SO₂ policy and input substitution under spatial monopoly

Shelby Gerking a, Stephen F. Hamilton b, *

a University of Central Florida, Orlando, FL 32816, United States
b California Polytechnic State University, San Luis Obispo, CA 93407, United States

Following the U.S. Clean Air Act Amendments of 1990, electric utilities dramatically increased their utilization of low-sulfur coal from the Powder River Basin (PRB). Recent studies indicate that railroads hauling PRB coal exercise a substantial degree of market power and that relative price changes in the mining and transportation sectors were contributing factors to the observed pattern of input substitution. This paper asks the related question: To what extent does more stringent SO₂ policy stimulate input substitution from high-sulfur coal to low-sulfur coal when railroads hauling low-sulfur coal exercise spatial monopoly power? The question underpins the effectiveness of incentive-based environmental policies given the essential role of market performance in input, output, and abatement markets in determining the social cost of regulation. Our analysis indicates that environmental regulation leads to negligible input substitution effects when clean and dirty inputs are highly substitutable and the clean input market is mediated by a spatial monopolist.

1. Introduction

Cap and trade systems for pollution control are much heralded, and rightly so, for their ability to marshal market forces to achieve pollution reductions at the lowest possible cost. An important aspect governing the performance of market-based approaches to environmental problems is the ability of pollution markets to align private incentives in the affiliated output, input and abatement markets that intersect with the policy. There is general consensus that the landmark success of the U.S. Clean Air Act Amendments of 1990, which achieved U.S. air quality objectives at a fraction of the anticipated
cost (U.S. Department of Energy, 2000), was driven largely by input substitution among U.S. electric utilities from high-sulfur coal to low-sulfur coal (see, e.g., Carlson et al., 2000). The substantial degree of input substitution that occurred between these coal deposits contributed to low SO₂ allowance prices, thereby defraying costly social adjustments that would otherwise have taken place in electricity and pollution abatement markets. Over the period 1990–2002, coal production from high-sulfur deposits in the Illinois Basin (Illinois, Indiana, and West Kentucky) declined by 42%, while low-sulfur coal production from the Powder River Basin (PRB) of Wyoming and Montana doubled and the number of utilities burning PRB coal more than tripled. This paper examines the potential role of the cap and trade system for SO₂ emission allowances in driving the observed pattern of input substitution in U.S. coal markets.

In principle, the creation of a pollution allowance market should stimulate a coordinated set of changes in the markets intersecting with environmental policy. To support a socially optimal resource allocation, allowance prices must be derived from an optimal set of adjustments in output markets (electricity), abatement markets (e.g., “scrubbers”), and input markets (high-sulfur coal, low-sulfur coal, natural gas, oil) that serve to equalize the marginal returns across activities. Because empirical studies of energy markets generally find household electricity demand to be highly price inelastic (Reiss and White, 2005), the ability to achieve environmental objectives with minimal disruption in consumer markets is facilitated by elastic supply conditions in abatement and input markets. For example, if coal deposits with different sulfur content are perfectly substitutable in electricity production and available at equal factor prices, SO₂ reductions could be met largely through input substitution, without the need for substantial investments in abatement equipment or output adjustments. In the event that the supply of low-sulfur coal was perfectly elastic, environmental objectives could be obtained at zero cost.

The main point of this paper is that spatial market power over low-sulfur coal deliveries to electric plants forecloses input substitution possibilities that would otherwise occur in response to SO₂ regulation. While it is evident that the effectiveness of environmental policy depends on the presence of market power in the markets for alternative inputs, our analysis highlights the pernicious effects of input market imperfection when deliveries are mediated by a spatial monopolist. Indeed, in the case of perfectly substitutable coals and perfectly elastic supply by mines, we show SO₂ policy to be completely ineffective in stimulating input substitution between high- and low-sulfur coals. The direct implication of this finding is that private incentives introduced by a market-based policy for SO₂ emissions appear to be largely confined to electricity markets, abatement markets, and input markets for alternative fuels besides low-sulfur coal. The indirect implication of this finding is that much of the apparent success of the U.S. Clean Air Act Amendments in achieving environmental objectives at lower-than-anticipated cost was due to input substitution from high-sulfur coal to low-sulfur coal among electric utilities that had little, if anything, to do with environmental policy. Instead, it appears that the timing of SO₂ policy was simply serendipitous. As Ellerman and Montero (1998) observe, the U.S. Clean Air Act Amendments of 1990 coincided with a period of substantial declines in the real prices of mine mouth PRB coal and rail transportation. These features, unlike SO₂ policy, are capable of stimulating the observed, substantial increase in aggregate deliveries of low-sulfur coal by a railroad monopoly market.

We frame our observations around a model of spatial monopoly power in the market for low-sulfur coal. The potential for railroads to exercise market power in the low-sulfur coal market is an important consideration, because virtually all low-sulfur coal in the U.S. is hauled eastward from Wyoming by a handful of railroads serving PRB mines.₁ Among railroads serving these lines, two firms – Burlington Northern Santa Fe (BNSF) and Union Pacific (UP) – currently initiate all transportation of PRB coal.₂ The potential for railroads to exercise spatial market power has been recognized since at least the case of Standard Oil (see Granitz and Klein, 1996), and evidence exists that railroads hauling PRB coal out of

₁ Alternative modes of coal transportation out of the PRB either are not cost-effective (e.g., trucking) or else do not exist (e.g., barges and coal slurry pipelines).
₂ The Chicago and Northwestern Railroad, which entered the Wyoming coal transportation market in the early 1980s, no longer serves the PRB. Also, the BNSF and UP do not always complete deliveries to all power plants because coal is frequently transshipped via other lines.
Wyoming indeed exercise a considerable degree of market power (Wolak and Kolstad, 1988; Busse and Keohane, 2007).

A large literature has emerged to examine the potential for imperfections in permit markets to erode the gains of market-based environmental policies (see, e.g., Hahn, 1984; Joskow et al., 1998; Montero, 1999, 2002). In light of this fact, it is somewhat surprising to note that little attention has been focused on market performance in the markets for abatement equipment and alternative inputs that intersect with cap-and-trade policies. Our analysis bridges this gap by examining the essential role of transportation markets for alternative inputs in mediating the performance of market-based environmental policies.

The remainder of the paper is organized as follows. In the next section, we develop a simple model of spatial intermediation by a monopoly railroad in the low-sulfur coal market. In Section 3, we calculate the effect of SO2 policy on the aggregate quantity of PRB coal delivered to electric utilities in the special case of perfectly substitutable coals. In Section 4, we conclude with some brief comments on the case of differentiated coal deposits and outline directions for further research.

2. The model

We consider a dominant firm-competitive fringe market structure in which utilities either buy low-sulfur coal from a monopoly railroad or purchase high-sulfur coal from a competitive industry. To focus the model on the implications of SO2 policy for the low-sulfur coal market, we limit the fuel portfolio available to utilities to two potential fuel inputs – low-sulfur coal and high-sulfur coal – and suppress the possibility of electric plants meeting the regulation through other compliance options. Implicitly, we are assuming that coal-fired utilities cannot burn natural gas or oil, and that switching from high-sulfur coal to low-sulfur coal is more cost-effective than avoiding fuel-switching altogether by installing post-combustion abatement equipment (“scrubbers”).3 Although slight differences exist in practice in the energy and sulfur content of each type of coal, depending on the particular mine from which the coal is sourced, we simplify the analysis by treating deposits of each type of coal as uniform in composition.

In the Midwestern U.S., high-sulfur coal is generally procured from deposits in reasonably close proximity to individual utilities. Relative to the market for low-sulfur coal, freight rates make up a smaller percentage of delivered prices for high-sulfur coal, resulting in delivered prices that do not vary substantially across locations in practice. To focus the model on the low-sulfur coal market, we assume that high-sulfur coal is ubiquitously available at a constant delivered price of \( p_h \). Monopoly prices for low-sulfur coal, in contrast, vary over space according to proximity of the utility demanding low-sulfur coal to the source mine.

2.1. Utility demand for PRB coal

Demand facing railroads for deliveries of low-sulfur coal is derived from the profit-maximization problem of electric utilities who seek to produce energy from alternative fuel inputs. To clarify our observations on the effect of a spatially intermediated input market, we assume the electricity market is competitive and utilities are homogeneous in all respects apart from their location in space.4 Utilities are arrayed spatially along the rail line and face different freight rates, and hence different delivered prices, for PRB coal.

The problem facing a utility at distance \( x \) is to select a quantity of PRB coal, \( q_L(x) \), and a quantity of high-sulfur coal, \( q_h(x) \), to maximize profit subject to environmental policy on SO2 emissions. Let \( \sigma \) denotes the SO2 emissions coefficient for high-sulfur coal, so that SO2 emissions for a utility at distance

---

3 As Carlson et al. (2000) observe, input substitution at U.S. electric utilities from high-sulfur coal to low-sulfur coal has proven to be substantially more cost-effective as a method of emissions control than the use of post-combustion abatement technology.

4 The implications of the model regarding incentives for fuel substitution among cost-minimizing electric utilities are qualitatively similar to those which would arise in a regulated monopoly electricity market. Under regulated monopoly, the ability of utilities to pass-through cost changes into electricity rates does alter the output effect of SO2 regulation in electricity markets. Extending the analysis to the electricity sector, which would complicate the model by adding an independently regulated consumer pricing structure, would provide a greater apparatus to sift through without fundamentally altering incentives for fuel substitution among utilities seeking to minimize fuel cost.
x can be defined as $e(x) = \sigma q_h(x)$. Under cap-and-trade policy, each utility is given an initial endowment of SO2 allowances (possibly zero) and must purchase an SO2 allowance for each unit of emissions above this level at the permit price, denoted $\lambda$.

The market price of an SO2 allowance is taken as given by each firm, but is determined endogenously by the size of the regulatory cap on emissions in the market. We represent the spatial market for coal inputs by arraying all electric utilities on a unit line segment and denote total SO2 emissions as $E = \int_0^1 e(x)dx = \sigma \int_0^1 q_h(x)dx$. The sum of all SO2 allowances held by utilities must meet the regulated level of $E$ under the emissions cap, which we denote by $E_0$.

Let $p_e$ denote the electricity price and $p(x)$ denote the delivered price of PRB coal at distance $x$. A utility with an initial endowment of $e_0$ allowances maximizes profits of

$$\pi(x) = p_x f_l(x; q_l(x), q_h(x)) - \sigma p_l(x) q_l(x) - p_h q_h(x) + \lambda(e_0 - e(x)),$$

subject to non-negativity constraints on the use of low-sulfur coal, $q_l(x) \geq 0$, and high-sulfur coal, $q_h(x) \geq 0$, and the link between high-sulfur coal use and SO2 emissions, $e(x) = \sigma q_h(x)$. An electric utility that is a net seller of emission allowances satisfies $e_0 - e(x) > 0$, and a utility that is a net purchaser of emission allowances satisfies $e_0 - e(x) < 0$.

The first-order necessary conditions for a maximum are

$$\Gamma_l \equiv p_e f_l(.) - p_l(0) \leq 0, \quad \Gamma_q q_l(0) = 0;$$

$$\Gamma_h \equiv p_e f_h(.) - p_h - \sigma \lambda \leq 0, \quad \Gamma_h q_h(0) = 0;$$

where $f_l(.)$ and $f_h(.)$ denote the marginal product of each type of coal in electricity production. Let $p_s = p_h + \sigma \lambda$ denote the “effective price” of high-sulfur fuel, which is the gross price of an input of high-sulfur coal inclusive of its implicit SO2 allowance requirement. Notice that the effective price of high-sulfur coal is independent of both distance and the initial allocation of SO2 allowances; hence the input mix at each utility between low-sulfur coal and high-sulfur coal depends only on the relative prices, $p_l(x)$ and $p_e$.

At an interior solution, the input mix for each utility equates its marginal rate of technical substitution with the ratio of prices, $p_l(x)/p_e$. In the event that delivered prices for low-sulfur coal rise over distance from the PRB, utilities along the rail line faced with a constant effective price of high-sulfur coal smoothly substitute away from low-sulfur coal and towards high-sulfur coal over space to an extent that depends on the elasticity of substitution between these fuels.

Depending on the value of the elasticity of substitution, a corner solution is also possible. Prevailing evidence suggests that low- and high-sulfur coal are highly substitutable inputs at electric plants (Gerking and Hamilton, 2008), implying a large elasticity of substitution. In the case of perfect substitutes, normalizing units for the energy (BTU) content of each type of coal, we can define $f_l(.) = f_h(.) = f(\cdot)$, and adjust units accordingly. By inspection of expressions (1) and (2), the conditional (inverse) demand for low-sulfur coal in the case of perfectly substitutable fuels is given by $p(q_l(x))$ for $p(q_l(x)) \leq p_e$ and zero otherwise.

Let $D^l(p_e, p_l(x), p_s)$ and $D^h(p_e, p_l(x), p_s)$ denote the demand functions for low- and high-sulfur coal that solve conditions (1) and (2) for the utility at location $x$. The demand for each type of coal at the electric plant depends on the electricity price, the delivered price of PRB coal and the effective price of high-sulfur coal.

2.2. The rail sector

Our model extends the framework of Greenhut and Ohta (1972) to consider spatial market power by a monopoly railroad over an endogenously determined service region. The railroad purchases low-sulfur coal at a constant price of $w$ per unit from competitive mines located at the origin of a rail line and delivers it at a marginal cost of $r$ per unit of distance ($x$) to identical electric utilities, which are assumed for analytic convenience to be uniformly distributed along a rail line of unit length. The total
cost of delivering coal to a utility at distance $x$ is $txD'(p_e, p_l(x), p_s)$, and the total cost of procuring and delivering an aggregate quantity $Q = \int_0^N D'(p_e, p_l(x), p_s)dx$ of PRB coal in the market accordingly is

$$c(Q) = \int_0^N (w + tx)D'(p_e, p_l(x), p_s)dx,$$

where $N \leq 1$ is the extensive margin of service. Fixed costs, which are necessary to justify the existence of railroad market power, play no role in the analysis and are consequently omitted.

The railroad’s problem is to select the number of utilities to serve, $N$, and a delivered price $p_l(x)$ for each utility in the service region $x \in [0, N^+]$. The railroad’s profit is

$$\pi(w, t, p_s) = \int_0^N (p_l(x) - tx - w)D'(p_e, p_l(x), p_s)dx.$$  \tag{3}$$

The first-order necessary conditions for a profit maximum are completely characterized by the Euler equation,

$$D'(p_e, p_l(x), p_s) + (p_l(x) - tx - w)D'(p_e, p_l(x), p_s) = 0, \quad \text{for } x \in [0, n],$$  \tag{4}$$
a boundary condition on interior pricing,

$$D'(p_e, p_l(x), p_s) = q'(p_s), \quad \text{for } x \in [n, N],$$  \tag{5}$$
and the transversality condition

$$(p_l(N) - tN - w)D'(p_e, p_l(N), p_s) = 0.$$  \tag{6}$$

where $0 < q'(p_s)$ is a fixed quantity delivered to locations $n \leq x \leq N$.

Conditions (4)–(6) have straightforward interpretations. Condition (4) defines the optimal monopoly price for delivery of PRB coal to a utility located at distance $x$. It states that the optimal monopoly price at distance $x$ be set so that the percent mark-up of delivered price over marginal cost, $tx + w$, is equal to the reciprocal of the demand elasticity. Delivered prices are higher for more distant utilities than for utilities in closer proximity to the source of PRB coal because the marginal cost of delivery rises in $x$, whereas price-cost margins narrow for more distant utilities as demand becomes more elastic at higher delivered prices. As delivered prices rise over space, utility demand for PRB coal decreases until one of two things happens: (i) a minimum shipment size is hit that satisfies Eq. (5); or (ii) demand falls smoothly to zero until the freight rate, $p_l(N) - w$, equates with unit transportation cost, $tN$, at the distance defined by Eq. (6). Eq. (6) defines the location of the critical utility, $x = N$, beyond which shipments to more distant utilities no longer contribute positively to railroad profit. Because a monopoly railroad can spatially price discriminate, railroad service continues until the delivered price is driven to marginal cost and profit falls to zero on the extensive margin of service.

There are two reasons why a minimum shipment size may bind on deliveries over space in Eq. (5). First, a minimum shipment quantity may emerge due to integer constraints, for instance when it is not practical to deliver a fraction of a rail car of coal. Second, the non-negativity constraint in utility demand Eq. (2) may bind, in which case a monopoly railroad might maintain delivery prices at the corner as deliveries continue over space. For example, when low- and high-sulfur coal are perfect substitutes, the delivered price for low-sulfur coal rises over distance to $p_s$ and then remains constant at that level until profits fall to zero on the extensive margin of service (see below). In either case, freight prices set by a monopoly railroad would rise over space in an interior pricing region over distances $0 < x < n$, and then remain constant at a level necessary to maintain demand over the remaining distances in the service region, $n \leq x \leq N$. Hereafter, we refer to the region of unconstrained monopoly pricing as “region I” and the region of constrained monopoly pricing as “region II”.

Consider the case of monopoly freight rates, $f(x) = p_l(x) - w$, in the case of perfectly substitutable fuels. As in a conventional dominant firm-competitive fringe model with perfectly elastic supply, demand for low-sulfur coal becomes horizontal at the effective price of high-sulfur coal, $p_s$, generating

---

6 Gerking and Hamilton (2008) classify shipments of less than 50 rail cars of PRB coal as intermittent shipments used for “test burns”, which suggests deliveries do not smoothly fall to zero along the rail line in practice.
a discontinuity in the marginal revenue schedule for low-sulfur coal. Equilibrium freight rates in region I are determined by the intersection of marginal revenue and marginal transportation cost for PRB coal. Freight rates rise smoothly over distance in this region until the delivered price of PRB coal rises to \( p_s \). At this point, the marginal revenue function facing the railroad is discontinuous; the railroad can continue serving utilities at further distances, but only at a delivered price at or below \( p_s \). Demand and marginal revenue are horizontal at price \( p_s \) in region II, because utilities have perfectly elastic demand for low-sulfur coal at this price. The ability to spatially price discriminate ensures that the railroad continues serving utilities in region II: At distance \( x = n \), a freight rate of \( p_l(N) = p_s \) more than compensates the railroad for the marginal cost of transporting it, \( p_l(N) - tN - w > 0 \), so that continuing service to utilities at distances beyond \( x = n \) contributes positively to profit. Service continues to region II utilities until transportation costs rise at location \( x = N \) to meet the terminal condition (6).

The second-order necessary conditions for a maximum are the Legendre condition,

\[
2D_l(p_e, p_l(x), p_s) + (p_l(x) - tx - w)D_{ll}^l(p_e, p_l(x), p_s) \leq 0,
\]

and the endpoint condition (see Caputo and Wilen, 1995),

\[
\Phi = -tD_l(p_e, p_l(N), p_s) \leq 0.
\]  

Throughout the remainder of the paper, we assume these conditions are strictly met.

Let \( p_l^r(x, t, w, p_e, p_s) \), \( n^r(t, w, p_e, p_s) \), and \( N^r(t, w, p_e, p_s) \) denote the solution to (4)–(6). With exogenous \( p_s \), the comparative statics for the delivery schedule for region I and region II utilities can then be derived from (4) and (5). Consider, first, the delivered price schedule over distance in region I. Dropping arguments for notational convenience, use of the implicit function theorem on (4) yields

\[
\frac{\partial p_l^-}{\partial x} = -\frac{tD_l}{2D_l + (p_l^- - tx - w)D_{ll}^l} > 0,
\]

where the inequality holds by condition (7). As the marginal cost of delivering a unit of coal increases over distance, the delivered price of PRB coal rises.

The outcome for the spatial pricing is as follows. The optimal freight schedule,

\[
f^*(x, t, w, p_e, p_s) = p_l^r(x, t, w, p_e, p_s) - w,
\]

rises gradually over distance by (9) until the region I boundary condition is met in Eq. (5), at which point deliveries continue at constant prices until a distance is reached where the freight rate equates with unit transportation cost, \( f^*(N^r(t, w, p_e, p_s), t, w, p_e, p_s) = tN^r(t, w, p_e, p_s) \), on the extensive margin of service. At this point, deliveries cease.

Fig. 1 compares the freight schedule under monopoly price discrimination to that which would emerge in a competitive transportation sector. Under competition, the freight rate, \( f^c(x) = tx \) rises

![Fig. 1. Freight schedule over distance under monopoly and competition.](image-url)
smoothly from zero at a rate of \( t \) over distance until a price is reached where the delivered price meets the terminal condition (6). Under monopoly pricing, the freight rate is piecewise concave, beginning at a higher level, and subsequently rising more slowly over distance. By Eq. (6), the terminal distance, \( N' \), must coincide in each case.

How do changes in SO2 policy alter the aggregate utilization of PRB coal when input substitution is mediated by a monopoly railroad sector? The effect of a binding regulatory cap on SO2 emissions is to raise the effective price of high-sulfur coal, which in turn causes utilities to change the cost-minimizing input mix in Eqs. (1) and (2) and to decrease electricity supply. Much of the effectiveness of environmental policy in achieving environmental objectives with minimal consumer harm depends on the ability of utilities to substitute away from high-sulfur coal and towards cleaner-burning inputs of low-sulfur coal to mitigate the decrease in energy supply as allowance prices rise in response to the policy.

We denote the aggregate quantity of PRB coal delivered by

\[
Q^* = -\int_{n}^{N} D^i(p_e, p_s^i(x), p_s)dx + (N - n^*)q^s(p_s). \tag{11}
\]

Differentiating Eq. (11) with respect to the allowance price and making the substitution \( D^i(p_e, p_s^i(n), p_s) = q^s(p_s) \) from Eq. (5) yields,

\[
\frac{\partial Q^*}{\partial \lambda} = \sigma \left[ \int_{n}^{N} D_s^i(p_e, p_s^i(x), p_s)dx + q^s(p_s) \frac{\partial N^g}{\partial p_s} + (N - n^*) \frac{\partial q^s(p_s)}{\partial p_s} \right]. \tag{12}
\]

In general, the ability of SO2 policy to stimulate input substitution among utilities towards low-sulfur coal depends on three effects in Eq. (12). On the intensive margin of service, the rise in the effective price of high-sulfur coal increases input demand for low-sulfur coal among region I utilities, which changes the aggregate quantity of PRB coal delivered to region I utilities by \( \sigma \int_{n}^{N} D^i_s(p_e, p_s^i(x), p_s)dx \) units. On the extensive margin of railroad service, the rise in the effective price of high-sulfur coal leads to two additional effects. First, there is a market expansion effect, which represents the outward shift in the extensive margin of service to more distant utilities. The expansion effect is given by \( \sigma q^s(p_s)\partial N^g/\partial p_s \) in (12). Because the extensive margin of railroad service is independent of market structure (see Eq. (6)), the expansion effect is identical under monopoly and competitive freight pricing. Second, under monopoly, there is a contraction effect, \( \sigma(N' - n^*)\partial q^s(p_s)/\partial p_s \), which represents a decrease in the equilibrium quantity of PRB coal delivered to region II utilities. The contraction effect arises when the minimum shipment size in Eq. (5) decreases in response to an increase in the equilibrium price of low-sulfur coal charged to region II utilities.

The relative magnitude of the expansion and contraction effects depends on the manner in which the minimum shipment constraint enters Eq. (5). If the minimum shipment quantity is independent of the delivered price of low-sulfur coal, for instance if the minimum shipment quantity is 50 rail cars irrespective of the delivered price, then there is no contraction effect. Alternatively, as we show below in the case of perfectly substitutable inputs, if the minimum shipment quantity is selected to prevent utilities from switching to high-sulfur coal in the PRB service territory, then the contraction effect operates against the expansion effect to limit aggregate deliveries of low-sulfur coal to region II utilities.

Making use of the implicit function theorem on Eq. (4), the effect of a marginal change in the price of SO2 allowances on delivered prices of low-sulfur coal in region I is

\[
\frac{\partial p_s^i(x)}{\partial \lambda} = \sigma \frac{\partial p_s^i(x)}{\partial p_s} = \frac{-\sigma(D^i_s + (p^i_s(x) - \tau x - w)D^i_s)}{2D^i_s + (p^i_s(x) - \tau x - w)D^i_s}. \tag{13}
\]

which takes the sign of \( D^i_s + (p^i_s(x) - \tau x - w)D^i_s \) by (7). Eq. (13) states that the delivered price of PRB coal to each utility in region I rises in response to a marginal increase in the SO2 allowance price provided that the outward shift in PRB demand \( (D^i_s > 0) \) is not coupled with a sufficiently strong
rotation effect that makes PRB demand more price elastic. Making use of (13), the effect of an allowance price increase on the quantity sold to each utility in region I is

$$\frac{\partial D_t^{\lambda}}{\partial \lambda} = \sigma \left( \frac{\partial p_s}{\partial \lambda} + D_t \left( \frac{\partial p_t^1(x)}{\partial \lambda} \right) \right) = \frac{\sigma \left[ \left( D_t^1 + (p_t^1(x) - tx - w)D_t^2 - D_t^3 \right) - 2D_t^1 + (p_t^1(x) - tx - w)D_t^3 \right]}{2D_t^1 + (p_t^1(x) - tx - w)D_t^3}$$

In general, this equation can take either sign; however, in the case of linear demand and imperfect substitutes, the delivered quantity increases for each utility in region I.

To derive the effect of a change in SO2 policy on the extensive margin served by the railroads, substitute $p_t^1(x, t, w, p_e, p_s)$ and $N^*(t, w, p_e, p_s)$ into Eq. (6) to get

$$\left[ \frac{\partial N^*}{\partial \lambda} - tN^* - w \right] D_t^3(p_e, p_t^1(N^*), t, w, p_e, p_s) = 0$$

Implicitly differentiating this equation and making use of (4) gives

$$\frac{\partial N^*}{\partial \lambda} = \frac{\sigma (p_t^1(N^*) - tN^* - w)D_t^3(p_e, p_t^1(N^*), p_s)}{tD(p_e, p_t^1(N^*), p_s)} > 0.$$

An increase in the price of SO2 allowances serves to expand the railroad’s service region.

The foregoing analysis isolates the input substitution effect of SO2 regulation in the spatially intermediated market for low-sulfur coal by suppressing abatement market and electricity market effects. In response to SO2 policy, the equilibrium increase in the effective price of high-sulfur coal integrates each of these responses. Following a rise in the input price of high-sulfur coal, utilities burning this fuel respond by supplying a lower quantity of electricity and engaging in greater abatement activities. Consumer electricity prices rise in response, which in turn would stimulate the use of low-sulfur coal in Eq. (1) in addition to the direct input substitution motivation resulting from changes in relative input prices.7

In the next section, we characterize the input substitution effect of SO2 regulation in the case of perfectly substitutable fuels. This is an important case, because the elasticity of substitution between high- and low-sulfur coal is likely to be quite high. The reason is that generating units at most U.S. power plants are engineered to burn different deposits of coal more or less intensively as relative fuel prices change. Coals obtained from different mines are commonly mixed with each other in an increasingly diversified fuel portfolio, and blending of low- and high-sulfur coals has occurred at many power plants since the early 1990s.8

### 3. Perfect Substitutes

To examine the role of spatial market intermediation on the performance of the low-sulfur coal market, consider the case in which low-sulfur coal and high-sulfur coal are perfect substitutes. Normalizing units to align the BTU content of each type of coal, we consider differences between coal deposits as arising only from differences in sulfur content, so that $f_i(.) = f_h(.) = f(.)$ in expressions (1) and (2).

For the case of perfectly substitutable fuels, it is analytically convenient to recast the profit-maximization problem of a monopoly railroad as a quantity choice problem. From expressions (1) and (2), the conditional (inverse) demand for low-sulfur coal facing the railroad is given by $p(q_l(x))$ for $p(q_l(x)) \leq p_s$ and zero otherwise. Railroad profit, accordingly, is

$$\pi(w, t, p_s) = \int_0^w (p(q_l(x)) - tx - w)q_l(x)dx + \int_w^N (p_s - tx - w)q^2dx,$$

where $q^2$ solves $p(q^2) = p_s$ in utility demand.

7 The magnitude of the energy price effect would differ according to whether the price adjustment in electricity markets occurred through a decrease in electricity supply (as modeled here) or according to pass-through relationships in a regulated monopoly electricity market.

8 Roughly 25% of coal-fired generating units burn a blend of PRB and eastern coal and blending generally does not occur at utilities in close proximity to PRB mines (Gerking and Hamilton, 2008).
The first-order necessary conditions for a profit maximum are
\[ p(q_l(x)) + q_l(x) p'(q_l(x)) - t x - w = 0, \quad \text{for} \ x \in [0, n], \tag{4'} \]
the region I boundary condition,
\[ p(q_l(x)) = p(q^d) = p_s, \quad \text{for} \ x \in [n, N], \tag{5'} \]
and the transversality condition,
\[ (p_s - tN - w)q^d = 0. \tag{6'} \]

Expressions (4′)–(6′) are analogous to expressions (4)–(6). The delivery schedule is set to equate marginal revenue with marginal cost for region I utilities in (4′). Delivered prices rise smoothly over this initial portion of the service region until some distance, \( n^* \), is reached where \( p(q_l(n^*)) = p_s \) in (5′), whereupon prices remain at this level – preempting utilities from switching into high-sulfur coal – until a distance is reached on the extensive margin of service where profit is driven to zero. At the extensive margin in (6′), railroad deliveries of PRB coal cease, and utilities at further distances on the line segment burn high-sulfur coal.

In equilibrium, the optimal freight schedule has two distinct spatial regions. In region I, utilities purchase a sufficiently large quantity of PRB coal that interior monopoly prices obtain, \( p(q_l(x)) \leq p_s \); however, profit is still positive at this utility because of the discontinuity in the marginal revenue schedule.\(^9\) Because spatial price discrimination is possible, the monopoly railroad continues to serve more distant region II utilities, but does so under the binding constraint that \( p(q^d) = p_s \). Region II deliveries continue to the distance \( N^* \), where \( p_s \) equates with marginal transportation cost, \( tN^* + w \), and profit is fully dissipated in (6′).

Next consider the effect of a change in the price of SO2 allowances on railroad shipments of low-sulfur coal. The aggregate quantity of PRB coal delivered by the railroad is
\[ Q^*_l = \int_0^{n^*} q^*_l(x) dx + (N^* - n^*)q^d(p_s). \tag{11′} \]
Differentiating Eq. (11′) with respect to the allowance price \( \lambda \) and factoring terms yields,
\[ \frac{\partial Q^*_l}{\partial \lambda} = \sigma \left[ \int_0^{n^*} \frac{\partial q^*_l}{\partial p_s}(x) dx + q^*_l(n) \frac{\partial N^*}{\partial p_s} + (N^* - n^*) \frac{\partial q^d(p_s)}{\partial p_s} \right] \tag{12′} \]
which is analogous to Eq. (12).

Notice that the first term on the right hand side of Eq. (12′) vanishes in the case of perfect substitutes, because demand for low-sulfur coal by utilities in the unconstrained monopoly pricing region is independent of \( p_s \). Nonetheless, an increase in the effective price of high-sulfur coal creates an expansion effect in the PRB coal market, which is the second term in the square brackets of (12′). The expansion effect can be derived by substituting \( N'(t, w, p_s) \) into Eq. (6′) to get \( p_s - tN'(t, w, p_s) - w = 0 \). Implicitly differentiating this equation yields \( \partial N'/\partial p_s = 1/t \), so that the expansion effect is given by
\[ \sigma q^*_l(n) \frac{\partial N^*}{\partial p_s} = \frac{\sigma q^d(n)}{t}. \tag{14} \]

The expansion effect of SO2 policy in (14) represents the sales of low-sulfur coal to utilities entering the extensive margin of railroad service in response to SO2 policy. An increase in the allowance price of \( d\lambda \) units raises the price of high-sulfur coal by \( \sigma d\lambda \) units, which stimulates an expansion on the extensive margin of railroad service of \( (\sigma/t) d\lambda \) units of distance, raising the total quantity of PRB coal deliveries by \( (\sigma q^*_l(n)/t) d\lambda \) units. The expansion effect of SO2 policy is identical under competition and monopoly freight pricing, because the terminal condition (6′) is identical in each case and \( q^*_l(n^*) = q^d(N^*) = q^d \) in equilibrium.

\(^9\) Formally, \( p(q^d) - tN^* - w = q^d p'(q^d) > 0 \) at \( x = n^* \) by Eq. (4′), which implies \( p(q^d) - tN^* - w q^d > 0 \).
The contraction effect can be found by substituting \( q^s = q^s(n^*) \) into boundary condition (5') to get
\[
p(q^s) = p_s.
\]
Dropping arguments for notational convenience and implicitly differentiating this equation gives
\[
\frac{\partial q^s}{\partial p_s} = \frac{\sigma(N^* - n^*)}{p^*}.
\]
An increase in the allowance price of \( d\lambda \) units raises the price of high-sulfur coal by \( \sigma d\lambda \) units, which bids up delivered prices for low-sulfur coal along the utility demand curve, reducing the quantity delivered to each of the \( N^* - n^* \) region II utilities served by the railroad by \( (\sigma / p'(q^s))d\lambda \) units. This contraction effect results in a decrease in the total quantity of PRB coal deliveries of \( (\sigma(N^* - n^*)/p'(q^s))d\lambda \) units in the railroads service region.

Summing terms in (14) and (15) gives
\[
\frac{\partial Q^{ss}}{\partial \lambda} = \frac{q^s(n^*) N^* - n^* p^*}{p^*} + \frac{p^* w^{ss}}{t p^*} = \frac{0}{0},
\]
where the latter equality holds by inspection of (4'). When spatial markets for low-sulfur coal are mediated by a monopoly railroad sector, environmental policies that raise the price of SO2 allowances are incapable of stimulating aggregate input substitution from high-sulfur coal to perfectly substitutable, low-sulfur coal. To better understand this outcome, consider the case of linear demand among electric utilities for coal, which is depicted in Fig. 2.

**Fig. 2.** Total quantity of PRB coal delivered under monopoly freight pricing.
After cap-and-trade regulation, total SO2 emissions decreases to some regulated level, \( E_1^* < E_0 \). The decrease in industry SO2 emissions occurs through a combination of input price effects that reduce the use of high-sulfur coal by utilities, \( q_1^* < q_0^* \), and fuel substitution effects from high-sulfur coal to PRB coal among utilities on the extensive margin of service, \( N_1 > N_0 \). For an arbitrary emissions cap of \( E_1^* \) units, the total amount of high-sulfur coal used in the regulated industry is represented in Fig. 2 by the shaded region, area \( N_1^*gh1 \). In equilibrium, total emissions satisfy \( E_1 = \sigma(1 - N_1^*)q_1^* \). Utilities located at distances between \( N_0 \) and \( N_1^* \) comply with the regulation by substituting away from high-sulfur coal to PRB coal and selling their SO2 allowances to utilities located at greater distances from the source mines for PRB coal. The outward expansion of the service region for PRB coal drives up the delivered price of PRB coal, and the expansion of the PRB service territory continues to distance \( N_1^* \) where the zero profit condition on the extensive margin of the railroad service region clears individual input demand for the remaining \( 1 - N_1^* \) utilities at the quantity \( (q_1^*)^\dagger \).\(^{10}\)

Notice that the permit price that emerges in the SO2 allowance market is independent of market structure in the PRB transportation sector. The reason is that the permit price is driven by fuel-switching behavior at the extensive margin of railroad service (region II) and utility demand for PRB coal is perfectly elastic for these utilities. The quantity of high-sulfur coal purchased by each utility located in the region \( 1 - N_1^* \) must clear the individual input demand at a level \( (q_1^*)^\dagger \) that exactly allocates aggregate SO2 emissions to meet the cap, and the effective price of high-sulfur coal, \( p_s \), must rise to clear demand at this quantity.

After cap-and-trade regulation on SO2, the delivered quantity schedule for PRB coal is depicted in the figure by \( b(\bar{x}) \). The change in the total delivered quantity of PRB coal in response to SO2 regulation can be seen in Fig. 2 as the sum of the expansion effect, area \( N_0^*fgN_1^* \), and the contraction effect, area \( abfe \).

To see why these areas must be equal in the case of perfectly substitutable coals, consider the linear (inverse) demand function \( p(x) = a - bq(x) \). In this case, the delivered quantity to a utility at distance \( x = 0 \) is \( q(0) = a/2b \) by Eq. (4'). The delivery quantity declines over distance at rate \( -t/2b \) in region I, and then remains constant thereafter at \( q_1^* \) in region II. Because the total service region expands by \( 1/t \), the magnitude of the expansion effect is \( q_1^*/t \).

The contraction effect can be decomposed as the area of the rectangle abfd less the triangle ade. The area of the rectangle is given by \( (N_0 - n_1^*)(q_0^* - q_1^*) \). The slope of the quantity schedule is \( -t/2b \), so that the quantity decrease along rail line of length \( n_1^* - n_0^* = 2/t \) is given by \( (q_0^* - q_1^*) = 1/b \). The length of the region II service area, \( N_0 - n_0^* \), can then be recovered in the linear case from Eqs. (4')–(6'). Evaluating (4') at \( n_0^* \) gives \( a - 2bq_0^* = tn_0^* \). Making use of (5'), this implies \( p_s - bq_0^* = tn_0^* \). From (6'), \( p_s = tn_0^* \), and combining these equations and canceling terms gives \( N_0 - n_0^* = (q_0^* - q_1^*)/t \). The area of rectangle abfd is thus \( q_0^*/t \). Now consider the triangle ade. Noting that the base of the triangle is \( n_1^* - n_0^* = 2/t \), its area is \( (q_0^* - q_1^*)/t \). The magnitude of the contraction effect, therefore, is \( q_0^*/t - (q_0^* - q_1^*)/t = q_1^*/t \), an amount that exactly offsets the expansion effect in the market.

Our analysis predicts no direct input substitution from high-sulfur coal to low-sulfur coal arises in response to SO2 regulation in the case of perfectly substitutable fuels and spatial monopoly pricing. Nevertheless, output effects in energy markets that lead to increased electricity prices would provide incentives for increased use of low-sulfur coal (and impact SO2 allowance prices). With downward-sloping electricity demand, SO2 policy reduces the total amount of coal combusted by electric utilities, which increases electricity prices and shifts the derived demand for coal outward at each utility. This effect would exist apart from the input substitution incentive we have isolated here. In Fig. 2, the rise in electricity prices would create a level effect in coal input demand, and the magnitude of the level effect at the margin (i.e., at quantity level \( q_1^* \) ) would be capitalized into SO2 allowance prices.

4. Evidence from PRB coal markets

Our analysis predicts that SO2 policy has little effect on the observed input substitution pattern from high-sulfur to low-sulfur coal when the transportation market is mediated by a spatial

\(^{10}\) Formally, under an emissions cap of \( E_1^* \) units the extensive margin of PRB service must satisfy \( N_1 = 1 - E_1/\sigma q_1^* \), which can be used together with Eqs. (5') and (6') to recover \( q_1^* \) and \( N_1^* = 1 - E_1/\sigma q_1^* \). For a binding emissions cap, the equilibrium allowance price solves Eq. (6'): \( \lambda = \left( tN_0^* + w - p_h \right)/\sigma \).
monopolist. Two empirical questions emerge from this observation. First, what does the available evidence suggest about spatial market power in the low-sulfur coal market? And, second, if the observed input substitution from high-sulfur to low-sulfur coal was not driven by SO2 policy, then what features of the PRB coal market were responsible for the shift?

There is considerable evidence that railroads hauling PRB exercise market power. In an earlier paper (Gerking and Hamilton, 2008) we examine data from the Carload Waybill Sample obtained from the U.S. Department of Transportation, Surface Transportation Board, for evidence of railroad market power in PRB deliveries. Along 353 observed transportation routes for PRB coal, all deliveries over the period 1988–1999 were initiated by one of two railroads (BNSF and UP), with market shares of 55.3% for BNSF and 44.7% for UP in 1999.11 Supplementing these data with information taken from Form 423 of the Federal Energy Regulatory Commission (FERC), we link the quantity of coal received by a power plant on each shipment to the delivered price and the source mine for the PRB coal, resulting in comprehensive transactional information on the spatial pricing of PRB coal over 1229 observations (mine-utility pairs). We examine Lerner indices, controlling for route-specific effects, and find that the average values of the Lerner indices fall with distance in the sample, a finding that concurs with the spatial pricing profile depicted in Fig. 1.12 Another prediction of the model is that SO2 policy would extend the length of region I and contract the length of region II for a spatial monopoly railroad, increasing market power overall. Controlling for route-specific effects, we indeed find that the annual average values of L* were 15% larger in 1999 than in 1988 (a difference that is significantly different from zero at the 1% level).

We estimate that roughly 90% of the power plants in the data set are region II plants that are subject to freight rates equal to \(ps - w\). For these plants, an exogenous shock in unit transportation cost in the railroad sector (\(t\)) has no affect on freight rates, and marginal changes in transportation cost alter freight rates only to electric utilities less than 550 miles distant from the source mine (the shortest 10% of all sample routes). This estimated relationship between marginal transportation cost and freight rates is inconsistent with the operation of a competitive railroad sector, in which a one-unit decrease in marginal cost leads to a one-unit decrease in the freight rate throughout the service region, irrespective of distance from the mine.

The estimated distinction between service regions I and II for a railroad monopoly in Gerking and Hamilton (2008) provides essential insight into the market features that supported the observed fuel substitution from high-sulfur to low-sulfur coal. In the railroad sector, Ellerman and Montero (1998) and Ellerman et al. (2000) estimate that real per ton/mile freight rates from the PRB to the Midwest fell by 44% over the period 1987–1993, and argue that these rate reductions occurred due to increased competition following deregulation in the Staggers Rail Act in 1980 and to significant productivity-enhancing improvements attained by railroads throughout the period. Controlling for distance, we find real rail rates on PRB coal shipments decreased by 22% over the period 1988–1999, with the remaining decline in freight rates explained by an expansion of the service territory on the extensive margin of railroad service to distant utilities that receive lower freight prices per ton mile (Gerking and Hamilton, 2008). Despite the Staggers Rail Act, railroad behavior over the 1988–1999 period appears to have become less competitive, as evidenced by the increase in the Lerner indices across routes of fixed length. Empirical evidence indicates that the real marginal cost of transporting coal declined by 36% over the 1988–1999 period in response to efficiency gains that followed the shift from steel railcars to lighter weight aluminum railcars in the late 1980s and 1990s. The decline in freight rates as transportation costs decreased in the rail sector.

Another causal factor that contributed to the shift among electric utilities to burning low-sulfur coal was a dramatic reduction in extraction costs for PRB coal. Over the period 1985–2000, coal extraction costs declined by 57% from $12.84/ton to $5.55/ton and the average real mine mouth price of PRB correspondingly coal fell by 64% (from $13.97/ton vs. $5.38/ton) (U.S. Department of Energy, 2008).

---

11 A transportation route is defined as a railhead at or near a particular mine to a particular power plant.
12 The estimated value of the Lerner index over the entire sample (all dates and routes) in Gerking and Hamilton (2008) is 0.37, which accords well with the estimated value of 0.41 over the period 2001–2006 in a recent study commissioned by the U.S. Department of Transportation (Christiansen Associates, Inc., 2008). Concern over railroad market power has led to Companion Bills (S 146 and HR 233) presently before the U.S. Congress to amend the railroad antitrust exemptions in the Clayton Act.
Energy Information Administration, various years). The dramatic decrease in the mine mouth price of PRB coal, combined with lower freight rates in the rail sector that occurred through independent efficiency improvements attained by railroads at about this same time, appear to have been the predominant factors that facilitated the landmark shift in the utilization of PRB coal at electric utilities in the period surrounding implementation of the U.S. Clean Air Act.

5. Conclusion

Several pertinent conclusions may be drawn from our analysis of the effect of SO$_2$ policy on input substitution from high-sulfur coal to low-sulfur coal. First, SO$_2$ regulation leads to an increase in the average transportation cost of delivered PRB coal. By increasing the (allowance-inclusive) price of high-sulfur coal, SO$_2$ regulation expands the service territory for PRB coal and increases the average distance each ton of coal is shipped. In the case of a monopoly railroad and perfectly substitutable fuels, the rise in average delivery cost for PRB coal is coupled with no change in total deliveries of PRB coal, so that SO$_2$ policy serves only to redistribute low-sulfur coal shipments from nearby utilities to those further away, reducing social returns. Second, SO$_2$ regulation exacerbates railroad market power by increasing the wedge between the delivered low-sulfur coal price and marginal shipment cost to (incumbent) utilities in region II. The reason is that freight rates rise by the same magnitude as the increase in the price of high-sulfur coal, whereas the cost schedule does not change.

Our observations on the limited ability of SO$_2$ regulation to stimulate increased PRB coal shipments through input substitution are most stark in the case of: (i) a uniform spatial distribution of utilities; (ii) perfect substitutability between low- and high-sulfur coal in electricity generation; and (iii) monopoly power by railroads hauling low-sulfur coal. While this situation provides a useful benchmark to consider the effect of environmental policy in spatially intermediated markets, more general models that relax these assumptions show that SO$_2$ policy can either raise or lower the aggregate quantity of PRB coal shipped. Under a non-uniform spatial distribution of utilities, SO$_2$ policy can potentially lead to negative input substitution – that is, a decrease in PRB coal deliveries – when the density of region II utilities decreases monotonically along the rail line towards the extensive margin of railroad service, as the contraction effect in this case would exceed the expansion effect of the policy.

Overall, in the case of U.S. SO$_2$ policy, the evidence suggests that sharp declines in mining cost and railroad transportation cost for PRB coal at around the time of the U.S. Clean Air Act Amendments are likely to have played a much larger role in facilitating the substitution from high-sulfur coal to low-sulfur, PRB coal than the environmental regulation. Indeed, it is possible that environmental policy did not contribute to the observed input substitution at all.

Acknowledgements

The authors thank Spencer Banzhaf, Michael Caputo, John Horowitz, Charles Kolstad, Wally Milon, and seminar participants at UC Berkeley, UC Santa Barbara, the University of Maryland, the University of Nebraska, the University of Massachusetts, and the NBER Summer Institute Workshop on Environmental Economics for numerous constructive suggestions.

References