Pollution markets with imperfectly observed emissions

Juan-Pablo Montero*

Catholic University of Chile and MIT

November 20, 2003

Abstract

I study the advantages of pollution permit markets over traditional standard regulations when the regulator has incomplete information on firms' emissions and costs of production and abatement (e.g., air pollution in large cities). Because the regulator only observes each firm's abatement technology but neither its emissions nor its output, there are cases in which standards can lead to lower emissions and, hence, welfare dominate permits. If permits can be optimally combined with standards, however, in many cases this hybrid policy converges to the permits-alone policy but (almost) never to the standards-alone policy. I then empirically examine these issues using evidence from a permits market in Santiago, Chile.

Keywords: asymmetric information, imperfect monitoring, pollution markets, permits

JEL classification: L51, Q28

*<jmontero@faceapuc.cl> Department of Economics, Catholic University of Chile, Vicuna Mackenna 4860, Santiago, Chile. The author is also Research Associate at the MIT Center for Energy and Environmental Policy Research (MIT-CEEPR). I am grateful to Luis Cifuentes, Denny Ellerman, Paul Joskow, Matti Liski, José Miguel Sánchez and seminar participants at Columbia, Econometric Society Meeting (UCLA), Global Development Network Conference (Cairo), Helsinki School of Economics, MIT, PUC and Universidad de Concepción for comments and discussions; Joaquín Poblete for research assistance; and MIT-CEEPR for financial support.

1 Introduction

In recent years, environmental policy makers are paying more attention to pollution markets (i.e., tradable emission permits) as an alternative to the traditional command-and-control (CAC) approach of setting emission and technology standards. A notable example is the 1990 U.S. acid rain program that implemented a nationwide market for electric utilities' sulfur dioxide (SO_2) emissions (Schmalensee et al., 1998; Ellerman et al., 2000). In order to have a precise estimate of the SO_2 emissions that are actually going to the atmosphere, the acid rain program requires each affected electric utility unit to install costly equipment that can continuously monitor emissions.[1]

This and other market experiences suggest that conventional permits markets are likely to be implemented in those cases where emissions can be closely monitored, which almost exclusively occurs in large stationary sources like electric power plants and refineries. Then, it is not surprising that, among other reasons, environmental authorities continue relying on CAC instruments to regulate emissions from smaller sources because compliance with such instruments only requires the authority to ensure that the regulated source has installed the required abatement technology or that its emissions per unit of output are equal or lower than a certain emissions rate standard.[2] In addition, some regulators view that a permits program where emissions cannot be closely monitored may result in higher emissions than under an alternative CAC regulation. Since compliance cannot be made contingent on firm's output or

[1] Another example with similar monitoring requirements is the Southern California RECLAIM program that implemented separated markets for nitrogen oxide (NO_x) and SO_2 emissions from power plants, refineries and other large stationary sources. This program did not include a market for volatile organic compounds (VOC) in large part because of the difficulties with monitoring actual emissions from smaller and heterogeneous sources (Harrison, 1999).

[2] Note that there are some credit-based trading programs aimed at curbing air pollution in urban areas working in the US (Tietenberg, 1985). New sources (or expansion of existing ones) must acquire emission credits to cover their emissions through, for example, shutting down existing plants or scrapping old vehicles. Although these programs have the merit of involving small sources, they are very limited in scope in the sense that they are embedded within an existing CAC regulation and are particularly designed to prevent further deterioration of air quality from the entry of new sources.

utilization, emissions could in fact be higher if the trade pattern is such that lower-output firms sell permits to higher-output firms.

Thus, pollution markets may appear inapproapriate for effectively reducing air pollution in large cities where emissions come from many small (stationary and mobile) sources rather than a few large stationary sources (e.g., air pollution in large cities like Mexico City, Sao Paulo in Brasil and Santiago-Chile). While one cannot neglect a priori that a permits market in which emissions (and ouput) are only imperfectly monitored may eventually lead to higher emissions than an alternative regulation, I think that rather than disregard pollution markets as a policy tool, the challenge faced by policy makers is when and how to implement these markets using monitoring procedures that are similar to those under CAC regulation.

Most of the literature on environmental regulation under asymmetric information deals with the case in which firms' costs are privately known but emissions are publicly observed (see Lewis (1996) for a survey). While there is some literature considering imperfect observation of emissions (e.g., Segerson, 1988; Fullerton and West, 2002),[3] there seems to be no work specifically studying the effect of imperfect information on emissions and costs on the design and performance of a permits market. Yet, it is interesting to observe that despite its limited information on each source's actual emissions (and costs), Santiago-Chile's environmental agency has already implemented a (small) market to control total suspended particulate (TSP) emissions from a group of about 600 stationary sources (Montero et al., 2002).[4] Based on estimates from annual inspection for technology parameters such as source's size and fuel type, Santiago's environmental regulator approximates each source's actual emissions by the maximum amount of emissions that the source could potentially emit in a given year.[5] I believe that a

[3] Fullerton and West (2002) study the control of vehicle emissions using a combination of taxes on cars and on gasoline as an alternative to an (unavailable) tax on emissions. Segerson (1988) study the control of emissions from (few) non-point sources using a "moral hazard in teams" approach.

[4] Sources affected by the TSP program are responsible for only 5% of 2000 TSP emissions in Santiago.

[5] As we shall see later, using the source's maximum emissions as a proxy does not prevent any adverse effects

close (theoretical and empirical) examination of this permits program represents a unique case study of issues of instrument choice and design that can arise in the practical implementation of environmental markets in which regulators face important information asymmetries.

In the next section (Section 2) I develop a theoretical model and start by showing that the regulator cannot implement the first-best when emissions are imperfectly measured (even in the absence of budget constraints and perfect cost information). In Section 3, I derive the optimal design of two policy instruments: emission standards and emission permits. I focus on these two policies rather than on more optimal ones not only because the latter include the use of nonlinear instruments and transfers to firms which has not been used in practice (Stavins, 2003; Hahn et al., 2003), but more importantly, because I want to specifically explore whether permits can still provide an important welfare advantage over traditional CAC regulation when emissions are imperfectly monitored.

In Section 4, I first discuss the conditions under which permits are welfare superior to standards. I show that permits provide firms not only with flexibility to reduce production and abatement costs but sometimes with perverse incentives to choose socially suboptimal combinations of output and abatement (something that would not occur if emissions were accurately measured). There are two situations in which the latter possibility can lead permits to higher emissions, and hence, be potentially welfare dominated by standards: (1) when firms with relatively large output ex-ante (i.e., before the regulation) are choosing low abatement (i.e., when there is a negative correlation between production and abatement costs), and (2) when firms doing more abatement find it optimal to reduce output ex-post. Because in deciding whether to use permits or standards, the regulator is likely to face this trade-off between cost

that the use of permits (instead of CAC regulation) could eventually have on aggregate emissions. The choice of proxy is an arbitrary matter because the number of permits being allocated can always be adjusted accordingly with no efficiency effects.

savings and possible higher emissions, I then discuss the advantages of implementing an optimal hybrid policy in which permits are combined with an optimally chosen standard. I find that in many situations the hybrid policy converges to the permits-alone policy but it almost never converges to the standards-alone policy.

In Section 5, I empirically examine the advantages of a permits program using emissions and output data from the Santiago's TSP permits program. I find evidence of large cost savings but also of higher emissions (about 6%) relative to what would have been observed under an equivalent standards policy. However, the welfare loss from higher emissions is only 8% of the welfare gain from lower abatement and production costs. Final remarks are in Section 6.

2 The model

Consider a competitive market for an homogeneous good supplied by a continuum of firms of mass 1. Each firm produces output q and emissions e of a uniform flow pollutant. To simplify notation, I assume that when the firm does not utilize any pollution abatement device $e = q$. Market inverse demand is given by $P = P(Q)$, where Q is total output and $P'(Q) \leq 0$. Total damage from pollution is given by $D(E)$, where E are total emissions and $D'(E) > 0$. Functions $P(Q)$ and $D(E)$ are known to the regulator.

A firm can abate pollution at a positive cost by installing technology x, which reduces emissions from q to $e = (1 - x)q$. Hence, the firm's emission rate is $e/q = 1 - x$. Each firm is represented by a pair of cost parameters (β, γ). A firm of type (β, γ) has a cost function $C(q, x, \beta, \gamma)$ where β and γ are firm's private information. To keep the model mathematically tractable, I assume that the cost function has the following quadratic form in the relevant output-abatement range[6]

$$C(q,x,\beta,\gamma) = \frac{c}{2}q^2 + \beta q + \frac{k}{2}x^2 + \gamma x + vxq \tag{1}$$

where c, k and v are known parameters common to all firms and $c > 0$, $k > 0$, $\Lambda \equiv ck - v^2 > 0$ and $v \gtreqless 0$.[7]

Function (1) incorporates two key cost parameters that are essential to understand firms' behavior under permits and standards regulation. One of these cost parameters is the correlation between β and γ (that we shall denote by ρ), which captures whether firms with higher output ex-ante (i.e., before the regulation) are more or less likely to install more abatement x. The other cost parameter is v, which captures the effect of abatement on output ex-post (note that we have constrained v to be the same for all firms, thus, a negative value of v would indicate that, on average, the larger the x the larger the increase in q ex-post). As we shall see, the values of the cost parameters v and ρ play a fundamental role in the design and choice of policy instruments when emissions are not closely monitored.

Although the regulator does not observe firms' individual values for β and γ, we assume that he knows that they are distributed according to the cumulative joint distribution $F(\beta,\gamma)$ on $\beta \in [\underline{\beta}, \overline{\beta}]$ and $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.[8] To simplify notation further and without any loss of generality I let E[β] =E[γ] = 0, where E[$\cdot$] is the expected value operator.[9] I also use the following notation:

[6] This approach was first introduced by Weitzman (1974).

[7] The parameter v can be negative, for example, if switching to a cleaner fuel saves on fuel costs but involves such a large retrofitting cost (i.e., high k) that no firm switches to the cleaner and cheaper fuel unless regulated.

[8] Note that we can easily add aggregate uncertainty to this formulation by simply letting $\beta^i = \beta^i + \theta$ and $\gamma^i = \gamma^i + \eta$, where θ and η are random variables common to all firms.

[9] Note that because β and γ are negative for some firms, one can argue that marginal costs can take negative values. This possibility is eliminated by assuming parameter values (including those in the demand and damage functions) that lead to interior solutions for q and x in which $\partial C/\partial q > 0$ and $\partial C/\partial x > 0$ for all β and γ. Furthermore, since these interior solution are assumed to fall within the range in which (1) is valid, what happens beyond this range is not relevant for the analysis of instrument design and choice that follows. Alternatively, one can let $\beta \in [0, \overline{\beta}]$ and $\gamma \in [0, \overline{\gamma}]$ with some further notation in the optimal designs but no change in the welfare comparisons.

$\text{Var}[\beta] \equiv \sigma_\beta^2$, $\text{Var}[\gamma] \equiv \sigma_\gamma^2$, $\text{Cov}[\beta,\gamma] \equiv \rho\sigma_\beta\sigma_\gamma$ and $F_{\beta\gamma} \equiv \partial^2 F(\beta,\gamma)/\partial\beta\partial\gamma$.

Firms behave competitively, taking the output clearing price P as given. Hence, in the absence of any environmental regulation, each firm will produce to the point where its marginal production cost equals the product price (i.e., $C_q(q,x,\beta,\gamma) = P$), and install no abatement technology (i.e., $x = 0$). Because production involves some pollution, this market equilibrium is not socially optimal. The regulator's problem is then to design a regulation that maximizes social welfare.

I let the regulator's social welfare function be

$$W = \int_0^Q P(z)dz - \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} C(q,x,\beta,\gamma) F_{\gamma\beta} d\beta d\gamma - D(E) \tag{2}$$

where $Q = \int_\beta \int_\gamma q(\beta,\gamma) F_{\gamma\beta} d\gamma d\beta$ is total output and $E = \int_\beta \int_\gamma (1 - x(\beta,\gamma)) q(\beta,\gamma) F_{\gamma\beta} d\gamma d\beta$ is total emissions. In this welfare function, the regulator does not differentiate between consumer and producer surplus and transfers from or to firms are lump-sum transfers between consumers and firms with no welfare effects.[10]

The firms' outputs and abatement technologies that implement the social optimum or first-best outcome are given by the following two first-order conditions

$$q : P(Q) - C_q(q,x,\beta,\gamma) - D'(E) \cdot (1 - x) = 0 \tag{3}$$

$$x : -C_x(q,x,\beta,\gamma) - D'(E) \cdot (-q) = 0 \tag{4}$$

If rearranged, eq. (3) says that the benefits from consuming an additional (small) unit of output is equal to its cost of production and environmental effects. Eq (4), on the other hand, says

[10] The model can be generalized by allowing the regulator to consider a weight $\mu \neq 1$ for firm profits and a shadow cost $\lambda > 0$ for public funds. However, this would not add much to our discussion.

that emissions should be reduced to the point where the marginal cost of emissions abatement (i.e., C_x/q) is equal to marginal damages (i.e., $D'(E)$).

Because of various information asymmetries between firms and the regulator it is not clear that the latter can design an environmental policy that can attain the first-best resource allocation. The regulator's problem then becomes to maximize (2) subject to information constraints (and sometimes administrative and political constraints as well). From the various possible type of information constraints, the case that have attracted most attention in the literature is that in which the regulator knows little or nothing about firms' costs (i.e., he may or may not know F) but can costlessly monitor each firm's actual emissions e (Kwerel, 1977; Dasgupta et al.,1980; Spulber, 1988; Lewis, 1996). These authors show that in many cases the first-best can be implemented. It is the case here, for example, if the regulator announces a (non-linear) emissions tax schedule $\tau(E)$ equal to $D'(E)$.[11]

In this paper, however, I am interested in the problem where the regulator cannot directly observe firms' actual emissions $e = (1 - x)q$; although he can costlessly monitor firms' abatement technologies or emission rates x. As in Santiago's permits program, this information asymmetry will be present when both continuous monitoring equipment is prohibitively costly and individual output q is not observable.[12] Thus, if the regulator asks for an output report from the firm, we anticipate that the firm would misreport its output whenever this was to its advantage. In this case, the regulator cannot implement the social optimum regardless of the information he or she has about firm's costs.[13]

[11] A competitive firm (β, γ) takes the price $P(Q)$ and tax rate $\tau(E) = D'(E)$ as given and maximize $\pi(q, x, \beta, \gamma) = P(Q)q - C(q, x, \beta, \gamma) - D'(E) \cdot (1 - x)q$ with respect to q and x. The firm's first order conditions for q and x are those given by (3) and (4).

[12] The regulator can nevertheless estimate total output Q from the observation of the market clearing price P.

[13] Consider the extreme situation in which regulator knows both β and γ. His optimal policy will be some function $T(x; \beta, \gamma)$ in the form of either a transfer from the firm or to the firm. Then, firm (β, γ) takes $P(Q)$ and $T(x; \beta, \gamma)$ as given and maximizes $\pi(q, x, \beta, \gamma) = P(Q)q - C(q, x, \beta, \gamma) - T(x; \beta, \gamma)$ with respect to q and x. It is not difficult to see that firm's first order conditions for q and x will always differ from (3) and (4) for any function $T(x; \beta, \gamma)$.

Even if the regulator has perfect knowledge of firm's costs and, therefore, can ex-post deduce firm's output based on this information and the observation of x, the fact that he cannot make the policy contingent on either emissions or output prevents him from implementing the first-best. In other words, the regulator cannot induce the optimal amounts of output and emissions with only one instrument (i.e., x).[14] Consequently, the regulator must necessarily content himself with less optimal policies.

3 Instrument design

Rather than considering a full range of policies, I focus here on those policies that are either currently implemented or have drawn some degree of attention from policy makers in the context of urban air pollution control (Stavins, 2003). I first study the optimal design of a traditional technology (or emission rate) standard and then the optimal design of a permits market. I leave the optimal design of the hybrid policy, which builds upon the individual designs, for Section 4. To keep the model simple, I make two further simplifications regarding the demand curve and the damage function. I let $P(Q) = P$ (constant) and $D(E) = hE$, where $h > v$.[15]

To visualize how the two policy designs depart from the social optimum, it is useful to compute the first-best and keep it as a benchmark. Plugging the above assumptions in (3) and (4), the first-best outcome is given by

$$x^* = \frac{(P - h - \beta)(h - v) - \gamma c}{ck - (h - v)^2} \tag{5}$$

[14] Since the regulator can have a good idea of total emissions E from air quality measures, one might argue that Holmström's (1982) approach to solving moral hazard problems in teams may apply here as well. However, in our context this approach is unfeasible because the large number of agents would require too big transfers; either from firms as penalties or to firms as subsidies.

[15] Besides that these assumptions facilitate the empirical estimation, they tend to offset each other, and if anything, they tend to favor the standards policy. Because costs are always lower under the permits policy, a backward sloping demand would lead to higher output, and hence, to higher surplus under such policy. On the other hand, because it is not clear a priory whether emissions are higher or lower under the permits policy, a convex damage function may or may not lead to higher environmental damages under such policy.

$$q^* = \frac{P - h - \beta + (h - v)x^*}{c} \tag{6}$$

It is immediate that $\partial x^*/\partial\beta < 0$, $\partial x^*/\partial\gamma < 0$, $\partial q^*/\partial\beta < 0$ and $\partial q^*/\partial\gamma < 0$. As expected, higher production and abatement costs lead to lower output and abatement levels.

3.1 Standards

The regulator's problem here is to find the emission rate standard x_s to be required to all firms that maximizes social welfare (subscript "s" denotes standards policy). The regulator knows that for any given x_s, firm (β, γ) will maximize $\pi(q, x_s, \beta, \gamma) = Pq - C(q, x_s, \beta, \gamma)$. Hence, firm's (β, γ) output decision will solve the first-order condition

$$P - cq - \beta - vx_s = 0$$

which provides the regulator with firm's output q as a function of the standard x_s

$$q_s(x_s) \equiv q_s = \frac{P - \beta - vx_s}{c} \tag{7}$$

Since x_s will be the same across firms, it is clear that production under a standard will also be suboptimal relative to the first-best q^* as q_s does not adapt to changes in γ.

Based on the welfare function (2), the regulator now solves

$$\max_{x_s} \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[Pq_s - C(q_s, x_s) - h \cdot (1 - x_s)q_s\right] F_{\beta\gamma} d\gamma d\beta$$

By the envelope theorem, the first-order condition is

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[-kx_s - \gamma - vq_s - h \cdot (1 - x_s)\frac{\partial q_s}{\partial x_s} + hq_s\right] F_{\beta\gamma} d\gamma d\beta = 0 \tag{8}$$

By replacing (7) and $\partial q_s/\partial x_s = -v/c$ into (8), the first-order condition (8) reduces to

$$-ckx_s - v \cdot (P - vx_s) + h \cdot (1 - x^s)v + h \cdot (P - vx_s) = 0$$

which leads to the optimal standard

$$x_s = \frac{P \cdot (h - v) + hv}{\Lambda + 2vh} \tag{9}$$

where $\Lambda \equiv ck - v^2 > 0$. Comparing this result to the first-best (5), it is interesting to observe that $x_s > x^*(\beta = 0, \gamma = 0)$. This indicates that even in the absence of production and abatement cost heterogeneity (i.e., $\beta = \gamma = 0$ for all firms), the standards policy still require firms to install more abatement technology than is socially optimal. Because $q_s(x^*) > q^*$, it is optimal to set x_s somewhat above x^* to bring output q_s closer to its optimal level q^*.

3.2 Permits

The regulator's problem now is to find the total number permits $\widetilde{e}_0$ to be distributed among firms that maximizes social welfare. Let R denote the equilibrium price of permits, which will be determined shortly.[16] The regulator knows that firm (β, γ) will take R as given and solve

$$\max_{q,x} \pi(q, x, \beta, \gamma) = Pq - C(q, x, \beta, \gamma) - R \cdot (\widetilde{e} - \widetilde{e}_0)$$

where $\widetilde{e} = (1-x)\widetilde{q}$ are firm's proxied emissions and $\widetilde{q}$ is some arbitrarily output or capacity level that is common to all firms (recall that the regulator only observes x). For example, $\widetilde{q}$ could be set equal to the maximum possible output that could ever be observed, which would occur

[16] Note that under a tax policy, the optimal price R will be the quasi-emissions tax. If we add aggregate uncertainty to the model, both policies will not be equivalent from an efficiency standpoint.

when $x = 0$ and $\beta = \underline{\beta}$. As we shall see later, the exact value of $\widetilde{q}$ turns out to be irrelevant because it simply works as a scaling factor. Note that if $\widetilde{e} < \widetilde{e}_0$ the firm will be a seller of permits.

From firms' first-order conditions

$$x: \quad -kx - \gamma - vq + R\widetilde{q} = 0 \tag{10}$$

$$q: \quad P - cq - \beta - vx = 0 \tag{11}$$

we have that firm's (β, γ) optimal abatement and output responses to R and $\widetilde{q}$ (or, more precisely, to $R\widetilde{q}$) are

$$x_p = \frac{R\widetilde{q}c - \gamma c - (P - \beta)v}{\Lambda} \tag{12}$$

$$q_p = \frac{P - \beta - vx_p}{c} \tag{13}$$

where the subscript "p" denotes permits policy. Comparing (12) and (13) with (5) and (6) illustrates the trade-off a regulator faces when implementing a permits program with imperfect observation of emissions. While $\partial q_p/\partial\beta$ and $\partial x_p/\partial\gamma$ are negative as in the first-best, $\partial q_p/\partial\gamma$ and $\partial x_p/\partial\beta$ are both positive when $v > 0$.[17]

Now we can solve the regulator's problem of finding the optimal $\widetilde{e}_0$. Since the market clearing condition is

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \widetilde{e} F_{\beta\gamma} d\gamma d\beta = \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} (1 - x_p)\widetilde{q} F_{\beta\gamma} d\gamma d\beta = \widetilde{e}_0 \tag{14}$$

and x_p is a function of $R\widetilde{q}$ as indicated by (12), it is irrelevant whether we solve for $R\widetilde{q}$ or $\widetilde{e}_0/\widetilde{q}$. Hence, we let the regulator solve (permits purchases and sales are transfers with no net welfare

[17] Note that $\partial q_p/\partial\beta = -k/\Lambda$, $\partial x_p/\partial\beta = \partial q_p/\partial\gamma = v/\Lambda$ and $\partial x_p/\partial\gamma = -c/\Lambda$.

effects)

$$\max_{R\widetilde{q}} \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[Pq_p(x_p(R\widetilde{q})) - C(q_p(x_p(R\widetilde{q})), x_p(R\widetilde{q})) - h \cdot (1 - x_p(R\widetilde{q}))q_p(x_p(R\widetilde{q})) \right] F_{\beta\gamma} d\gamma d\beta$$

By the envelope theorem, the first-order condition is

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[-(1 - x_p)h \frac{\partial q_p}{\partial (R\widetilde{q})} + hq_p \frac{\partial x_p}{\partial (R\widetilde{q})} - R\widetilde{q} \frac{\partial x_p}{\partial (R\widetilde{q})} \right] F_{\beta\gamma} d\gamma d\beta = 0 \tag{15}$$

By plugging $\partial q_p / \partial (R\widetilde{q}) = [\partial q_p / \partial x_p][\partial x_p / \partial (R\widetilde{q})]$, $\partial q_p / \partial x_p = -v/c$, (12) and (13) into (15), the first-order condition can be rearranged to obtain the optimal permits price

$$R\widetilde{q} = \frac{Ph(kc + v^2) + hv\Lambda}{(\Lambda + 2hv)c} \tag{16}$$

which, in turn, allows us to obtain the optimal permits allocation $\widetilde{e}_0 / \widetilde{q}$ by simply replacing (16) in (12) and that in (14). As in the standards policy, this allocation induces, on average, more abatement than is socially optimal (note that $x_p(0,0) > x^*(0,0)$).

We can now replace $R\widetilde{q}$ in (12) and (13) to obtain expressions for q_p and x_p that are more readily comparable to q_s and x_s (see eqs. (7) and (9)). After some algebra, the following expressions are obtained

$$x_p = x_s + \frac{v\beta - c\gamma}{\Lambda} \tag{17}$$

$$q_p = \frac{P - vx_s}{c} - \frac{k\beta - v\gamma}{\Lambda} = q_s - \frac{v^2\beta - cv\gamma}{c\Lambda} \tag{18}$$

If firms are homogeneous (i.e., $\beta = \gamma = 0$ for all firms), it is not surprising that $x_p = x_s$ and $q_p = q_s$ and that both regulations provide the same welfare. As firms become more heterogenous, x and q move in different directions depending on the regulatory regime and the value of v (it will also depend on ρ as we will see next). In fact, when $v > 0$ and firms differ on γ, firms' abatement decisions x tend to remain closer to the social optimum x^* under the permits regulation than under the CAC regulation since $\partial x^*/\partial \gamma$ and $\partial x_p/\partial \gamma$ are both negative.[18] However, as firms differ on β, firms' abatement decisions remain closer to x^* under the CAC regulation since $\partial x^*/\partial \beta$ and $\partial x_p/\partial \beta$ have opposite signs. A similar trade-off can be found analyzing firms' production decisions to changes in β and γ. Because this eventual output/abatement "misalignment" may result in higher emissions under the permits policy, in deciding whether or not to implement a permits program, the regulator will inevitably face a trade-off between abatement flexibility and possible higher emissions. I study this trade-off more formally in the next section.

4 Instrument choice

The regulator must now decide wether to use permits or standards or, eventually, a combination of both. Although by construction a hybrid policy cannot be welfare dominated by either single-instrument policy, it is useful to compare first single-instrument policies and then explore the extra gains, if any, of implementing a hybrid policy.

4.1 Permits versus standards

For a regulator that is limited to use a single instrument, the difference in the social welfare between the optimal permits policy and the optimal standards policy is

$$\Delta_{ps} = W_p(\widetilde{e}_0/\widetilde{q}) - W_s(x_s) \tag{19}$$

[18] Note that $\partial x^*/\partial \gamma = \partial x_p/\partial \gamma$ for $v = 0$.

where $\widetilde{e}_0$ is the optimal number of permits normalized by some $\widetilde{q}$ and x_s is the optimal standard. The normative implication of (19) is that if $\Delta_{ps} > 0$, the regulator should implement the permits policy.

To explore under which conditions this is the case, we write (19) as

$$\Delta_{ps} = \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[Pq_p - C(q_p, x_p) - (1 - x_p)q_p h - Pq_s + C(q_s, x_s) + (1 - x_s)q_s h \right] F_{\beta\gamma} d\gamma d\beta \quad (20)$$

where q_p, x_p, q_s and x_s can be expressed according to (17) and (18). Since $Q_p = Q_s = (P - vx_s)/c$, eq. (20) can be re-written as

$$\Delta_{ps} = \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[\{C(q_s, x_s) - C(q_p, x_p)\} + \{(1 - x_s)q_s - (1 - x_p)q_p\} h \right] F_{\beta\gamma} d\gamma d\beta \quad (21)$$

Recalling that $e = (1-x)q$, the first curly bracket of the right hand side of (21) is the difference in costs between the two policies, whereas the second curly bracket is the difference in emissions that multiplied by h gives the difference in pollution damages.

If we plug (17) and (18) into (21), after some algebra (21) becomes

$$\Delta_{ps} = \frac{v^2\sigma_\beta^2 - 2cv\rho\sigma_\beta\sigma_\gamma + c^2\sigma_\gamma^2}{2c\Lambda} - \frac{h \cdot \left(kv\sigma_\beta^2 - (kc + v^2)\rho\sigma_\beta\sigma_\gamma + cv\sigma_\gamma^2 \right)}{\Lambda^2} \quad (22)$$

and after collecting terms, it reduces to

$$\Delta_{ps} = A_1\sigma_\beta^2 + A_2\sigma_\gamma^2 + A_3\rho\sigma_\beta\sigma_\gamma \quad (23)$$

where $A_1 = (v^2\Lambda - 2ckhv)/2c\Lambda^2$, $A_2 = (c\Lambda - 2chv)/2\Lambda^2$ and $A_3 = (ckh + hv^2 - v\Lambda)/\Lambda^2 > 0$. Since A_1, A_2 and ρ can be either positive, negative or zero,[19] the sign of (23) will depend upon

[19] Recall that for interior solutions in all cases we must have $ck > (h - v)^2$, $ck > v^2$, and $h > v$.

the value of the different parameters. As the heterogeneity across firms decreases (i.e., σ_β^2 and σ_γ^2 approach zero), however, the welfare difference between the two policies tend to disappear.[20]

The ambiguous sign of (23) should not be surprising given the trade-off between flexibility and potential higher emissions that we identified in the previous section. Expression (22) illustrates this trade-off more clearly. The first term is the difference in costs between the two policies. Since $-1 \leq \rho \leq 1$, this term is always positive which indicates that the optimal permits policy is always less costly than the optimal standards policy. The second term is the difference in damages, which can either be positive, negative or zero depending on the value of the different parameters of the cost function. Hence, a permits policy will always lead to cost savings but it can also lead to higher emissions.

The magnitude of Δ_{ps} depends on the value of the different parameters of the model, but its sign is governed by the key cost parameters v and ρ. If $v = \rho = 0$, for example, the second term of (22) is zero, so $\Delta_{ps} > 0$. But if $v > 0$ and $\rho < 0$, this second term becomes negative and depending on the value of h relative to other parameters, Δ_{ps} may become negative as well (it would be so, for example, for $\sigma_\beta = \sigma_\gamma$, $v = 0$ and $\rho < -c/h$). More generally, one can draw a line $\ell_{\Delta=0}$ in the (v, ρ)-space that solves $\Delta_{ps}(v, \rho) = 0$. These results can be summarized in the following proposition

Proposition 1 *The permits policy can lead to either higher or lower welfare than the standards policy. A necessary condition for permits to be welfare dominated by standards is that either* $v > 0$ *or* $\rho < 0$.

To understand this proposition one must first recognize that a permits program can create "perverse" incentives for shifting output from cleaner to dirtier firms resulting in higher total emissions. The latter could happen under two circumstances: (i) when firms doing more abate-

[20] Note that the welfare difference remains if we consider regulator's aggregate uncertainty.

ment are at the same time decreasing output relative to other firms (i.e, when $v > 0$) and (ii) when highly utilized firms are doing less abatement (i.e., when $\rho < 0$). A low (or negative) value of v, on the other hand, reduces both the effect of the environmental regulation on the firm's output under either policy (see (7) and (13)) and the effect of production cost heterogeneity (i.e., β) on firms' abatement decisions under the permits policy (see (12)). A positive correlation between production and abatement costs (i.e., $\rho > 0$) also makes the permits policy more attractive because when both cost parameters (β and γ) are either simultaneously high or low, output and abatement remain closer to the first-best.

Contrary to what occur when emissions are perfectly monitored, Proposition 1 indicates that neither permits nor standards is the appropriate policy choice in all cases. Because of this ambiguity, there seems to be room for a hybrid policy to improve upon either single-instrument policy. In fact, in the classic problem of prices versus quantities under uncertainty pioneered by Weitzman (1974), Roberts and Spence (1980) shows that a hybrid policy that combines permits with taxes/subsidies is always superior to either permits or taxes/subsidies alone. The reason for this result is that in Weitzman (1974) taxes are always superior in terms of costs (lower expected compliance costs) while permits are always superior in terms of emissions (higher expected environmental benefits). In our problem, however, permits are always superior in terms of costs but standards are not always superior in terms of emissions, so it remains to be seen whether and when a hybrid policy would provide a net welfare gain.

4.2 The hybrid policy

The regulator's design problem here is to find the allocation $\tilde{e}_0^h$ of permits (together with an utilization factor $\tilde{q}^h$) and an emission standard x_s^h that maximize social welfare. The superscript "h" stands for hybrid policy. Under this hybrid policy, firms are free to trade permits but each of them cannot reduce less than x_s^h. As in the permits-alone policy, there will be a price $R^h\widetilde{q}^h$ (the scaling factor $\widetilde{q}^h$ can be the same as in the permits-alone policy; its actual value is not relevant).

Depending on the values of β and γ, the standard x_s^h will be binding for some firms. Thus, we can divide the universe of firms in two groups: those that comply with the permits policy ($x^h \equiv x_p^h$ and $q^h \equiv q_p^h$) and those that comply with the standards policy ($x^h \equiv x_s^h$ and $q^h \equiv q_s^h$). Note that the latter group of firms must still cover its (proxied) emissions with permits, which will ultimately affect the equilibrium price of permits. The values of β and γ of those firms that are in the frontier that divides the two groups of firms satisfy $x_p^h = x_s^h$. From the first order condition (12), we know that

$$x_p^h = \frac{R^h\widetilde{q}^h c - \gamma c - (P-\beta)v}{\Lambda}$$

which combined with $x_p^h = x_s^h$ gives us such frontier

$$\gamma = \frac{v}{c}\beta - \frac{\Lambda}{c}x_s^h + R^h\widetilde{q}^h - \frac{Pv}{c} \equiv \widehat{\gamma}(\beta) \tag{24}$$

Because the initial allocation of permits does not affect behavior, firms with $\gamma \geq \widehat{\gamma}(\beta)$ will behave as in the standard-alone policy and firms with $\gamma < \widehat{\gamma}(\beta)$ will behave as in the a permits-alone policy. Then, the regulator's problem can be written as

$$\max_{x_s^h, \widetilde{e}_0^h/\widetilde{q}^h} \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}(\beta)} \left[Pq_p^h - C(q_p^h, x_p^h, \beta, \gamma) - h(1-x_p^h)q_p^h\right] F_{\beta\gamma} d\gamma d\beta +$$
$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}(\beta)}^{\overline{\gamma}} \left[Pq_s^h - C(q_s^h, x_s^h, \beta, \gamma) - h(1-x_s^h)q_s^h\right] F_{\beta\gamma} d\gamma d\beta \tag{25}$$

where q_s^h, x_p^h and q_p^h can be obtained, respectively, from (7), (12) and (13) (simply add the superscript "h"). The regulator ensures full compliance, so the market clearing condition is (all firms must cover their emissions with permits)

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}(\beta)} (1 - x_p^h)\widetilde{q}^h F_{\beta\gamma} d\gamma d\beta + \int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}(\beta)}^{\overline{\gamma}} (1 - x_s^h)\widetilde{q}^h F_{\beta\gamma} d\gamma d\beta = \widetilde{e}_0^h$$

Since $\widetilde{q}^h$ is a constant and x_p^h depends on $R^h\widetilde{q}^h$, it is clear again that it is irrelevant to solve for either the $R^h\widetilde{q}^h$ or $\widetilde{e}_0^h/\widetilde{q}^h$. As before, we solve for $R^h\widetilde{q}^h$. Following the derivations for the single-instrument policies, the first order conditions for the optimal hybrid policy are (we allow for corner solutions)

$$\begin{aligned} R^h\widetilde{q}^h : & \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}(\beta)} \left[-(1 - x_p^h)h\frac{\partial q_p^h}{\partial(R^h\widetilde{q}^h)} + hq_p^h\frac{\partial x_p^h}{\partial(R^h\widetilde{q}^h)} - R^h\widetilde{q}^h\frac{\partial x_p^h}{\partial(R^h\widetilde{q}^h)} \right] F_{\beta\gamma} d\gamma d\beta \\ & + \int_{\underline{\beta}}^{\overline{\beta}} \left[Pq_p^h - C(q_p^h, x_p^h, \beta, \widehat{\gamma}(\beta)) - h(1 - x_p^h)q_p^h \right] F_\beta(\beta, \widehat{\gamma}(\beta)) d\beta \\ & - \int_{\underline{\beta}}^{\overline{\beta}} \left[Pq_s^h - C(q_s^h, x_s^h, \beta, \widehat{\gamma}(\beta)) - h(1 - x_s^h)q_s^h \right] F_\beta(\beta, \widehat{\gamma}(\beta)) d\beta \leq 0 \\ & [= 0 \text{ if } R^h\widetilde{q}^h > 0; < 0 \text{ if } R^h\widetilde{q}^h = 0] \end{aligned} \tag{26}$$

$$\begin{aligned} x_s^h : & \int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}(\beta)}^{\overline{\gamma}} \left[-kx_s^h - \gamma - vq_s^h - (1 - x_s^h)h\frac{\partial q_s^h}{\partial x_s^h} + hq_s^h \right] F_{\beta\gamma} d\gamma d\beta \\ & - \frac{\Lambda}{c} \int_{\underline{\beta}}^{\overline{\beta}} \left[Pq_p^h - C(q_p^h, x_p^h, \beta, \widehat{\gamma}(\beta)) - h(1 - x_p^h)q_p^h \right] F_\beta(\beta, \widehat{\gamma}(\beta)) d\beta \\ & + \frac{\Lambda}{c} \int_{\underline{\beta}}^{\overline{\beta}} \left[Pq_s^h - C(q_s^h, x_s^h, \beta, \widehat{\gamma}(\beta)) - h(1 - x_s^h)q_s^h \right] F_\beta(\beta, \widehat{\gamma}(\beta)) d\beta \leq 0 \\ & [= 0 \text{ if } x_s^h > 0;\ < 0 \text{ if } x_s^h = 0] \end{aligned} \tag{27}$$

Since at $\widehat{\gamma}(\beta)$ we have that $x_s^h = x_p^h$ and hence, $q_s^h = q_p^h$, (26) and (27) reduce, respectively, to

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}(\beta)} \left[-(1 - x_p^h) h \frac{\partial q_p^h}{\partial x_p^h} + h q_p^h - R^h \widetilde{q}^h \right] F_{\beta\gamma} d\gamma d\beta \leq 0 \tag{28}$$

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}(\beta)}^{\overline{\gamma}} \left[-k x_s^h - \gamma - v q_s^h - (1 - x_s^h) h \frac{\partial q_s^h}{\partial x_s^h} + h q_s^h \right] F_{\beta\gamma} d\gamma d\beta \leq 0 \tag{29}$$

where $\partial q_p^h / \partial x_p^h = \partial q_s^h / \partial x_s^h = -v/c$.

The solution of the optimal hybrid policy can result in a corner solution if either (28) or (29) is negative at the optimum. When (28) is negative, the hybrid policy converges to the standards-alone policy (i.e., $R^h \widetilde{q}^h = 0$ and $x_s^h = x_s$) and when (29) is negative, the hybrid policy converges to the permits-alone policy (i.e., $R^h \widetilde{q}^h = R\widetilde{q}$ and $x_s^h = 0$). Formal exploration of these possibilities lead to the following proposition

Proposition 2 *The optimal hybrid policy converges to the optimal permits-alone policy for many combinations of v and ρ (e.g., $v = \rho = 0$). If abatement cost heterogeneity is significantly smaller than production cost heterogeneity (more precisely, if $-\underline{\gamma}/\overline{\beta} < (h - v)/c$), the optimal hybrid policy converge to the optimal standards-alone policy for $\rho = -1$.*

Proof. See the Appendix. ■

Contrary to what happens in Roberts and Spence (1980) and Kwerel (1980), in this model the combination of instruments does not necessarily leads to higher welfare. In those papers, taxes are superior in terms of costs (i.e., lower expected costs) but permits are superior in terms of benefits (i.e., lower expected emissions),[21] so an optimal combination of instruments strictly dominates either single instrument. Here, permits are not only less costly than standards but also they can lead to lower emissions, so it is reasonable to think that in such cases standards

[21] See also Weitzman (1974).

may add nothing to the policy design.[22] It is not so obvious, however, that in some cases, although very few, permits may also add nothing to the policy design.

The exact shape of the region in which the hybrid policy dominates either single-instrument policy depends on the parameter values. A simple numerical exercise may be useful. In Figure 1, line $\ell_{h=p}$ indicates the combinations of v and ρ for which the hybrid policy just converges to the permits policy for the following parameters values: $P = k = c = 4$, $h = 2$, $\overline{\beta} = 2$, $\underline{\beta} = -2$, $\overline{\gamma} = 1$, $\underline{\gamma} = -1$.[23] The figure also includes the line $\ell_{\Delta=0}$ (i.e., combinations of v and ρ that yield $\Delta_{ps} = 0$) and the line $\ell_{\Delta E=0}$ (i.e., combinations of v and ρ for which the permits policy and the standards policy yield the exact same level of emissions). One can distinguish three regions in the figure. To the left of $\ell_{h=p}$, there are those combinations for which the hybrid policy coincides with the permits policy. As the first row of Table 1 shows, if $v = -0.5$ and $\rho = 0.6$, for example, social net benefits (W) are 33% higher under the permits policy than under the standards policy. Note also that in some places of this region the hybrid policy does not improve upon the permits-alone policy despite the fact that emissions are higher than under a standards-alone policy. The logic behind this result is that the introduction of some binding standard (in combination with permits) would not only reduce emissions but also increase production and abatement costs. And in this particular region, the latter effect dominates.

The second region —between the lines $\ell_{h=p}$ and $\ell_{\Delta=0}$— includes all those combinations for which the hybrid policy is superior to the permits-alone policy, which in turn, is superior to the standards-alone policy. For example, if $v = 0.6$ and $\rho = 0.6$, the welfare gain from implementing the hybrid policy (Δ_h) is 12.6% of Δ_{ps}, as shown in the second row of the table. It is interesting to observe that despite welfare may not increase by much, policy designs are

[22] A similar situation would occur in the analysis of permits and taxes if the slope of the marginal benefit curve were assumed negative instead of positive, as normally the case. Taxes would be superior to permits in terms of both cost and benefits and there would be no gains from adding permits into the tax design.

[23] The simulation is carried out with only four type of firms: $(\overline{\beta}, \overline{\gamma})$,$(\overline{\beta}, \underline{\gamma})$, $(\underline{\beta}, \overline{\gamma})$ and $(\underline{\beta}, \underline{\gamma})$. Also, the value of the different parameters limit the range of v to $[-0.5, 0.7]$.

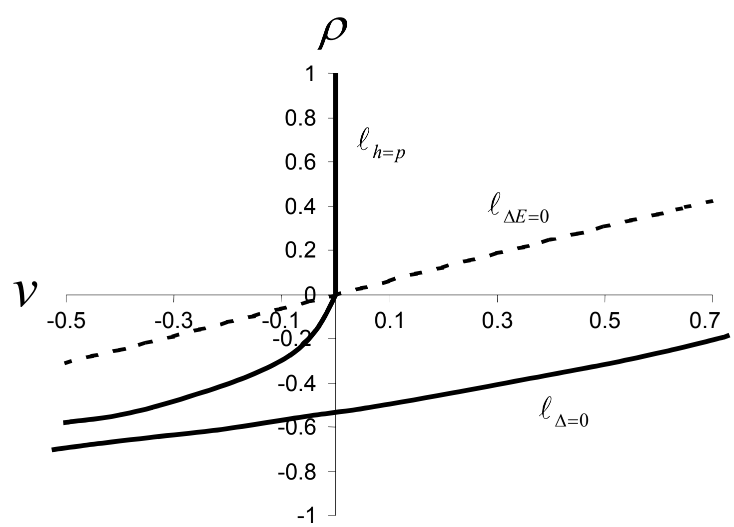

Figure 1: Hybrid and single-instrument policies

quite different (the hybrid policy includes a standard that is almost half the one in the standards-alone policy; though the equilibrium permit price do not vary much). Finally, the third region —to the right of $\ell_{\Delta=0}$— includes those combinations of v and ρ for which the hybrid policy is welfare superior to the standards-alone policy, which in turn, is superior to the permits-alone policy. Here, the gain from implementing the hybrid policy as opposed to the standards-alone policy is substantial, 32.5% of $|\Delta_{ps}|$.[24]

Table 1. Hybrid and single-instrument policies: design and welfare

v	ρ	x_s	$R\widetilde{q}$	x_s^h	$R^h\widetilde{q}^h$	W_s	Δ_{ps}	Δ_h
-0.5	0.6	0.65	2.08	0	2.08	123.64	41.08	0
0.6	0.5	0.38	2.07	0.18	1.99	82.04	13.66	1.72
0.7	-0.5	0.36	2.10	0.21	1.49	79.74	-6.37	2.07

[24] Note that despite that $\sigma_\gamma = 0.5\sigma_\beta$, there is no region in Figure 1 where the hybrid policy converges to the standards-alone policy.

5 An empirical evaluation

In this section, I use the experience from Santiago's total suspended particulate emissions (TSP) trading program to evaluate the advantages (if any) of using permits when emissions are imperfectly observed. Because firms are not required to provide the regulator with information on production and abatement costs, I apply the theoretical framework previously developed (including the assumption of a benevolent regulator given his choice of h) to infer the cost structure of the firms affected by the TSP program and other parameters such as h and P based on observables such as emission rates and utilization. This information is then used to compare the actual performance of the TSP permits program with the performance of a hypothetical and equivalent standards policy (and with a hybrid policy). The equivalent standards policy is constructed under the assumption that if the regulator would have to introduce a standard he will do it optimally using the same value of h (together with the other parameters) that he used to implement the permits policy.[25]

5.1 The TSP program

The city of Santiago has constantly presented air pollution problems since the early 1980s. The TSP trading program, established in March of 1992, was designed to curb TSP emissions from the largest stationary sources in Santiago (industrial boilers, industrial ovens, and large residential and commercial heaters) whose emissions are discharged through a duct or stack at a flow rate greater than or equal to 1,000 m^3/hr. Because sources were too small to require sophisticated monitoring procedures, the authority did not design the program based on sources' actual emissions but on a proxy variable equal to the maximum emissions that a source could emit in a given period of time if it operates without interruption.

[25] Allowing for a regulator with objective functions and parameter values that dependen on the instrument under consideration introduces new elements to the policy analysis that go well beyond the scope of this paper.

The proxy for emissions (expressed in kg of TSP per day) used by the authority in this particular program was defined as the product of emissions concentration (in mg/m^3) and flow rate (in m^3/hrs) of the gas exiting the source's stack (multiplied by 24 hrs and 10^{-6} kg/mg to obtain kg/day).[26] Although the regulatory authority monitors each affected source's concentration and flow rate once a year,[27] emissions $\widetilde{e}$ and permits $\widetilde{e}_0$ are expressed in daily terms to be compatible with the daily TSP air quality standards. Thus, a source that holds one permit has the right to emit a maximum of 1 kg of TSP per day indefinitely over the lifetime of the program.

Sources registered and operating by March 1992 were designated as existing sources and received grandfathered permits equal to the product of an emissions rate of 56 mg/m^3 and their flow rate at the moment of registration. New sources, on the other hand, receive no permits, so must cover all their emissions with permits bought from existing sources. The total number of permits distributed (i.e., the emissions cap) was 64% of aggregate (proxied) emissions from existing sources prior to the program. After each annual inspection, the authority proceeds to reconcile the estimated emissions with the number of permits held by each source (all permits are traded at a 1:1 ratio). Note that despite permits are expressed in daily terms, the monitoring frequency restricts sources to trade permits only on an annual or permanent basis.[28]

[26] In terms of our model, this is equivalent as to make $\widetilde{q}$ equal to the maximum possible output, which in our case is $(P - \underline{\beta})/c$. But note that the program would have worked equally well with an either higher or lower $\widetilde{q}$. The use of a different $\widetilde{q}$ only requires to adjust the number of quasi-permits $\widetilde{e}_0$ to be distributed such that $R\widetilde{q}$ remains at its optimal level.

[27] There are also random inspections to enforce compliance throughout the year.

[28] In addition, the authority introduced an emission rate standard of 112 mg/m^3 (as in the hybrid design) to be comply by all stationary sources. It is not clear in the case of the TSP sources, however, whether this standard was actually enforced or whether sources could simply buy extra permits. By 1997 several sources (29 out of 576) are above the standard, some of which with earlier emission rates below the standard.

5.2 The data

The data for the study was obtained from PROCEFF's databases for the years 1993 through 1999.[29] Each database includes information on the number of sources and their dates of registration, flow rates, fuel types, emission rates and utilization (i.e., days and hours of operation during the year). While information on flow rates, fuel types and emission rates is directly obtained by the authority during its annual inspections, information on utilization is obtained from firms' voluntary reports.[30] The 1993 database contains all the information, including the flow rate used to calculate each source's allocation of permits, before the program became effective in 1994. Table 2 presents a summary of the data. The first two rows show the proportion of existing and new sources.[31]

[INSERT TABLE 2 HERE OR BELOW]

The next rows of Table 2 provide information on the evolution of flow rates, emission rates and utilization. The large standard deviations indicate that these three variables vary widely across sources in all years.[32] In order to comply with the TSP trading program, affected sources can hold permits, reduce emissions or do both. They can reduce emissions by either decreasing their size (i.e., flow rate) or their emission rates, through either fuel switching (for example, from wood, coal, or heavy oil to light oil, liquid gas, or natural gas) or the installation of end-of-pipe technology such as filters, electrostatic precipitators, cyclones, and scrubbers.[33] Sources

[29] PROCEFF is the government office responsible for enforcing the TSP program.

[30] Since utilization has no effect at all on the source's compliance status, there is no reason to beleive that firms have incentives to misreport their true utilization. For the same reason, this information is available for most but not all sources.

[31] It is interesting to point out that by 1999 36% of the affected sources were new sources despite the fact they did not receive any permits.

[32] It may seem strange to observe some flow rates below the 1,000 (m^3/hr) mark. In general, these are existing sources for which their flow rates were wrongly estimated above 1,000 (m^3/h) at the time of registration. Nevertheless, these sources chose to remain in the program to keep the permits they had already received.

[33] Note that for most sources their flow rates do not change over time.

do not gain anything, in terms of emissions reduction, by changing their utilization level (i.e., days and hrs of operation), because by definition it is assumed to be at 100%. Given that the authority controls for the size of the source (i.e., flow rate) at the moment of permits allocation and monitoring, in terms of our theoretical model changes in emission rates is captured by x_p and utilization by q_p.

The last two rows of Table 2 show data on total emissions and permits.[34] Although 1994 was in principle the first year of compliance with the program, trading activity did not occur until 1996 when compliance was more effectively enforced (Montero et al, 2002). The emissions goal of the TSP program was only achieved by 1997 (total emissions below total permits);[35] year after which natural gas became available from Argentina at unexpectedly attractive prices that many affected sources switched to this cleaner fuel leaving the cap of 4,087.5 permits largely unbinding.[36] Consequently, the empirical evaluation that follows is mainly based on the 1997 data and to a lesser extent on the 1998 data.[37]

5.3 Preliminary estimation of Δ_{ps}

Before attempting an estimate of each of the parameters of (23), it is useful to derive first an alternative (and simpler) expression to estimate Δ_{ps}. From (17) and (18) it is immediate that

$$\text{Var}[x_p] = \frac{1}{\Lambda^2}(v^2\sigma_\beta^2 + c^2\sigma_\gamma^2 - 2vc\rho\sigma_\beta\sigma_\gamma) \tag{30}$$

[34] A few permits were retired from the market in 1997 as the authority revised the eligibility of some sources to receiving permits (Montero et al., 2002).

[35] The fact that aggregate quasi-emissions in 1997 are somewhat below the cap should not be interpreted as either overcompliance or unbinding regulation. One explanation is that firms tend to hold a few extra permits as an insurance against some measurement uncertainty (inherent to a monitoring precedure of this sort). A second explanation is the uncertainty associated to the revision of the initial allocation of permits carried out by the authority since the beginning of the program. The 1997 allocation drop is, in fact, the result of such revision.

[36] This is consistent with the fact that inter-firm trading activity stopped by mid 1998.Obviously, intra-firm trading activity has continued as new sources come into operation.

[37] It should be mentioned that in addition to the TSP program the authority introduced an emission rate standard of 112 mg/m^3 (as in the hybrid design). It seems, however, that this either was only enforced by 1998 or became unbinding after the arrival of natural gas.

$$\mathrm{Cov}[x_p, q_p] \equiv \mathrm{E}[x_p q_p] - \mathrm{E}[x_p]\mathrm{E}[q_p] = \frac{1}{\Lambda^2}[(kc + v^2)\rho\sigma_\beta\sigma_\gamma - kv\sigma_\beta^2 - vc\sigma_\gamma^2] \tag{31}$$

Note that (31) is the negative value of the difference in total emissions, E, between the permits and the standards policy, that is $E_p - E_s = -\mathrm{Cov}[x_p, q_p]$. In turn, eqs. (30) and (31) allow us to re-write $\Delta_{ps} = A_1\sigma_\beta^2 + A_2\sigma_\gamma^2 + A_3\rho\sigma_\beta\sigma_\gamma$ as

$$\Delta_{ps} = \frac{\Lambda}{2c}\mathrm{Var}[x_p] + h\mathrm{Cov}[x_p, q_p] \tag{32}$$

As in (22), the first term in (32) is the difference in costs savings between the permits and the standards policies, which is always positive, while the second term, $h\mathrm{Cov}[x_p, q_p]$, is the difference in environmental benefits.

However, (32) cannot be immediately applied to our data because of the wide heterogeneity in emission rates before the TSP program was implemented. While this heterogeneity does not affect the permits policy design (and firms' behavior with regard to x_p and q_p), it does affect the standards policy design in that the regulator now sets an emission rate standard r_s instead of the reduction rate standard x_s. Depending of their counterfactual emission rates (i.e., rates before the regulation), some firms will reduce more than others.

Heterogeneity in counterfactual rates can be easily incorporated into the theoretical model by assigning to each firm some $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ and letting $e = rq$, where $r = 1 + \alpha - x > 0$ is the emissions rate, $1 + \alpha$ is the counterfactual rate and $\mathrm{E}[\alpha] = 0$. Since it is likely to be cheaper to reduce emissions for sources with high counterfactual emission rates, it is reasonable to expected a strong (negative) correlation between α and γ (to be confirmed in the econometrics results below). Introducing these extensions to the model and allowing possible correlations between α and β, and between α and γ, the welfare difference between the permits policy and the (optimal) standard r_s is[38]

$$\Delta_{ps} = \frac{\Lambda}{2c}\text{Var}[r_p] - h\text{Cov}[r_p, q_p] \tag{33}$$

where $r_p = (1+\alpha-x_p)$ is the observed emissions rate under the permits policy. Note that now $E_p - E_s =$Cov$[r_p, q_p]$.

The sign of (33) the can then be explored by looking at the covariance matrix for the emission rate (r_p) and utilization (q_p). Using the flow rate as a weight to control for size differences across sources, the weighted statistics for the 1997 data (499 obs.) are Var$[q_p] = 0.112$, Var$[r_p] = 0.211$ and Cov$[r_p, q_p] = 0.026$ (to work with dimensionless variables hereafter, emission rates are divided by its 1993 mean value of 94.9 mg/m^3)[39] and for the 1998 data (543 obs.) the weighted statistics are, respectively, 0.111, 0.056 and 0.005. Although these figures do not allows us to sign Δ_{ps} yet, they indicate that emissions has been somewhat larger than what would have been under an equivalent standards policy. Since $E_p =$E$[r_p q_p]$ and the weighted value of E$[r_p q_p]$ in 1997 is 0.445, emissions would have been 0.419 under the equivalent standard of 0.663 (the latter is the weighted value of E$[r_p]$ in 1997).

The 1997 figures also show that Var$[r_p]$ is more than eight times larger than Cov$[r_p, q_p]$, raising the possibility that the higher emissions may be more than offset by cost savings. To test for this possibility, however, more information on various parameters is required.

5.4 Estimation of parameters

A more precise estimate of Δ_{ps} requires values of v, c, k, h and P; information that is to be recovered from the data described in Table 2 (no detailed information on production and abatement costs is available elsewhere; at least to my knowledge). I start with the estimation

[38] The optimal emission rate standard is $r_s = 1 - x_s$, where x_s is as in (9). Note that now the actual reduction under the standards policy is $1 - r_s + \alpha$.

[39] The unweighted statistics are, respectively, 0.101, 0.221 and 0.004.

of v. Based on first order conditions (10) and (11), v is obtained by estimating the following simultaneous-equation system

$$REDUC_i = a_0 + a_1 UTIL_i + a_2 FLOW93_i + a_3 EMRTE93_i + a_4 ENDPIPE_i + a_5 INDUST_i + a_6 STATE_i + \varepsilon_i \quad (34)$$

$$UTIL_i = b_0 + b_1 REDUC_i + b_2 UTIL93_i + a_2 FLOW93_i + b_4 INDUST_i + b_5 STATE_i + u_i \quad (35)$$

where i indexes sources, ε^i and u^i are error terms whose characteristics will be discussed shortly, and the different variables relate to those in (10)–(11) as follows. $REDUC$ corresponds to x_p, i.e., the level of reduction under the permits policy. $REDUC$ is calculated as the difference between the source's counterfactual emission rate $(1+\alpha)$ and its actual emission rate (r_p).[40] I use the 1993 as the counterfactual year,[41] so $EMRTE93$ is the counterfactual emissions rate.

The variable $UTIL$ corresponds to q_p, i.e., the level of utilization or output. As in the theoretical model, the TSP program's authority does not observe $UTIL$, and therefore, he cannot use it for monitoring and enforcement purposes. Put it differently, because the regulator only observes source's flow rate and emissions rate, he has only control over changes in emissions due to changes in source's size (i.e., flow rate) and emission rates but not over changes in emissions due to changes in utilization.

The variables $FLOW93$, $EMRTE93$, $ENDPIPE$, $INDUST$ and $STATE$ included in (34) are intended to capture differences in abatement costs across sources (i.e., γ). $FLOW93$ is the

[40] Recall that emission rates are normalized by the 1993 mean.

[41] Results do not qualitatively change when I use 1995 as the counterfactual year; year in which I have a few more data points.

source's flow rate in 1993. If there are any scale economies associated with pollution abatement we should expect more abatement from bigger sources (i.e., larger $FLOW93$), other things equal (I also use $FLOW93^2$ and $\ln FLOW93$).[42] Similarly, I expect that a source with a high emission rate before the TSP program (i.e., high $EMRTE93$) should face more abatement possibilities and hence lower costs. Conversely, I expect a source already equipped with some end-of-pipe abatement technology required by previous (and source specific) regulation to be less likely to reduce emissions. Hence, I introduce the dummy variable $ENDPIPE$ that equals 1 if the source has any type of end-of-pipe abatement technology by 1993. I also introduce the dummy variables $INDUST$ and $STATE$ to see whether there is any difference in abatement costs (or abatement behavior) between industrial sources ($INDUST = 1$) and residential/commercial sources, and between state or municipality owned sources ($STATE = 1$) and privately owned sources.[43]

The variables $UTIL93$, $FLOWRTE93$, $INDUST$ and $STATE$ included in (35) are intended to capture differences in production costs across sources (i.e., β). $UTIL93$ is the source's utilization in 1993 and serves as a proxy for the level of utilization that would have been observed in the absence of the TSP program and of changes in exogenous factors (e.g., input prices, demand, etc.).[44] Since, on average, utilization has been increasing overtime, $FLOW93$ should capture whether expansion in larger units is relatively cheaper than in smaller units. For the same reason, I also include $INDUST$ and $STATE$.

An estimate of the sign (and relative value) of v can then be inferred from either $a_1 = -v/k$ or $b_1 = -v/c$. Since $UTIL$ and $REDUC$ enter as endogenous variables in (34)–(35),

---

[42] I use the 1993 flow rate instead of the actual flow rate to control for possible endogeneity problems. However, results are virtually the same when I use the actual flow rate. This is in part because the firm's flow rate barely change overtime (the drop in average flow rates shown in Table 2 is mainly due to changes in one particular large firm).

[43] For example, $INDUST = 0$ and $STATE = 1$ for the boiler of the central heating system of a public hospital.

[44] To work with a larger dataset I use the 1995 utilization for 66 sources. This should not baised the results in any particular way since the TSP program was not effectively enforced until 1996 (see Table 2).

however, their correlations with the error terms ε_i and u_i would produce biased OLS estimators. Therefore, I employ a two-stage least squares (2SLS) estimation procedure to obtain unbiased estimates. 2SLS results for equations (34) and (35) are presented in Table 3 (first-stage results are omitted).

[INSERT TABLE 3 HERE OR BELOW]

The first three columns of Table 3 show the results for the 1997 data. Results in column (1) indicate that the coefficients of $UTIL$ and $REDUC$ (i.e., a_1 and b_1, respectively), although positive, are not significantly different from zero. Because our theoretical model assumes that all firms are expected to produce, on average, the same amount of output ($\text{E}[q_p] = (P - vx_s)/c$), however, one can argue that these coefficients may provided a biased estimation of v by not taking into account the fact that firms are of different sizes. One could further argue that the true value of v may be even of different sign because of the coefficients of $FLOW93$ and $FLOW93^2$ in the reduction equation indicate that reduction decreases with size throughout the relevant range. To control for such possibility, I run a weighted 2SLS regression using the 1997 flow rate as weight. The new estimates, which are reported in columns (2) and (3), are not very different from the unweighted estimates, confirming that the interaction term v in equation (1) is not statistically different from zero.

The last three columns of Table 3 show the 2SLS results for the 1998 data [weighted estimates are in columns (5) and (6)]. In particular, we observe that the coefficients of $UTIL$ and $REDUC$ in column (5) are positive and significantly different from zero at the 10% level. This negative value of v can be attributed in large part to the arrival of natural gas at relatively low prices by the end of 1997.[45] Although the 1998 results must be carefully interpreted because of the

[45] In fact, 112 the 144 affected sources that switched to natural gas in 1998 increased or maintained its utilization relative to 1997.

apparently slack cap, they prove to be useful to illustrate the estimation of Δ_{ps} when v is different than zero, as we shall see next.

We can now use the value of v to obtain an estimate for the remaining parameters of the model, and hence, for Δ_{ps}. Following the 1997 econometric results, let us consider first the case in which $v = 0$. When this is the case, we have that $h = cR\widetilde{q}/P$ from (16), $R\widetilde{q} = k\mathrm{E}[x_p]$ from (12), and $P = c\mathrm{E}[q_p]$ from (13). Replacing the 1997 (weighted) statistics for $\mathrm{E}[x_p] = 0.203$ and $\mathrm{E}[q_p] = 0.631$ in the expression for h, (33) reduces to

$$\Delta_{ps}|_{97,v=0} = \frac{k}{2}\mathrm{Var}[r_p] - \frac{k\mathrm{E}[x_p]}{\mathrm{E}[q_p]}\mathrm{Cov}[r_p, q_p] = (0.1055 - 0.0084)k = 0.097k$$

These numbers not only indicate that the permits policy is welfare superior to an equivalent standards policy but more importantly, that the welfare loss from higher emissions is only 8% of the welfare gain from cost savings.

Based on the 1998 results contained in column (5) of Table 3, let us consider now the case in which $v < 0$. Normalizing v to $-a$ (where a is some positive number), from the coefficients of $UTIL$ and $REDUC$ we obtain, respectively, $k = 1.86a$ and $c = 15.87a$ (which in turn, yields $\Lambda = 28.52a^2$). In addition, by simultaneously solving (12) and (13) for P and $R\widetilde{q}$ with $\mathrm{E}[x_p] = 0.466$ and $\mathrm{E}[q_p] = 0.669$, we get $P = 10.15a$ and $R\widetilde{q} = 0.20a$ that replaced into (16) gives $h = 0.31a$. Plugging these numbers and the corresponding statistics into (33), we finally obtain $\Delta_{ps}|_{98,v<0} = (0.0503 - 0.0016)a = 0.049a$. This result, while qualitatively similar to the 1997 result, shows an even smaller welfare loss from higher emissions —only 3% of the welfare gain from cost savings.

5.5 Gains from a hybrid policy

To explore the design of and the gains from implementing a hybrid policy, we need, in addition to the previous parameters, some idea about the properties of the cumulative joint distribution $F(\alpha, \beta, \gamma)$. Since we cannot recover $F(\alpha, \beta, \gamma)$ from the data, I will simply estimate $\boldsymbol{\sigma} = (\sigma_\alpha, \sigma_\beta, \sigma_\gamma)$ and $\boldsymbol{\rho} = (\rho_{\alpha\beta}, \rho_{\alpha\gamma}, \rho_{\beta\gamma})$ and then impose a multinomial distribution of 12 types consistent with such estimates.[46] Values for $\boldsymbol{\sigma}$ and $\boldsymbol{\rho}$ are obtained by simultaneously solving $\text{Var}[x_p]$, $\text{Var}[r_p]$, $\text{Var}[q_p]$, $\text{Cov}[x_p, r_p]$, $\text{Cov}[x_p, q_p]$ and $\text{Cov}[r_p, q_p]$ (see, e.g., eqs. (30) and (31)). The solution is of the form $\boldsymbol{\sigma}(c, k, v)$ and $\boldsymbol{\rho}(c, k, v)$ that combined with the previous estimates of c, k and v allows us to obtain $\boldsymbol{\sigma} = (\sigma_\alpha, \sigma_\beta, \sigma_\gamma)$ and $\boldsymbol{\rho} = (\rho_{\alpha\beta}, \rho_{\alpha\gamma}, \rho_{\beta\gamma})$. Note that if $v = 0$, $\boldsymbol{\sigma}$ remains a function of k and c, which cannot be independently estimated. Conversely, if $v \neq 0$, we can leave $\boldsymbol{\sigma}$ and all the other parameters of the model (i.e., k, c, P and h) as a function of v, as we did in the previous section for the 1998 numbers. Thus, instead of choosing any arbitrary combination of k and c, I use the results of column (2) of Table 3 and set $c/k = 25.35$ for 1997.

The results for 1997 are $\boldsymbol{\sigma}/k = (0.63, 8.48, 0.60)$ and $\boldsymbol{\rho} = (-0.04, -0.72, -0.08)$, and for 1998 are $\boldsymbol{\sigma}/a = (0.63, 5.34, 1.29)$ and $\boldsymbol{\rho} = (0.11, -0.90, -0.39)$, where $a = -v$. Looking at Figure 1 and the 1997 numbers ($v = 0$ and $\rho_{\beta\gamma} < 0$), there seems to be room, although little, for extra benefits from the hybrid policy.[47] However, the heterogeneity in counterfactual emission rates and the negative correlation between α and γ (consistent with the coefficient of $EMRTE93$) shift the line $\ell_{h=p}$ in Figure 1 to the right leaving the cost structure of the TSP firms within the region in which the hybrid policy converges to the permits policy.[48] I also find for 1998

[46] The types are $(\overline{\alpha}, \overline{\beta}, 0)$, $(\overline{\alpha}, \underline{\beta}, 0)$,..., $(0, \underline{\beta}, \underline{\gamma})$. If anything this distribution would favor the introduction of a hybrid policy because it puts more weight on extreme combinations of α, β and γ.

[47] In fact, if we impose $\sigma_\alpha = 0$, and hence $\rho_{\alpha\beta} = \rho_{\alpha\gamma} = 0$, the hybrid policy design is $r_s^h = 1.62$ (or 153 mg/m^3) and $R^h \tilde{q}^h = 0.97 R\tilde{q}$ (the equivalent standards is $r_s = 0.88$) and the increase in net benefits over Δ_{ps} is less than 1%.

[48] A negative $\rho_{\alpha\gamma}$ reduces the advantages of introducing a standard because many of the sources that are supposed to be hit by the standard (those with high α) are likely to be already reducing under the permits-alone policy.

that the hybrid policy converges to the permits-alone policy.

6 Conclusions

I have developed a model to study the design and performance of pollution markets (i.e., tradable permits) when the regulator has imperfect information on firms' emissions and costs. A salient example is the control of air pollution in large cities where emissions come from many small (stationary and mobile) sources for which continuous monitoring is prohibitely costly. In such a case the cost-savings superiority of permits over the traditional command and control approach of setting technology and emission standards is no longer evident. Since the regulator only observes a firm's abatment technology but neither its emissions nor its output (utilization), permits could result in higher emissions if firms doing more abatement are at the same time reducing output relative to other firms and/or if more highly utilized firms find it optimal to abate relatively less. Because of this trade-off between cost savings and possible higher emissions, I also examined the advantages of a hybrid policy that optimally combines permits and standards. I found that in many cases the hybrid policy converges to the permits-alone policy but it almost never converges to the standards-alone policy.

Having studied the theoretical advantages of pollution markets (whether alone or in combination with a standard), I then used emissions and output data from Santiago-Chile's TSP permits program to explore the implications of the theoretical model. I found that the production and abatement cost characteristics of the sources affected by the TSP program are such that the permits policy is welfare superior. The estimated cost savings are only partially offset (about 8%) by a moderate increase in emissions relative to what would have been observed under an equivalent standards policy. In addition, I found no gain from implementing a hybrid policy for this particular group of sources.

Since sources under the TSP program are currently responsible for less than 5% of total TSP emissions in Santiago, the model developed here can be used to study how to expand the TSP program to other sources of TSP that today are subject to command and control regulation. A good candidate would be powered-diesel buses which are responsible for 36.7% of total TSP emissions. According to Cifuentes (1999), buses that abate emissions by switching to natural gas are likely to reduce utilization relative to buses that stay on diesel and that older-less utilized buses are more likely to switch to natural gas. In the understanding that switching to natural gas is a major abatement alternative, both of these observations would suggest that the optimal way to integrate buses to the TSP program is by imposing, in addition to the allocation of permits, an emission standard specific to buses. It may also be optimal to use different utilization factors ($\widetilde{q}$) for each type of source. These and related design issues deserve further research.

References

[1] Cifuentes, L. (1999), Costos y beneficions de introduccir gas natural en el transporte público en Santiago, mimeo, Catholic University of Chile.

[2] Dasgupta, P., P. Hammond and E. Maskin (1980), On imperfect information and optimal pollution control, *Review of Economic Studies* 47, 857-860.

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

[4] Fullerton, D. and S.E. West (2002), Can taxes on cars and on gasoline mimic an unavailable tax on emissions?, *Journal of Environmental Economics and Management* 43, 135-157.

[5] Hahn, R.W., S.M Olmstead and R.N. Stavins (2003), Environmental regulation in the 1990s: A retrospective analysis, *Harvard Environmental Law Review* 27, 377-415.

[6] Harrison, David Jr. (1999), Turning theory into practice for emissions trading in the Los Angeles air basin, in Steve Sorrell and Jim Skea (eds), *Pollution for Sale: Emissions Trading and Joint Implementation*, Edward Elgar, Cheltenham, UK.

[7] Holmström, B. (1982), Moral hazard in teams, *Bell Journal of Economics* 13, 324-340.

[8] Kwerel, E. (1977), To tell the truth: Imperfect information and optimal pollution control, *Review of Economic Studies* 44, 595-601.

[9] Lewis, T. (1996), Protecting the environment when costs and benefits are privately known, *RAND Journal of Economics* 27, 819-847.

[10] Montero, J.-P., J.M. Sánchez, R. Katz (2002), A market-based environmental policy experiment in Chile, *Journal of Law and Economics* XLV, 267-287.

[11] Roberts, M. and M. Spence (1976), Effluent charges and licenses under uncertainty, *Journal of Public Economics* 5, 193-208.

[12] Schmalensee, R, P. Joskow, D. Ellerman, J.P. Montero, and E. Bailey (1998), An interim evaluation of sulfur dioxide emissions trading, *Journal of Economic Perspectives* 12, 53-68.

[13] Segerson, K. (1988), Uncertainty and incentives for nonpoint pollution control, *Journal of Environmental Economics and Management* 15, 87-98.

[14] Spulber, D. (1988), Optimal environmental regulation under asymmetric information, *Journal of Public Economics* 35, 163-181.

[15] Stavins, R. (2003), Experience with market-based environmental policy instruments, in *Handbook of Environmental Economics*, Karl-Göran Mäler and Jeffrey Vincent eds., Amsterdam: Elsevier Science.

[16] Tietenberg, T (1985), *Emissions Trading: An Exercise in Reforming Pollution Policy*, Resources for the Future, Washington, D.C.

[17] Weitzman, M. (1974), Prices vs quantities, *Review of Economic Studies* 41, 477-491.

Appendix: Proof of Proposition 2

The proof is divided in four parts. The first three parts demonstrate that there exist several combinations of v and ρ for which the hybrid policy converges to the permits-alone policy (note that a more general proof would require specifying $F(\beta, \gamma)$, but this, in turn, would make the proof particular to the specified distribution function). Part 1 demonstrates so for $v = \rho = 0$. Part 2 demonstrates that for $v = 0$ and $\rho \neq 0$, the hybrid policy converges to the permits-alone policy only if $\rho > 0$. Part 3 demonstrates that for $v \neq 0$ and $\rho = 1$, the hybrid policy always converges to the permits-alone policy if $v < 0$ but not necessarily so if $v > 0$. Part 4, which finishes the proof, demonstrates that for $\rho = -1$ and some values of v the hybrid policy converges to the standards-alone policy if $-\underline{\gamma}/\overline{\beta} < (h - v)/c$.

Part 1. For $\rho = 0$, $F(\beta, \gamma)$ can be expressed as $F^{\beta}(\beta)F^{\gamma}(\gamma)$ where F^{β} and F^{γ} are independent cumulative distribution functions for β and γ respectively (f^{β} and f^{γ} are their corresponding probability density functions). Thus, when $v = \rho = 0$, $\widehat{\gamma}$ is independent of β and the first order conditions (28) and (29) become, respectively

$$\int_{\underline{\beta}}^{\overline{\beta}} \left[hq_p^h - R^h \widetilde{q}^h \right] F^{\gamma}(\widehat{\gamma}) f^{\beta} d\beta \leq 0 \tag{36}$$

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}}^{\overline{\gamma}} \left[-kx_s^h - \gamma + hq_s^h \right] f^\beta(\beta) f^\gamma(\gamma) d\gamma d\beta \leq 0 \tag{37}$$

where $q_p^h = q_s^h = (P - \beta)/c$. Before solving for $R^h \widetilde{q}^h$ and x_s^h, note that from first order condition (11) we have that $-kx_p^h - \gamma + R^h \widetilde{q}^h = 0$ for $\gamma \leq \widehat{\gamma}$ and that $-kx_s^h - \gamma + R^h \widetilde{q}^h < 0$ for $\gamma > \widehat{\gamma}$ (otherwise the standard would not be binding). Since (36) cannot be positive, its solution yields $R^h \widetilde{q}^h = Ph/c = R\widetilde{q}$ (see (16)). On the other hand, developing (37) we obtain

$$\int_{\widehat{\gamma}}^{\overline{\gamma}} \left[-kx_s^h - \gamma + \frac{Ph}{c} \right] f^\gamma(\gamma) d\gamma \leq 0$$

But since $Ph/c = R^h \widetilde{q}^h$ and $-kx_s^h - \gamma + R^h \widetilde{q}^h < 0$, the solution of (37) is a corner solution with $x_s^h = 0$.

Part 2. When $v = 0$ and $\rho \neq 0$, $\widehat{\gamma}$ is independent of β and since $R^h \widetilde{q}^h$ must be necessarily positive (otherwise (28) would be positive), the first order condition (28) reduces to

$$R^h \widetilde{q}^h - \frac{Ph}{c} + \frac{h \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}} \beta F_{\gamma\beta} d\gamma d\beta}{c \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}} F_{\gamma\beta} d\gamma d\beta} = 0 \tag{38}$$

where its last term, which we will denote by ξ to save on notation, is positive (negative) for $\rho < 0$ ($\rho > 0$) and zero otherwise.

On the other hand, the first order condition (29) reduces to

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}}^{\overline{\gamma}} \left[-kx_s^h - \gamma + \frac{Ph}{c} - \frac{h\beta}{c} \right] F_{\gamma\beta} d\gamma d\beta \leq 0 \tag{39}$$

After replacing Ph/c according to (38), (39) can be rewritten as

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}}^{\overline{\gamma}} \left[-kx_s^h - \gamma + R^h \widetilde{q}^h + \xi \right] F_{\gamma\beta} d\gamma d\beta - \frac{h}{c} \int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}}^{\overline{\gamma}} \beta F_{\gamma\beta} d\gamma d\beta \leq 0 \tag{40}$$

where its last term (without the negative sign in front) has the exact opposite sign of ξ. Now, from condition (11) we know that $-kx_s^h - \gamma + R^h\widetilde{q}^h < 0$ for $\gamma > \widehat{\gamma}$, so if $\rho < 0$ there must be an interior solution (i.e., $x_s^h > 0$) because ξ is positive. To see this, let us imagine we increase the standard x_s^h from cero to the point where we reach the very first firm for which there is no difference between complying with the standard and buying permits. At this point $-kx_s^h - \gamma + R^h\widetilde{q}^h = 0$ but $\xi > 0$, so condition (40) would be positive. It would be beneficial then to increase x_s^h a bit further until (40) reaches cero. Conversely, if $\rho > 0$, ξ is negative, so the solution of (40) is indeed a corner solution with $x_s^h = 0$.

Part 3. When $v \neq 0$ and $\rho \neq 0$, it is useful to first to express $q_p^h(R^h\widetilde{q}^h, \beta, \gamma)$, $x_p^h(R^h\widetilde{q}^h, \beta, \gamma)$ and $q_s^h(x_s^h, \beta)$ according to (7), (12) and (13) (simply add the superscript "h") and then substitute them into (28) and (29) to obtain, respectively, the first order conditions

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\gamma}}^{\widehat{\gamma}(\beta)} \left[(R\widetilde{q} - R^h\widetilde{q}^h) \left(\frac{\Lambda + 2hv}{\Lambda} \right) + \frac{2hv}{\Lambda}\gamma - \frac{h(ck + v^2)}{c\Lambda}\beta \right] F_{\gamma\beta} d\gamma d\beta \leq 0 \tag{41}$$

$$\int_{\underline{\beta}}^{\overline{\beta}} \int_{\widehat{\gamma}(\beta)}^{\overline{\gamma}} \left[(x_s - x_s^h) \left(\frac{\Lambda + 2hv}{c} \right) - \gamma - \frac{(h - v)}{c}\beta \right] F_{\gamma\beta} d\gamma d\beta \leq 0 \tag{42}$$

To demonstrate now that the hybrid policy converges to the permits-alone policy if $\rho = 1$ and $v < 0$, it is sufficient to show that the extra benefits from introducing $x_s^h > 0$ that is just binding for a marginal firm are negative when $R^h\widetilde{q}^h = R\widetilde{q}$. For $\rho = 1$ and $v < 0$, the cost parameters of the marginal firm are $\overline{\beta} > 0$ and $\overline{\gamma} > 0$ and the extra benefits are (from (42))

$$(x_s - x_s^h) \left(\frac{\Lambda + 2hv}{c} \right) - \overline{\gamma} - \frac{(h - v)}{c}\overline{\beta} \tag{43}$$

Since this marginal firm is indifferent between permits and standards, i.e., $\widehat{\gamma}(\overline{\beta}) = \overline{\gamma}$, we have from (24) that

$$\overline{\gamma} = \frac{v}{c}\overline{\beta} - \frac{\Lambda}{c}x_s^h + R\widetilde{q} - \frac{Pv}{c} \tag{44}$$

We also know from (9) and (16) that $R\widetilde{q}c = x_s\Lambda + Pv$. Using this to substitute $R\widetilde{q}$ into (44), we obtain an expression for $x_s - x_s^h$ that replaced into (43) leads to

$$\frac{2hv}{\Lambda}\overline{\gamma} - \frac{h(ck + v^2)}{c\Lambda}\overline{\beta} < 0 \tag{45}$$

which demonstrates that the extra benefits are negative.

For $\rho = 1$ and $v > 0$, on the other hand, the marginal firm is still $(\overline{\beta}, \overline{\gamma})$ and the extra benefits are also given by (45). But because $v > 0$, these extra benefits can now be either negative or positive, in which case it is optimal to implement a hybrid policy with $x_s^h > 0$ (using a similar procedure , it can be demonstrated that when $\rho = 1$, the authority always wants to use some permits).

Part 4. Let us now demonstrate that the hybrid policy may converge to the standards-alone policy for $\rho = -1$ and some values of v.[49] As analogous to Part 3, this will be the case if the extra benefits from introducing $R^h\widetilde{q}^h > 0$ that is just binding for a marginal firm are negative when $x_s^h = x_s$. When $\rho = -1$, the cost parameters of the marginal firm are $\overline{\beta} > 0$ and $\underline{\gamma} < 0$ and the extra benefits are (from (41))

$$(R\widetilde{q} - R^h\widetilde{q}^h)\left(\frac{\Lambda + 2hv}{\Lambda}\right) + \frac{2hv}{\Lambda}\underline{\gamma} - \frac{h(ck + v^2)}{c\Lambda}\overline{\beta} \tag{46}$$

Following the same procedure as in Part 3 to obtain an expression for $R\widetilde{q} - R^h\widetilde{q}^h$, (46) reduces to $-\underline{\gamma} - (h - v)\overline{\beta}/c$, which is negative (and hybrid policy converges to standards-alone policy) if $-\underline{\gamma}/\overline{\beta} < (h - v)/c$.■

[49] Note that to work with interior solutions v cannot be any arbitrary value.

TABLE 2. Summary statistics for all affected sources: 1993–1999.

Variable	1993	1995	1996	1997	1998	1999
No. of sources						
Existing	635	578	504	430	365	365
New	45	112	127	146	221	208
Total Affected	680	690	631	576	566	573
Flow rate (m^3/h)						
Average	4,910.7	4,784.1	4,612.6	4,062.1	4,213.9	4,146.6
Standard dev.	15,058.8	14,908.0	15,490.9	9,498.6	13,091.0	11,793.5
Max.	261,383.9	261,304.7	261,304.7	182,843.0	207,110.6	183,739.5
Min.	499.2	204.3	204.3	493.3	216.9	165.6
Emission rate (mg/m^3)						
Average	94.9	83.1	78.5	54.7	31.1	27.8
Standard dev.	88.1	77.8	76.8	43.0	21.1	18.5
Max.	702.0	698.2	674.0	330.7	110.0	108.2
Min.	1.5	1.5	3.4	3.6	2.9	4.6
Utilization (%)*						
Average	39.4	48.0	47.1	49.2	51.7	53.7
Standard dev.	30.3	31.5	31.7	31.8	32.0	32.3
Max.	100	100	100	100	100	100
Min.	0	0	0	0	0	0
No. of obs.	278	463	457	499	543	542
Total proxied-emissions (kg/day)	7,051.9	6,320.9	5,094.4	3,535.0	1,975.3	1,665.0
Total permits (kg/day)	4,604.1	4,604.1	4,604.1	4,087.5	4,087.5	4,087.5

Source: Elaborated from PROCEFF's databases

* An utilization of 100% corresponds to 24 hrs of operation during 365 days a year. As indicated by the No. of observations, utilization figures are not based on all sources (recall that information on utilization is not required for monitoring and enforcement purposes).

Figure 2: Table 2

TABLE 3. 2SLS estimates for the reduction and utilization equations

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Reduction Equation						
UTIL	0.078	0.137	0.087	0.256*	0.539*	0.308
	(0.153)	(0.175)	(0.175)	(0.132)	(0.309)	(0.322)
FLOW93	-0.789**	-0.788***		-0.937***	-1.090***	
	(0.330)	(0.275)		(0.345)	(0.422)	
$FLOW93^2$	0.270**	0.271**		0.346***	0.373**	
	(0.131)	(0.111)		(0.129)	(0.151)	
ln(FLOW93)			-0.088***			-0.093***
			(0.032)			(0.031)
EMRTE93	0.741***	0.717***	0.698***	0.987***	0.944***	0.940***
	(0.094)	(0.115)	(0.116)	(0.019)	(0.039)	(0.035)
ENDPIPE	-0.058	-0.191	-0.027	-0.128	-0.032	0.182
	(0.198)	(0.251)	(0.140)	(0.083)	(0.100)	(0.129)
INDUST	-0.008	0.079	0.120	0.014	-0.031	0.023
	(0.077)	(0.149)	(0.153)	(0.042)	(0.061)	(0.056)
STATE	-0.137	-0.193**	-0.193**	-0.105**	-0.083	-0.118
	(0.106)	(0.084)	(0.082)	(0.050)	(0.077)	(0.074)
Constant	-0.390***	-0.474***	0.217	-0.420***	-0.512***	0.272
	(0.075)	(0.116)	(0.201)	(0.050)	(0.115)	(0.211)
Utilization Equation						
REDUC	-0.003	0.005	-0.017	0.012	0.063*	0.054
	(0.030)	(0.032)	(0.039)	(0.029)	(0.035)	(0.039)
UTIL93	0.560***	0.567***	0.532***	0.364***	0.313***	0.275***
	(0.055)	(0.064)	(0.061)	(0.064)	(0.093)	(0.089)
FLOW93	0.401***	0.384***		0.416	0.417***	
	(0.130)	(0.096)		(0.267)	(0.129)	
$FLOW93^2$	-0.095**	-0.087***		-0.101	-0.098**	
	(0.048)	(0.033)		(0.101)	(0.045)	
ln(FLOW93)			0.078***			0.090***
			(0.012)			(0.011)
INDUST	0.069*	0.038	0.022	0.141***	0.077	0.044
	(0.037)	(0.045)	(0.049)	(0.044)	(0.057)	(0.057)
STATE	-0.077*	-0.042	-0.038	-0.039	-0.117**	-0.098*
	(0.045)	(0.045)	(0.043)	(0.058)	(0.051)	(0.055)
Constant	0.158***	0.182***	-0.407***	0.221***	0.293***	-0.376***
	(0.034)	(0.045)	(0.100)	(0.042)	(0.057)	(0.090)
No. Obs.	344	344	344	288	288	288

Notes: First-stage results are omitted. White corrected standard errors are in parenthesis. Columns (2), (3), (5) and (6) present weighted estimates (the 1997 flow rate is the weight in (2) and (3) and the1998 flow rate in (5) and (6))
* significant at 10%, ** significant at 5%, *** significant at 1%

Figure 3: Table 3