Multipollutant markets

Juan-Pablo Montero

I study the optimal design of marketable permit systems to regulate various pollutants (e.g., air pollution in urban areas) when the regulator lives in a real world of imperfect information and incomplete enforcement. I show that the regulator should have pollution markets integrated through optimal exchange rates when the marginal-abatement cost curves in the different markets are steeper than the marginal-benefit curves; otherwise he should keep markets separated. I also find that incomplete enforcement reduces the advantage of market integration.

1. Introduction

In recent years, environmental policy makers are paying more attention to environmental markets as an alternative to the traditional command-and-control 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). The U.S. Environmental Protection Agency’s (EPA) emissions-trading policy represents another and older attempt to implement environmental markets to mitigate air pollution problems in urban areas across the country (Hahn, 1989; Foster and Hahn, 1995). In addition, a few less-developed countries are also beginning to experiment in different forms with emissions trading (World Bank, 1997; Monerot, Sánchez, and Katz, forthcoming; Stavins, forthcoming).

The above experiences show that regulators always favor simple regulatory designs that can be implemented in practice over more optimal ones that generally involve nonlinear instruments and transfers to (or from) firms. Within this context of good policy design, however, it is surprising how little attention regulators and policy analysts have paid to multipollutant markets and the possibility of interpollutant trading in those cases where more than one pollutant is being controlled. Once markets have been set up, interpollutant trading requires defining some exchange rate through which emissions permits from the different markets can be traded.

In the U.S. Acid Rain Program, which controls not only SO$_2$ but also nitrogen oxides (NO$_x$), there was some discussion about the possibility of trading SO$_2$ for NO$_x$ emissions and vice versa that never prospered. The EPA’s emissions-trading policy implemented in Los Angeles,
which controls five air pollutants, does include a provision that in principle allows interpollutant trading; in practice, however, it has never been used. I do not have a complete explanation for this regulator’s resistance to having pollution markets be more integrated. Some regulators argue that because the environmental and economic consequences of interpollutant trading are rarely known with certainty (as in this article), the appropriate exchange rate will be too difficult to estimate. Others argue that interpollutant trading could make their current enforcement efforts even less effective by potentially shifting too many emissions from one pollutant to another. We shall see below that neither concern should per se prevent the integration of pollutant markets but rather be part of the policy design.

Because in any urban air pollution control program, as in some other environmental problems, the design and implementation of good environmental policy necessarily involve more than one pollutant (Eskaeland, 1997), the study of marketable permit systems to simultaneously regulate various pollutants becomes very relevant. If the regulator has perfect information about costs and benefits of pollution control, it is not difficult to show that the regulator can implement the first best through the allocation of marketable permits to the different markets without the need for interpollutant trading. In the real world, however, regulators must design and implement policies in the presence of significant uncertainty about costs and benefits (Weitzman, 1974; Lewis, 1996), and usually, under incomplete enforcement as well (Russell, 1990; Malik, 1990). The objective of this article is to study the optimal design of multipollutant markets in such a context.

The optimal design specifies permit allocations to each market and an exchange rate, if any, at which permits from two markets can be traded. Results indicate that the regulator should allow interpollutant trading and have markets fully integrated as long as the marginal-cost curves are steeper than the marginal-benefit curves. This result is analogous to the result obtained by Weitzman (1974), so a similar rationale applies here. Interpollutant trading provides firms with more compliance flexibility, making the cost of control more certain, but at the same time it makes the amount of control in each market more uncertain. Thus, when marginal-cost curves are steeper than the marginal-benefit curves, the regulator should pay more attention to the cost of control rather than the amount of control and, therefore, have markets integrated.

The presence of incomplete enforcement reduces the advantage of market integration as the result of two opposing effects. On the one hand, incomplete enforcement makes the amount of control relatively less uncertain when markets are integrated than when they are not, which increases the advantage of market integration. On the other hand, incomplete enforcement softens both quantity-based market designs, i.e., separated versus integrated markets, making them resemble nonlinear instruments as in Roberts and Spence (1976). When costs are higher than expected, firms do not buy permits but choose not to comply and face an expected penalty fee. While both designs become more flexible in the presence of incomplete enforcement in the sense that the amount of control adapts to cost shocks, the “separated-markets” design becomes relatively more flexible than the “integrated-markets” instrument, because the latter already gave firms flexibility to diversify costs across markets. This flexibility effect dominates the first effect, reducing the advantage of market integration.

---

4 The EPA’s emissions-trading policy covers all significant stationary sources of pollution for five principal air pollutants: carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), particulate matter (PM), and SO2.

5 See South Coast Air Quality Management District’s Program Summary and Rules of 1993, Rule 1309(g), Interpollutant Offsets.

6 Communication with Enrique Calfucura at Chile’s National Comission for the Environment (October 2000).

7 Even if enforcement levels in the markets to be integrated are different, the use of an appropriate exchange rate that controls for enforcement differences could make interpollutant trading a good policy option.

8 Global warming is another good example because current policy proposals include the control of various pollutants besides carbon. Yet another interesting example is George W. Bush’s proposal to simultaneously implement cap-and-trade programs on electric utilities for NOx, SO2, and mercury (Air Daily, October 3, 2000, p. 1.)

9 In fact, this article was motivated by the current interest of Chile’s National Commission for the Environment in exploring quantity-based market instruments for simultaneously controlling various air pollutants in Santiago (mainly PM10 and NOx).
The rest of the article is organized as follows. In Section 2 I present the model and explain firms’ compliance behavior for both market designs: separated and integrated markets. In Section 3 I introduce uncertainty and derive optimal market designs. In Section 4 I compare both designs and discuss the conditions under which one design provides higher expected welfare than the other. Concluding remarks and policy implications are offered in Section 5.

2. The model

Following Montero (forthcoming), I develop a simple multiperiod model with an infinite horizon that captures the basics of multipollutant markets under uncertainty and incomplete enforcement. Consider two flow pollutants, 1 and 2 (e.g., PM10 and NOx), that are to be regulated by implementing two pollution markets. Let $x_i$ be the number of permits that the regulator distributes or auctions off in market $i$ ($i = 1, 2$) in each period and $\alpha$ be the exchange rate, if any, at which permits from market 1 and 2 can be traded (a firm in market 1 can cover 1 unit of pollutant 1 with $1/\alpha$ permits from market 2).

In each market there is a continuum of firms of mass 1 such that in the absence of regulation, each of these firms emits one unit of pollution per period. Pollution in market $i$ can be abated at a cost $c_i$ per period. The value of $c_i$ differs across firms according to the (continuous) density function $g_i(c_i)$ and cumulative density function $G_i(c_i)$ defined over the interval $[c_i, \bar{c}_i]$. These functions are commonly known by both firms and the welfare-maximizing regulator. Although the regulator does not know the control cost of any particular firm, he can derive the aggregate abatement cost curve in market $i$, $C_i(q_i)$, where $0 \leq q_i \leq 1$ is the aggregate quantity of emissions reduction in any given period. The regulator also knows that the benefit curve from emissions reduction in market $i$ in any given period is $B_i(q_i)$. As usual, I assume that $B'(q) > 0$, $B''(q) \leq 0$, $C'(q) > 0$, $C''(0) \geq 0$, $B'(0) > C'(0)$, and $B'(q) < C'(q)$ for $q$ sufficiently large.

The regulator is also responsible for ensuring individual firms’ compliance whether markets are integrated or separated. As in Kaplow and Shavell (1994) and Livernois and McKenna (1999), firms are required to monitor their own emissions and submit a compliance status report to the regulator. Emissions are not observed by the regulator except during costly inspection visits, when they can be measured accurately. Thus, some firms may have an incentive to report themselves as being in compliance when, in reality, they are not. The compliance report also includes details of permit transfers, which are assumed to be tracked at no cost by the regulator. For example, if firm A submits a report with one unit of pollution and a “false” permit transfer from firm B, this can easily be identified, since B would not report a transfer for which it does not get paid. To corroborate the truthfulness of reports received, however, the regulator must observe emissions, which is a costly process.

The regulator lacks sufficient resources to induce full compliance, so to verify reports’ truthfulness he randomly inspects those firms reporting compliance through pollution reduction to monitor their emissions (or check their abatement equipment). Each firm in market $i$ that is reporting compliance faces a probability $\phi_i$ of being inspected. Firms found to be in disagreement with their reports are levied a fine $f_i$ and brought under compliance in the next period. To come under compliance, firms can either reduce pollution or buy permits. Firms reporting non-compliance face the same treatment, so it is always in a firm’s best economic interests to report

---

10 Note that because I am assuming that all parameters in the model (some of which are imperfectly known by the regulator) remain constant over time, the optimal market design will not require banking and borrowing of permits.
11 The aggregate cost curve is $C(q) = \int_0^q c dG$, where $y = G^{-1}(q)$. Note that $C'(q) = y$, $C'(0) = \bar{c}_1$ and $C''(q) = 1/g(y)$.
12 Note that in the absence of interpollutant trading and full enforcement, $q_i = 1 - x_i$.
13 Alternatively, we can simply say that the cost of inspection is large enough that full compliance is not socially optimal (Becker, 1968).
14 A regulator’s power to bring a noncompliant firm under compliance is discussed by Livernois and McKenna (1999). To make sure that a noncompliant firm found submitting a false report is in compliance during the next period (but not necessarily the period after), we can assume that the regulator always inspects the firm during that next period and will, if he finds the firm to be out of compliance again, raise the penalty to something much more severe.
compliance, even if that is not the case. 15 Finally, I assume that the regulator does not alter his policy of random inspections in response to information acquired about firms’ type, so each firm submitting a compliance report declaring a reduction of emissions faces a constant probability \( \phi_i \) of being inspected.

Before describing firms’ compliance behavior under incomplete enforcement, it is worth indicating here that to keep the model simple, I will later introduce the following assumptions: \( B^\prime_i = B^\prime_i, C^\prime_i = C^\prime_i \), and \( \phi_1 = \phi_2 \). Without much loss of generality, this symmetry will prevent us from relying on numerical solutions.

□ Compliance when markets are separated. When markets are designed to work separately, the regulator specifies independently the number of permits to be auctioned off (or freely distributed) in each market, i.e., \( x^s_1 \) and \( x^s_2 \) (superscript \( s \) refers to separated markets). Before we study optimal designs, let \( p^*_i \) be the auction-clearing (or equilibrium) price in market \( i \).

Each firm seeks the compliance strategy that minimizes its expected discounted cost of compliance. Depending on the value of \( \phi_i, F_i, p^*_i \) (assume for the moment that \( \phi_i F_i < p^*_i \)), its marginal-abatement cost \( c_i \), a firm will follow one of two possible strategies: (i) compliance and submission of a truthful report \( (SCT) \), and (ii) noncompliance and submission of a false report declaring compliance \( (SNF) \). Compliance can be achieved by either reducing pollution or buying a permit. Because the horizon is infinite, a firm following a particular strategy at date \( t \) will find it optimal to follow the same strategy at date \( t + 1 \). The date subscript is therefore omitted in the calculations that follow. The subscript \( i \) is also omitted.

Consider first the case in which a firm has relatively low control costs, that is, \( c < p^* \). Such a low-cost firm will never consider buying permits as part of its compliance strategy. If \( SCT \) is its optimal strategy, it will comply by reducing pollution. Conversely, if \( SNF \) is its optimal strategy, should it be found submitting a false compliance report, it will return to compliance by reducing pollution instead of buying permits.

The expected discounted cost of adopting strategy \( SCT \) (compliance and truth-telling) for a low-cost firm is given by

\[
Z^\ell_{CT} = c + \delta Z^\ell_{CT},
\]

where \( \delta \) is the discount rate and superscript \( \ell \) signifies a low-cost firm. In this period, the firm incurs a cost \( c \) from pollution reduction, and during the next period, the firm incurs the present value of following \( SCT \) again. Solving (1) gives

\[
Z^\ell_{CT} = \frac{c}{1 - \delta}.
\]

The expected discounted cost of adopting strategy \( SNF \) (noncompliance and false reporting) for the same low-cost firm (i.e., \( c < p^* \)) is given by

\[
Z^\ell_{NF} = 0 + \phi(F + \delta c + \delta^2 Z^\ell_{NF}) + (1 - \phi)(\delta Z^\ell_{NF}).
\]

In this period, the firm incurs no abatement costs. If the firm is found to have submitted a false report, which happens with probability \( \phi \), the firm must immediately pay the fine \( F \) and return to compliance during the next period by reducing pollution at cost \( c \) (which is cheaper than buying permits at price \( p^* \)). After that, the firm follows \( SNF \) again, with an expected cost of \( Z^\ell_{NF} \). If the firm is not inspected, which happens with probability \( 1 - \phi \), the firm incurs no cost during this period, and in the next period it follows \( SNF \) again, with an expected cost of \( Z^\ell_{NF} \). Solving (3) gives

\[
Z^\ell_{NF} = \frac{\phi(F + \delta c)}{(1 - \delta)(1 + \phi \delta)}.
\]

15 Noncompliance and truth telling could be a feasible strategy if firms reporting noncompliance were subject to a fine lower than \( F \). See Kaplow and Shavell (1994) and Livernois and McKenna (1999) for details.

A low-cost firm is indifferent between following $S_{CT}$ or $S_{NF}$ if $Z^l_{CT} = Z^l_{NF}$. Letting $\bar{c}$ be the marginal cost that makes $Z^l_{CT} = Z^l_{NF}$, we have that

$$\bar{c} = \phi F$$

is the “cutoff” point for a truthful compliance report when $c < p^s$. Thus, if $\bar{c} \leq c \leq \bar{c}$, the firm follows $S_{CT}$, whereas if $\bar{c} < c < p^s$, the firm follows $S_{NF}$.

Consider now the case of a high-cost firm, that is, a firm for which $c \geq p^s$. Such a firm will never consider reducing pollution as part of its compliance strategy. If $S_{CT}$ is its optimal strategy, it will comply by buying permits. Conversely, if $S_{NF}$ is its optimal strategy and it is found submitting a false compliance report, it will return to compliance by buying permits instead of reducing pollution. As before, the expected discounted cost of adopting strategy $S_{CT}$ (compliance and truth telling) for a high-cost firm is given by

$$Z^h_{CT} = p^s + \delta Z^h_{CT}.$$  (6)

In this period, the firm incurs a cost $p^s$, and during the next period the firm incurs the present value of following $S_{CT}$ again. Solving (6) gives

$$Z^h_{CT} = \frac{p^s}{1 - \delta}. $$  (7)

The expected discounted cost of adopting strategy $S_{NF}$ (noncompliance and false reporting) for a high-cost firm is given by

$$Z^h_{NF} = 0 + \phi (F + \delta p^s + \delta^2 Z^h_{NF}) + (1 - \phi)(\delta Z^h_{NF}).$$  (8)

In this period, the firm incurs no abatement costs. If the firm is found to have submitted a false report, which happens with probability $\phi$, the firm must immediately pay the fine $F$ and return to compliance next period by buying permits (which is cheaper than reducing pollution). After that, the firm follows $S_{NF}$ again, with an expected cost of $Z^h_{NF}$. If the firm is not inspected, which happens with probability $1 - \phi$, the firm does not incur any cost in this period, and in the next period it follows $S_{NF}$ again with an expected cost of $Z^h_{NF}$. Solving (8) gives

$$Z^h_{NF} = \frac{\phi (F + \delta p^s)}{(1 - \delta)(1 + \phi \delta)}. $$  (9)

Because $\phi F < p^s$ by assumption, it is not difficult to show that $Z^h_{NF} < Z^h_{CT}$, so a high-cost firm will always follow $S_{NF}$.

Firms’ compliance behaviors can be grouped according to their abatement costs as follows: compliant firms have very low abatement costs (i.e., $\bar{c} \leq c \leq \bar{c}$) and always comply by reducing emissions; noncompliant firms have medium and high costs (i.e., $\bar{c} < c \leq \bar{c}$). A noncompliant firm that is inspected returns to compliance by either reducing pollution if its abatement cost is in the medium range (i.e., $\bar{c} < c \leq p^s$) or buying permits if its abatement cost is high (i.e., $p^s < c \leq \bar{c}$). Note that the above compliance characterization breaks down if $\phi F \geq p^s$. In such a case, there will be full compliance: low-cost firms (i.e., $\bar{c} \leq c \leq p^s$) will reduce pollution all the time, and high-cost firms (i.e., $p^s < c \leq \bar{c}$) will always buy permits. Although $\phi F \geq p$ is possible for low inspection costs and high fines, in this article I am interested in the case of partial compliance, or incomplete enforcement. Note also that if $\phi = 1$ and $F < p^s$, it is still possible to have a fraction of noncompliant firms.

Because of partial compliance, the effective amount of pollution reduction in any given period is expected to be

$$q^e_{x^s} = G(\bar{c}) + \gamma [G(p^s) - G(\bar{c})].$$  (10)
where the first term of the right-hand side represents reductions from low-cost compliant firms and the second term represents reductions from a fraction $\gamma = \phi/(1 + \phi)$ of formerly noncompliant firms that came into compliance this period by reducing one unit of pollution (subscript “c” refers to effective amount).\(^{16}\)

Similarly, the effective control costs incurred by firms are expected to be

$$C_e(x^t) = \int_{c}^{\tilde{c}} c dG + \gamma \int_{c}^{\phi} c dG. \quad (11)$$

Note that as $\phi$ and/or $F$ increases, $\tilde{c}$ approaches $p^*$ and $C_e(\cdot)$ approaches $C(\cdot)$.

Because the regulatory design does not specify $p^*$ but rather the number of permits to be supplied, $x^t$, it remains to find $p^*$ as a function of $x^t$, that is, $p^*(x^t)$. Assuming rational expectations, the market-clearing condition is\(^{17}\)

$$x^t = \gamma[1 - G(p^*)]. \quad (12)$$

The left-hand side of (12) is the total number of permits supplied by the regulator, while the right-hand side is purchases from high-cost firms (i.e., $c > p^*$) following $SNF$ that in this period come under compliance by buying permits instead of reducing pollution.

Solving (12) gives

$$p^*(x^t) = G^{-1}(1 - \frac{x^t}{\gamma}), \quad (13)$$

where $G^{-1}(1 - x^t/\gamma)$ can be viewed as the marginal cost $c$ just after $1 - x^t/\gamma$ units of pollution have been reduced.\(^{18}\) Since the equilibrium price of permits under full compliance would be $G^{-1}(1 - x^t)$, which occurs when firms are in compliance all the time (in this model, when $\gamma = 1$), it is immediate that incomplete enforcement lowers $p^*$. The reason for this result is simply that noncompliance and permits are (imperfect) substitutes, which depresses their demand and price. Note also that if $\gamma = 1$, $G(p^*) = 1 - x^t$ and $q_e(x) = 1 - x^t$.

\[]

**Compliance when markets are integrated.** When markets are designed to work together, the regulator specifies simultaneously the number of permits to be auctioned off in each market, $x^1_t$ and $x^2_t$ respectively, and the exchange rate $\alpha$ at which permits from market 1 can be traded for permits from market 2 (superscript $t$ refers to integrated markets). Note that because permits are fully tradeable across markets, it is irrelevant how the regulator allocated the total number of permits, $x^1_t + \alpha x^2_t$, between the two markets. In other words, the regulator just needs to specify $x^1_{12} = x^1_t + \alpha x^2_t$ and $\alpha$.

Firms’ compliance when markets work together is not much different from the previous analysis except for the market-clearing conditions. If $p^1_t$ and $p^2_t$ are the clearing prices in markets 1 and 2, respectively, the “integrated-markets”–clearing condition is

$$x^1_{12} = \gamma_1[1 - G_1(p^1_t)] + \alpha \gamma_2[1 - G_2(p^2_t)]. \quad (14)$$

The left-hand side of (14) is the total number of permits from both markets expressed as permits of market 1. The first term of the right-hand side is purchases from high-cost firms in market 1

\(^{16}\)To determine $\gamma$, let $K_t + N_t = 1$, where $K_t$ is the number of noncompliant firms (i.e., firms that follow $SNF$) that are in compliance at time $t$ and $N_t$ is the number of noncompliant firms that are out of compliance at $t$. Provided that $K_t = \phi N_{t-1}$ and $N_{t-1} = 1 - K_{t-1}$, at steady state we have that $K_t = K_{t-1} = K = \phi/(1 + \phi)$.

\(^{17}\)The same market-clearing condition results under grandfathered permits. If each firm receives $x^t$ permits, then $x^t G(\tilde{c}) + x^t \gamma(G(p^t) - G(\tilde{c})) + x^t (1 - \gamma) (1 - G(\tilde{c})) = (1 - x^t) \gamma (1 - G(p^t))$. Sellers are on the left-hand side: compliant firms, a fraction $\gamma$ of noncompliant firms in compliance today, and a fraction $(1 - \gamma)$ of noncompliant firms. Buyers are on the right-hand side: noncompliant firms in compliance today.

\(^{18}\)Note that for a uniform distribution of $g(c) = 1/C'' = 1/\bar{c} - \tilde{c}$, we have that $p^t = \bar{c} - C'' x^t / \gamma$. 

In the real world, regulators must choose policy goals and instruments in the presence of significant uncertainty about both \( B(q) \) and \( C(q) \). It is true, however, that while both the regulator and firms are uncertain about the true shape of the benefit curve, firms generally know or have a better sense than the regulator of the true value of their costs.

So far I have not assumed any particular shape for the benefit and cost curves. To keep the model tractable after the introduction of uncertainty, however, I follow Weitzman (1974) and Baumol and Oates (1988) in considering linear approximations for the marginal-benefit and cost curves and additive uncertainty. Then, let the expected benefit and cost curves in market \( i \) in any given period be, respectively,

\[
B_i(q_i) = b_i q_i + \frac{B''_i}{2} q_i^2 \\
C_i(q_i) = c_i q_i + \frac{C''_i}{2} q_i^2,
\]

where \( b_i \equiv B'_i(0) > 0, \) \( B''_i < 0, \) and \( C''_i \equiv \bar{C}_i - \bar{C}_j > 0 \) are all fixed coefficients.\(^{19}\)

Next, to capture the regulator’s aggregate uncertainty about the true shape of these curves at the time of market design and implementation, let his prior belief for the marginal-benefit curve be \( B_i^i(q_i, \eta_i) = B_i(q_i) + \xi_i, \) where \( \eta_i \) is a stochastic term such that \( E[\eta_i] = 0 \) and \( E[\eta_i^2] > 0. \) In addition, for the marginal-cost curve, let his prior belief be \( C_i^i(\theta_i) = c_i + \theta_i, \) where \( \theta_i \) is another stochastic term such that \( E[\theta_i] = 0 \) and \( E[\theta_i^2] > 0. \) I assume that \( \theta_i \) is common to all individual costs \( c_i \in \{ c, \bar{c} \} \), which introduces aggregate uncertainty in the marginal-cost curve, \( C_i^i(q_i), \) by the amount \( \theta_i, \) that is, \( C_i^i(q_i, \theta_i) = C_i^i(q_i) + \theta_i. \) Recall that the realization of \( \theta_i \) is observed by all firms in market \( i \) before they make and implement their compliance (and production) plans. Because \( \theta_i \) (and also \( \eta_i \)) does not vary over time, one can argue that the regulator may (imperfectly) deduce its value with a lag from the aggregate behavior of firms. However, I assume that he adheres to the original regulatory design from the beginning of time.\(^{20}\)

\[ \square \]

Designing separated markets. The regulator needs to specify the number of permits \( x_i \) and \( x'_i \) to be allocated in each market, respectively, in any given period. He considers each market separately and solves (I omit subscript \( i \))

\[
\max_{x} W(x) = E \left[ B(q(x^*, \theta), \eta) - C_i(x^*, \theta) \right].
\]  

\(^{19}\) Note first that the linear marginal-cost curve results simply from a uniform distribution for \( g_i(c_i). \) Further, the notation \( b_i \) is meant to be consistent with \( \bar{c}_j \) in the cost curve.

\(^{20}\) Alternatively, we can say that new sources of uncertainty arise continually, so the issue of uncertainty is never resolved. For example, we can let \( \theta \) and \( \eta \) follow random walks. The computation of compliance strategies would be the same, but the variance of \( \theta \) and \( \eta \) would grow linearly with time.
where \( q^e_\alpha(x, \theta) \) and \( C^e_\alpha(x, \theta) \) can be derived from (10) and (11) as \((g(c) = 1/C'' = 1/(\bar{c} - c))\)

\[
q^e_\alpha(x^i, \theta) = \frac{1}{C''} c + \gamma \int_{c_{\min}}^{c_{\max}} \frac{1}{C''} dc + \gamma \int_{c_{\min}}^{c_{\max}} \frac{1}{C''} dc,
\]

\[
C^e_\alpha(x^i, \theta) = \frac{1}{C''} c + \gamma \int_{c_{\min}}^{c_{\max}} \frac{1}{C''} dc + \gamma \int_{c_{\min}}^{c_{\max}} \frac{1}{C''} dc,
\]

and

\[
p^\prime = p^\prime(x^i, \theta) = p^\prime(x^i) + \theta = \frac{\gamma \bar{c} - x^i C''}{\gamma} + \theta.
\]

Substituting (17)–(19) into (16), the first-order condition for \( x^i \) reduces to

\[
E \left[ b + \eta + B'' q^e_\alpha(x^i, \theta) - p^\prime(x^i, \theta) \right] = 0,
\]

where

\[
q^e_\alpha(x^i, \theta) = q^e_\alpha(x^i) - \frac{1 - \gamma}{C''} \theta = 1 - x^i - \frac{(1 - \gamma)(\bar{c} - F)}{C''} - \frac{1 - \gamma}{C''} \theta.
\]

The first-order condition (20) indicates that at the optimum the expected marginal benefit is equal to the expected equilibrium price of permits (i.e., expected marginal cost). Furthermore, because of the linear approximations and additive uncertainty, the optimal amount of permits \( x^i \) is independent of \( \eta \) and \( \theta \). Equation (21), on the other hand, shows that under incomplete enforcement \((\gamma < 1)\) the effective amount of reduction, \( q^e_\alpha(x^i, \theta) \), does depend on the value of \( \theta \). If costs are higher than expected \((\theta > 0)\), firms will reduce their level of compliance, and consequently, the effective amount of reduction. Under complete enforcement \((\gamma = 1)\), however, the amount of (effective) reduction is fixed and equal to \(1 - x^i\), which is simply baseline emissions minus the number of auctioned permits.

\( \Box \) Designing integrated markets. If both markets are designed to work together, the regulator needs to specify the total number of permits \( x^e_{12} = x^e_1 + \alpha x^e_2 \) to be allocated and the exchange rate \( \alpha \) at which permits can be traded. Then, he considers both markets simultaneously and solves

\[
\max_{x^{e}_{12, \alpha}} W^e = E \left[ \sum \left( B_i(q^e_{\alpha}(x^e_{12}, \alpha, \theta_i, \theta_{-i}), \eta_i) - C^e_{\alpha}(x^e_{12}, \alpha, \theta_i, \theta_{-i}) \right) \right],
\]

where \( q^e_{\alpha} \) and \( C^e_{\alpha} \) can be obtained, respectively, from (17) and (18) by simply replacing \( p^\prime \) with the corresponding price \( p^i \). From (14) and (15) and assuming \( C^e_1 = C^e_2 = C'' \), these prices are

\[
p^i_1(x^e_{12}, \alpha, \theta_1, \theta_2) = \frac{\gamma (\bar{c}_1 + \alpha \bar{c}_2) - x^e_{12} C''}{\gamma (1 + \alpha^2)} + \frac{\theta_1 + \alpha \theta_2}{1 + \alpha^2}
\]

\[
p^i_2(x^e_{12}, \alpha, \theta_1, \theta_2) = \alpha p^i_1(x^e_{12}, \alpha, \theta_1, \theta_2).
\]

Even in this already simple model of two pollutants, the solution of (22) and subsequent comparison with the solution of (16) requires numerical solutions unless some further simplifications are made.

**Proposition 1.** If \( B_1'' = B_2'' = B'' \), \( C_1'' = C_2'' = C'' \), \( \phi_1 = \phi_2 \), and \( \theta_1 \) and \( \theta_2 \) are i.i.d. and not correlated with \( \eta \), the optimal design when markets are integrated is

\[
x^e_{12} = x^e_1 + \alpha x^e_2,
\]
and
\[
\alpha = \frac{E[B'_2(q^*_e(x^*_e, \theta_2), \eta_2)]}{E[B'_1(q^*_e(x^*_e, \theta_1), \eta_1)]} = \frac{E[p^*_2(x^*_e, \theta_2)]}{E[p^*_1(x^*_e, \theta_1)]}.
\]

**Proof.** See the Appendix.

Because of the symmetry of the problem, the results under Proposition 1 are very intuitive.\(^{21}\)

The first result indicates that the total number of permits is the same under either market design.

The second result indicates that the exchange rate at which permits from market 1 and 2 can be traded is exactly equal to the ratio of expected marginal damages in the optimal separated-markets design, which must also be equal to the ratio of expected prices. Since I do not impose restrictions on the values of \(b_i\) and \(c_i\), the value of \(\alpha\) can be equal to, greater than, or lower than one. In fact, if \(b_2 > b_1\) and \(c_2 > c_1\), the optimal value of \(\alpha\) will be greater than one.

Using the results of Proposition 1, we can now easily compare prices and quantities (i.e., amount of effective reduction) under both market designs. Prices are given by
\[
\begin{align*}
 p'_1 &= p^*_1 - \frac{\alpha^2 \theta_1 - \alpha \theta_2}{1 + \alpha^2}, \\
 p'_2 &= p^*_2 + \frac{\alpha \theta_1 - \theta_2}{1 + \alpha^2},
\end{align*}
\]

where \(p^*_i\) is given by (19). While expected prices do not vary with market design (i.e., \(E[p'_i] = E[p^*_i]\)), actual prices generally do so. For example, if \(\theta_1 > 0\) and \(\theta_2 = 0\), \(p'_1 < p^*_1\). The equilibrium price \(p'_1\) does not go up as much because under integrated markets those firms with costs between \(p'_1\) and \(p^*_1\) will find it cheaper to buy permits from market 2 than reducing emissions themselves.

Quantities, on the other hand, are given by
\[
\begin{align*}
 q^*_e &= q^*_e - \frac{(\alpha^2 \theta_1 - \alpha \theta_2)\gamma}{(1 + \alpha^2)\gamma'}, \\
 q^*_e &= q^*_e + \frac{(\alpha \theta_1 - \theta_2)\gamma}{(1 + \alpha^2)\gamma'},
\end{align*}
\]

where \(q^*_e\) is given by (21). Again, while expected quantities do not vary with market design (i.e., \(E[q^*_e] = E[q^*_e]\)), actual quantities generally do so. For example, if \(\theta_1 > 0\) and \(\theta_2 = 0\), \(q^*_e < q^*_e\) for the same reasons laid out above.

Because \(\theta\)'s are i.i.d., by having markets integrated, firms have more flexibility to comply, which ultimately makes the price \(p'_1\) (or marginal cost) in each market less uncertain (this leads to cost savings in expected terms). At the same time, however, the actual reductions \(q^*_e\) and \(q^*_e\) that will take place in each market become more uncertain (this leads to benefit losses in expected terms). In deciding whether to have markets integrated and to allow interpollutant trading, the regulator will inevitably face this tradeoff between cost savings and benefit losses. I study this tradeoff more formally in the next section.

### 4. Integrated versus separated markets

To find the optimal policy design, let us start by writing the difference in expected welfare between the two market designs (integrated and separated markets):
\[
\Delta_{ts} = W'_{12}(x^*_{12}, \alpha) - (W'_{11}(x^*_1) + W'_{22}(x^*_2)),
\]

where \(x^*_{12}, \alpha, x^*_1,\) and \(x^*_2\) are at their optimal values. The normative implication of (29) is that if \(\Delta_{ts} > 0\), the optimal policy design is to have both markets integrated.

---

\(^{21}\) The proposition still holds if \(E[\eta_1 \theta_1] = E[\eta_2 \theta_2]\).
To explore under which conditions this is the case, we conveniently rewrite (29) as

$$
\Delta_{ts} = E \left[ \sum_{i=1}^{n} \left( \{ B_i(q_r^i, \eta_i) - B_i(q_c^i, \eta_i) \} - \{ C_r^i - C_c^i \} \right) \right].
$$

The first curly bracket of the right-hand side of (30) is the difference in environmental benefits provided by the two market designs, whereas the second curly bracket is the difference in abatement costs. Introducing the same simplifying assumptions under Proposition 1, (30) becomes

$$
\Delta_{ts} = E \left[ \sum_{i=1}^{n} \left( \left( \tilde{p}_i + \eta_i \right)(q_r^i - q_c^i) + \frac{B''}{2} \left( (q_r^i)^2 - (q_c^i)^2 \right) \right) - \left\{ \gamma \int p_i^c \frac{c}{C''} dc \right\} \right].
$$

where \( p_i^r, p_i^c, q_r^i, \) and \( q_c^i \) can be obtained from (25)–(28).

Taking the expectation and assuming that \( E[\eta_i, \theta_i] = 0 \) and \( E[\eta_i, \theta_{i-1}] = 0 \), expression (31) reduces to

$$
\Delta_{ts} = \gamma (2 - \gamma) E[\eta^2] B'' + \gamma E[\theta^2] C'',
$$

where the first term of the right-hand side is the difference in expected benefits and the second term is the difference in expected costs. Finally, rearranging (32) leads to

$$
\Delta_{ts} = \frac{\gamma E[\eta^2]}{2(C'')^2} (2 - \gamma) B'' + C'',
$$

where \( \gamma = \phi/(1+\phi) < 1 \) is the fraction of noncompliant firms that are in compliance today, \( E[\eta^2] \) represents regulator’s cost uncertainty, and \( B'' < 0 \) is the slope of the marginal-benefit curves, and \( C'' > 0 \) is the slope of the marginal-cost curves. We can summarize the above result in the following proposition:

**Proposition 2.** If \( B'' = B''_1 = B''_2 = B''_3 \), \( C'' = C''_1 = C''_2 = C''_3 \), and \( \gamma \) are i.i.d. and not correlated with \( \eta \), the optimal policy design under full enforcement \((\gamma = 1)\) is to have both markets integrated as long as \( C'' > |B''| \). Under incomplete enforcement, however, the optimal design is to have markets integrated only if \( C'' > (2 - \gamma) |B''| \).

The first result stated in Proposition 2 is that under full enforcement \((\gamma = 1)\) the regulator should allow interpollutant trading as long as the marginal-cost curves are steeper than the marginal-benefit curves. This result is analogous to the result obtained by Weitzman (1974) when he compared price (e.g., taxes) and quantity (e.g., tradable permits) instruments. Weitzman found that if the marginal-cost curve was steeper than the marginal-benefit curve, the regulator should pay more attention to the cost of control than the amount of control (i.e., emissions reduction) and should, therefore, use the price instrument.

The same rationale applies to our multipollutant markets story. Interpollutant trading provides more flexibility to firms in case costs are higher than expected, but at the same time, it makes the amount of control in each market more uncertain (see (32)). Then, if the marginal-cost curves are steeper than the marginal-benefit curves, the regulator should pay more attention to cost than to the amount of control and, therefore, have markets integrated. On the other hand, if the marginal-benefit curves are steeper than the marginal-cost curves, the regulator should pay more attention to the amount of control in each market and, therefore, have markets separated.

The second result under Proposition 2 is that if incomplete enforcement \((\gamma < 1)\) has important effects on the multipollutant markets design.\(^{22}\) Since \( 2 - \gamma > 1 \), (33) indicates that incomplete

\(^{22}\) Note that because \( \phi \leq 1 \), \( \gamma \)’s upper bound is 1/2 in this model. But as the enforcement power (regulator’s ability to bring and keep noncompliant firms under compliance) increases, \( \gamma \)’s upper limit approaches one. That would be the case in this model, if we assume, for example, that the regulator is able to keep the noncompliant firm under compliance for more than one period.
enforcement reduces the advantage of market integration: the regulator should allow interpollutant trading only if the marginal-cost curves are $2 - \gamma$ times steeper than the marginal-benefit curves.

There are two effects that lead to this result. The first term of the right-hand side of (32) captures the first effect: the gains in expected benefits from market separation are reduced under incomplete enforcement ($\gamma (2 - \gamma) < 1$) because $q_{i2}^{*}$ becomes relatively less uncertain than $q_{i1}^{*}$. The second term captures the second effect: the gains in expected cost savings from market integration are reduced under incomplete enforcement ($\gamma < 1$ by definition) because both market designs adapt to some extent to unexpected costs. Because $\gamma (2 - \gamma) > \gamma$, the second effect dominates, so the overall advantage of market integration is reduced. In addition, note that as enforcement weakens ($\gamma$ falls), the welfare difference between the two market designs shrinks and disappears when there is no compliance at all (i.e., $\gamma = 0$).

Finally, we can relax some of the assumptions on correlations between the different stochastic terms. If $E[\eta_1 \theta_1] > 0$, a negative term enters into (32), increasing the advantage of separated markets; otherwise, benefit uncertainty does not intervene. In addition, if $E[\theta_1 \theta_{-1}] > 0$, a term of opposite sign enters into (32), reducing the welfare difference between the two market designs.26

5. Conclusions and policy implications

Because in many environmental problems the design and implementation of good policy necessarily involves more than one pollutant, I have developed a simple model to study the optimal design of environmental markets (i.e., tradeable emissions permits) to simultaneously regulate various pollutants when the regulator lives in a real world of imperfect information and incomplete enforcement. I found that if the marginal-abatement cost curves are relatively steeper than the marginal-benefit curves, which seems to be the case for some urban air pollutants (Cifuentes, 2001), the regulator should have multipollutant markets integrated through optimal exchange rates, unless enforcement is too weak, in which case the regulator should keep markets separated.

The results of the article are also relevant for a regulator that is implementing a multipollutant offset system where new firms must compensate for all their emissions by buying emission-reduction credits from existing firms. For example, this regulator should address the question of whether a firm can compensate for all its PM10 and NOx emissions with PM10 credits. This question also involves specifying the appropriate exchange rate between PM10 and NOx credits. Note that because the level of control in each market is not necessarily at its (ex ante) optimum, the exchange rate may depart from the recommendations in Section 3.

I hope that the results of this article provide the basis for empirical and applied work on the design of multipollutant markets in more real settings. In such cases the model should consider, among other things, atmospheric interaction among pollutants, spatial and temporal characteristics of the pollutants, joint production of pollutants, monitoring heterogeneity, and, possibly, institutional constraints. If we also let cost and benefit curves vary over time (e.g., $\theta$ and $\eta$ following random walks), the model should also allow for the possibility of banking and borrowing of permits. I leave an application of this model to the case of air pollution in Santiago, Chile for further research.

Appendix

Proof of Proposition 1. I proceed in two steps. Since I have already shown in the text that first-order conditions for $x_{11}^{*}$ and $x_{22}^{*}$ are independent of the stochastic variables $\eta$ and $\theta$ (see (20)), I first demonstrate that Proposition 1 holds under certainty, and then I demonstrate that the first-order conditions for $x_{12}^{*}$ and $a$ are independent of $\eta$ and $\theta$.

23 A discussion of whether this correlation is more likely to be positive or negative is found in Stavins (1996).

24 For $E[\eta_1 \theta_1] = E[\theta_2 \theta_2]$, the extra term is $-\gamma E[\eta_1 \theta_1]/C''$.

25 Whether this correlation is positive or negative is an empirical question. It is very likely to be positive if we are considering PM2.5 and PM10, but it is not so clear if we are considering NOx and SO2.

26 The extra term is $-a\gamma(2 - \gamma)B'' + C''E[\theta_1 \theta_1]/((1 + a^2)(C''))$. Note that because $a/(1 + a^2) \leq 1/2$ for all $a > 0$ and $E[\theta_1 \theta_1] < E[\theta_1 \theta_1] = E[\theta_2 \theta_2]$, this extra term can never revert the policy choice prescribed by (33).
Under certainty, the first-order conditions for \( x'_t \) and \( \alpha \) are, respectively,

\[
\begin{align*}
(\beta_1 + B''q'_1 - p'_1) \frac{\partial p'_1}{\partial x'_{12}} + (\beta_2 + B''q'_2 - p'_2) \frac{\partial p'_2}{\partial x'_{12}} &= 0 \\
(\beta_1 + B''q'_1 - p'_1) \frac{\partial p'_1}{\partial \alpha} + (\beta_2 + B''q'_2 - p'_2) \frac{\partial p'_2}{\partial \alpha} &= 0.
\end{align*}
\]

(A1)

(A2)

Since both \( \partial p'_1 / \partial x'_{12} \) and \( \partial p'_2 / \partial \alpha \) are different from zero, the solution of the above system of equations satisfies

\[
\beta_1 + B''q'_1 - p'_1 = 0.
\]

(A3)

But (A3) is the first-order condition for \( x'_t \) (see (20)), which implies that under certainty we have

\[
p'_1 = p'_1.
\]

(A4)

Using (A4) and \( p'_2 = \alpha p'_1 \) (by an arbitrage condition of integrated markets), it follows that \( \alpha = p'_2 / p'_1 \), where \( p'_2 \) is equal to the value of \( E[p'_2] \) obtained from (19). Thus, we have that under certainty,

\[
p'_1 = \frac{\gamma x'_1 - x'_1 C''}{\gamma},
\]

(A5)

therefore

\[
p'_1 + \alpha p'_2 = (1 + \alpha^2)p'_1 = \frac{\gamma (\bar{c}_1 + \alpha \bar{c}_2) - (x'_1 + \alpha x'_2) C''}{\gamma}.
\]

(A6)

Comparing (A6) with the deterministic part of (23), it follows that \( x'_1 = x'_1 + \alpha x'_2 \) under certainty.

Proceeding with the second step of the proof, the first-order conditions for \( x'_t \) and \( \alpha \) under uncertainty are, respectively,

\[
E \left[ \sum_{i=1,2} \left( \beta_i + \eta_i + B''q'_i(\theta, \theta_{-i}) - p'_i(\theta, \theta_{-i}) \right) \frac{\partial p'_i(\theta, \theta_{-i})}{\partial x'_{i2}} \right] = 0
\]

(A7)

\[
E \left[ \sum_{i=1,2} \left( \beta_i + \eta_i + B''q'_i(\theta, \theta_{-i}) - p'_i(\theta, \theta_{-i}) \right) \frac{\partial p'_i(\theta, \theta_{-i})}{\partial \alpha} \right] = 0,
\]

(A8)

where

\[
q'_i(\theta_1, \theta_2) = A_1 + \frac{(\gamma - 1 - \alpha^2)\theta_1 + \alpha \gamma \theta_2}{(1 + \alpha^2)C''}
\]

(A9)

\[
q'_2(\theta_1, \theta_2) = A_2 + \frac{\alpha \gamma \theta_1 + (\alpha^2 \gamma - 1 - \alpha^2) \theta_2}{(1 + \alpha^2)C''}
\]

(A10)

\[
p'_i(\theta_1, \theta_2) = D_1 + \frac{\theta_1 + \alpha \theta_2}{1 + \alpha^2}
\]

(A11)

\[
p'_2(\theta_1, \theta_2) = D_2 + \frac{\alpha \theta_1 + \alpha^2 \theta_2}{1 + \alpha^2}
\]

(A12)

\[
\frac{\partial p'_1(\theta_1, \theta_2)}{\partial x'_{12}} = -\frac{C''}{(1 + \alpha^2)\gamma} = \frac{1}{\alpha} \frac{\partial p'_1(\theta_1, \theta_2)}{\partial x'_{12}}
\]

(A13)

\[
\frac{\partial p'_1(\theta_1, \theta_2)}{\partial \alpha} = H_1 + \frac{-2 \alpha^2 \gamma \theta_1 + \gamma^2 (1 - \alpha^2) \gamma \theta_2}{[(1 + \alpha^2)\gamma]^2}
\]

(A14)

\[
\frac{\partial p'_2(\theta_1, \theta_2)}{\partial \alpha} = H_2 + \frac{\gamma^2 (1 - \alpha^2) \theta_1 + 2 \alpha \gamma^2 \theta_2}{[(1 + \alpha^2)\gamma]^2}
\]

(A15)

\( A_1, D_1, \) and \( H_1 \) are deterministic terms.

Since \( \theta_1 \) and \( \theta_2 \) are i.i.d. and not correlated with either \( \eta_1 \) or \( \eta_2 \), we are only interested in \( E[\theta'_1] \) terms. On the one hand, the first-order condition (A7) does not include any \( E[\theta'_1] \) terms because \( \partial p'_1 / \partial x'_{12} \) is independent of \( \theta_i \) and \( \theta_{-i} \), as shown by (A13). The first-order condition (A8), on the other hand, does include several \( E[\theta'_1] \) terms because \( \partial p'_1 / \partial \alpha \) depends on \( \theta_i \) and \( \theta_{-i} \), as indicated by (A14) and (A15). However, the multiplicative interaction of \( \partial p'_1 / \partial \alpha \) with \( \theta'_1 \) and \( \partial p'_2 / \partial \alpha \) with \( p'_1 \) results in a total of eight \( E[\theta'_1] \) and \( E[\theta'_2] \) terms that cancel out when \( E[\theta'_1] = E[\theta'_2] \). This demonstrates...
that (A8) is also independent of \( \eta \) and \( \theta \) and, consequently, finishes the proof of Proposition 1. (Note that if we let \( \theta_i \) be correlated with \( \eta_i \), there will be two additional \( E[\theta_i \eta_i] \) terms in (A8) that cancel out when \( E[\theta_1 \eta_1] = E[\theta_2 \eta_2] \).) Q.E.D.

References


