M. (2002). “The Effect of Sulfur Regulation on the US Electric Power Industry: A Generalized Cost Approach,” *Energy Economics*, 24:491-508.
Lee, M. (2005). “The Shadow Price of Substitutable Sulfur In the US Electric Power Plant: A Distance Function Approach,” *Journal of Environmental Management.* 77: 104-110.
Leibenstein, H. (1966) “Allocative efficiency vs X efficiency”, *American Economic Review*. 56: 392-415
Mann, H. B., and D. R. Whitney (1947). “On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other,” *Annals of Mathematical Statistics.* 18: 50-60.
Meeusen W. and J. van den Broeck (1977). “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error,” *International Economic Review.* 18: 435–444.
Morrison P., J. Catherine, W.E. Johnston and G.A.G. Frengley (2000). “Efficiency in New Zealand Sheep and Beef Farming: The Impacts of Regulatory Reform,” *The Review of Economics and Statistics*. 82(2): 325-337.
Olley, G.S. and A. Pakes (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*. 64(6): 1263-1297.
Park, H.M. (2006), “Univariate Analysis and Normality Test Using SAS, STATA and SPSS”. The Trustees of Indiana University. Available online at www.indiana.edu/~statmath.
Paskura, C.A. (2006). “Decomposing Electric Power Plant Emissions Within a Joint Production Framework,” *Energy Economics*. 28: 26-43.
Pearson, A. V., and H.O. Hartley (1972). *Biometrica Tables for Statisticians, Vol 2*, Cambridge, England, Cambridge University Press.
Piacenza, M. (2006). “Regulatory Contracts and Cost Efficiency: Stochastic Frontier Evidence from the Italian Local Public Transport,” *Journal of Productivity Analysis.* 25: 257-277.
Piacenza, M. (2006). “Regulatory Contracts and Cost Efficiency: Stochastic Frontier Evidence from the Italian Local Public Transport,” *HERMES.* Working Paper *2.*
Piesse, J. and C. Thirtle (2000). “A Stochastic Frontier Approach to Firm Level Efficiency, Technological Change, and Productivity during the Early Transition in Hungary”, *Journal of Comparative Econoimcs.* 28: 473-501.
Pitt, M. and L. Lee (1981). “The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry,” *Journal of Development Economics*. 9: 43-64.
Pittman, R. (1983). “Multilateral Productivity Comparisons with Undesirable Outputs,” *The Economic Journal*. 93(372): 883-891.
Pollitt, M. (1996). “Ownership and Efficiency in Nuclear Power Production,” *Oxford Economic Papers*. 48(2): 342-360.
Porter, M. (1991). "America”s Green Strategy," *Scientific American*, April (264): 168.

Porter, M.E. and C. van der Linde (1995). “Toward a new conception of the environment competitiveness relationship,” *Journal of Economic Perspectives*. 9(4): 97–118.
Reifschneider, D. and R. Stevenson (1991). “Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency”, *International Economic Review*. 32(3): 715-723.
Rezek, J. and B.F. Blair (2005). “Abatement Cost Heterogeneity in Phase I Electric Utilities,” *Contemporary Economic Policy*. 23(2): 324-340.
Richmond, J. (1974). “Estimating the Efficiency of Production,” *International Economic Review*. 15(2): 515-521.
Rossi, M (Royston, J. P. (1983). “Some techniques for assessing multivariate normality based on the Shapiro-Wilk *W,” Applied Statistic*s. 32: 121–133.
Rungsuriyawiboon, S. and T. Coelli (2004). “Regulatory Reform and Economic Performance in US Electricity Generation,” *Centre for Efficiency and Productivity Analyisis,* Working Paper No: 06/2004.
Sarkadi, K. (1975). “The Consistency of the Shapiro-Francia Test,” *Biometrika*. 43: 456-458.
Schmalensee, R, P.L. Joskow, A.D. Ellerman, J.P. Montero, and E.M. Bailey (1998). "An Interim Evaluation of Sulfur Dioxide Emissions Trading," *Journal of Economic Perspectives*, 12(3): 53-68.
Seiford, L.M., & R.M. Thrall (1990) "Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis," *Journal of Econometrics* vol. 46: pp. 7-38.
Seitz, W.D. (1971). “Productive Efficiency in the Steam-Electric Generating Industry,” *Journal of Political Economy*. 79(4): 878-886.
Shapiro, S. S. and R.S. Francia (1972). “An Approximate Analysis of Variance Test for Normality,” *Journal of the American Statistical Association*. 67: 215–216.
Shapiro, S. S. and M.B. Wilk (1965). “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*. 52: 591–611.
Shephard, R.W. (1953). *Cost and Production Functions*, Princeton: Princeton University Press.
Shephard, R.W. (1970). *Theory of Cost and Production Functions.* Princeton: Princeton University Press.
Shih, J-S , W. Harrington, W. A. Pizer, and K. Gillingham (2004). “Economies of Scale and Technical Efficiency in Community Water Systems,” *Resources for the Future*, Discussion Paper 04-15.
Simon, H. (1955). “A Behavioral Model of Rational Choice,” *Quarterly Journal of Economics.* 6: 99-118.
Simon, H. (1957), *Models of Man*. NY: John Wiley.
Stern, N., 2006, *Review on the Economics of Climate Change*, H.M. Treasury, UK, October. Available online at: http://www.sternreview.org.uk.
Trainer, F. E. (T.), (2005), The Simpler Stevenson, R.E. (1980). “Likelihood Functions for Generalized Stochastic Frontier Estimation,” *Journal of Econometrics*. 13(1): 57-66.
Stigler, G. (1976). “The Xistence of X-efficiency,” *American Economic Review*. 66: 213–6.
Suer, B. (1995). “Technological Change and Productivity in the UK Food, Drink and Tobacco Industries,” *International Journal of Manpower.* 16(1): 32-45.
Swinton, J. R. (1998). "At What Cost Do We Reduce Pollution? Shadow Prices of Reduction,” *The Energy Journal.* 19: 63-83.
Thanassoulis, E. (2001). *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software* , Boston: Kluwer Academic Publishers.

Timmer, C.P. (1971). "Using a Probabalistic Frontier Production Function to Measure Technical Efficiency," *Journal of Political Economy*. 79(4): 776-794.
Tuthill, L. (2008). "Interfuel Substitution and Technical Change in the US Electricity Generating Industry Under the Tradable Sulphur Allowance Scheme: 1990-2004," OIES Working Paper, EV 40.
Vardanyan, M. and D-W Noh (2006). "Approximating Pollution Abatement Costs via Alternative Specifications of a Multi-Output Production Technology of the US Electric Utility Industry," *Journal of Environmental Management.* 80(2): 177-190.
Wang, H-J (2002). "Heteroscedasticity and Non-Monotonic Efficiency Effects of A Stochastic Frontier Model," *Journal of Productivity Analysis*. 18:241-253.
Wang, H-J (2003). "A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan," *Journal of Business and Economic Statistics*. 21(3): 406-419.
Wang, H-J and P. Schmidt (2002). "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency," *Journal of Productivity Analysis.* 18(2): 129-144.
"What Statistical Analysis Should I Use? Statistical Analyses Using Stata." UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu/stat/stata/whatstat/whatstat.htm (accessed January 24, 2008).
Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics* 1: 80-83.
Winsten, C.B. (1957) "Discussion on Mr. Farrell"s Paper," *Journal of the Royal Statistical Society*, Series A, General, 120, Part 3, 282-284.
Yaisawarng, S. and J.D. Klein (1994). "The Effects of Sulfur Dioxide Controls on Productivity Changein the US Power Industry," *The Review of Economics and Statistics*. 76(3): 447-460.