Multipollutant markets

Juan-Pablo Montero*

Catholic University of Chile and MIT

February 20, 2001

Abstract

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.

*Assistant Professor of Industrial Economics at the Department of Industrial Engineering of the Catholic University of Chile (PUC), and Research Associate at the MIT Center for Energy and Environmental Policy Research (MIT-CEEPR). Send any correspondence to: Catholic University of Chile, Casilla 306, Correo 22, Santiago, Chile. Phone: 56-2-686-4613, fax: 56-2-552-1608; and e-mail: jp-monter@ing.puc.cl. Thanks to Luis Cifuentes, Joe Kruger and participants in a seminar at PUC for comments and information; and the Chile’s National Commision for the Environment (CONAMA) and MIT-CEEPR for financial support. Remaining errors and omissions are my own responsibility.
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; Montero et al., 2000).1

The above experiences show that regulators always favor simple regulatory designs that can be implemented in practice over more optimal ones that generally involve non-linear instruments and transfers to (or from) firms.2 Within this context of good policy design,3 however, it is surprising the little attention that regulators and policy analysts have payed 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 emission permits from the different markets can be traded.

In the U.S. Acid Rain Program, which not only controls SO$_2$ but also NO$_x$, there

---

1Further experiences of environmental markets can be found in Stavins (2001) and EPA (2001).
2For a complete survey of optimal environmental regulation see Lewis (1996).
3I borrow the word “good” from Schmalensee (1989) to mean feasible to implement and more efficient than command-and-control regulation.
was some discussion about the possibility of trading SO$_2$ for NO$_x$ emissions and vice versa that never prospered.$^4$ The EPA’s emissions trading policy implemented in Los Angeles, which controls five air pollutants,$^5$ does include a provision that, in principle, allows interpollutant trading;$^6$ but in practice, it has never being used. There are at least two reasons that seem to explain regulators’ resistance to have pollution markets more integrated. The first reason is the uncertainties regulators face in evaluating the environmental and economic consequences of interpollutant trading and, hence, in finding the right exchange rate for the trade between two pollutants. The second reason is that some regulators fears that interpollutant trading would make their current enforcement efforts less effective (i.e. there will be less compliance than otherwise).$^7$

Because in any urban air pollution control program, like in some other environmental problems,$^8$ the design and implementation of good environmental policy necessarily involves more than one pollutant (Eskeland, 1997), the study of marketable permit systems to simultaneously regulate various pollutants becomes very relevant.$^9$ 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 per-

$^4$See Clean Air Act Amendments of 1990, Public Law 101-594, title IV, § 403(c), Interpollutant trading.

$^5$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 (NO$_x$), particulate matter (PM) and SO$_2$.

$^6$See Rule 1309(g), Interpollutant offsets, in SCAQMD (1993).

$^7$Communication with Enrique Calfucura at Chile’s National Comission for the Environment (October, 2000).

$^8$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 of simultaneously implement cap-and-trade programs on electric utilities for NO$_x$, SO$_2$, mercury and CO$_2$ (Air Daily, Vol. 7, No. 189, October 3, 2000).

$^9$In fact, this paper was motivated by the current interest of Chile’s National Comission for the Environment in exploring quantity-based market instruments for simultaneously controlling various air pollutants in Santiago (mainly PM10, CO and NO$_x$).
mits 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 concerning 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 permits 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 vs. integrated markets, making them to resemble non-linear instruments.\footnote{See Roberts and Spence (1976) on the advantages of two-part instruments.} 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 adapt to cost shocks, the “separated markets” design becomes relatively more flexible than the “integrated markets” instrument, because the latter already provided firms with flexibility to diversify costs across markets. This flexibility effect dominates the first effect reducing the advantage of market integration.

The rest of the paper 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 (2001), I develop a simple multi-period model of infinite horizon that captures the basics of multipollutant markets under uncertainty and incomplete enforcement. Consider two pollutants 1 and 2 (e.g. PM10 and NO$_x$) that are to be regulated by implementing two pollution markets. Let $x_i$ be the number permits that the regulator distribute or auction off in market $i$ ($= 1, 2$) in each period and $\alpha$ 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 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.\footnote{The aggregate cost curve is $C(q) = \int_0^q c dG$, where $y = G^{-1}(q)$. Note that $C'(q) = y$, $C'(0) = c$, and $C''(q) = 1/g(y)$.} 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 $B'(q) > 0$, $B''(q) \leq 0$, $C'(q) > 0$, $C''(q) \geq 0$, $B'(0) > C'(0)$, 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 be easily 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

\footnote{Note that in the absence of interpollutant trading and full enforcement $q_i = 1 - x_i$.}
observe emissions, which is a costly process.

The regulator lacks sufficient resources to induce full compliance,\textsuperscript{13} therefore, in order 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.\textsuperscript{14} To come under compliance, firms can either reduce pollution or buy permits. Firms reporting noncompliance face the same treatment, so it is always in a firm’s best economic interests to report compliance, even if that is not the case.\textsuperscript{15} Finally, I assume that the regulator does not alter its policy of random inspections in response to information acquired about firms’ type, so each firm submitting a compliance report 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: $B_1'' = B_2''$, $C_1'' = C_2''$ and $\phi_1 = \phi_2$. Without much loss of generality, this symmetry will prevent us from relying on numerical solutions.

\textsuperscript{13}Alternatively, we can simply say that the cost of inspection is large enough that full compliance is not socially optimal (Becker, 1968).

\textsuperscript{14}Regulator’s enforcement power to bring a non-compliant firm under compliance is discussed by Livernois and McKenna (1999). To make sure that a non-compliant 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 in the case the firm is found to be out of compliance again, the regulator raises the penalty to something much more severe.

\textsuperscript{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.
2.1 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. \(x^1_s\) and \(x^2_s\) (superscript “s” refers to separated markets). Before studying 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 \((S_{CT})\), and (ii) noncompliance and submission of a false report declaring compliance \((S_{NF})\). Compliance can be achieved by either reducing pollution or paying the tax. 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, \(c < p^a\). Such a low-cost firm will never consider buying permits as part of its compliance strategy. If \(S_{CT}\) is its optimal strategy, it will comply by reducing pollution. Conversely, if \(S_{NF}\) 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 \(S_{CT}\) (compliance and truth-telling)
for a low-cost firm is given by

\[ Z_{CT}^l = c + \delta Z_{CT}^l \]  

(1)

where \( \delta \) is the discount rate and superscript “\( l \)” 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 \( S_{CT} \) again. Solving (1) gives

\[ Z_{CT}^l = \frac{c}{1 - \delta} \]  

(2)

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

\[ Z_{NF}^l = 0 + \phi(F + \delta c + \delta^2 Z_{NF}^l) + (1 - \phi)(\delta Z_{NF}^l) \]  

(3)

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 \( S_{NF} \) again, with an expected cost of \( Z_{NF}^l \). If the firm is not inspected, which happens with probability \( 1 - \phi \), the firm incurs no cost during this period, and next period follows \( S_{NF} \) again, with an expected cost of \( Z_{NF}^l \). Solving (3) gives

\[ Z_{NF}^l = \frac{\phi(F + \delta c)}{(1 - \delta)(1 + \phi \delta)} \]  

(4)
A low-cost firm is indifferent between following $S_{CT}$ or $S_{NF}$ if $Z_{CT}^l = Z_{NF}^l$. Letting $\tilde{c}$ be the marginal cost that makes $Z_{CT}^l = Z_{NF}^l$, we have that

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

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

$$Z_{CT}^h = p^s + \delta Z_{CT}^h$$

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

$$Z_{CT}^h = \frac{p^s}{1 - \delta}$$

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

\[ Z_{hNF}^h = 0 + \phi(F + \delta p^s + \delta^2 Z_{hNF}^h) + (1 - \phi)(\delta Z_{hNF}^h) \]  

(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 \( Z_{hNF}^h \). If the firm is not inspected, which happens with probability \( 1 - \phi \), the firm does not incur any cost in this period, and next period follows \( S_{NF} \) again with an expected cost of \( Z_{hNF}^h \). Solving (8) gives

\[ Z_{hNF}^h = \frac{\phi(F + \delta p^s)}{(1 - \delta)(1 + \phi \delta)} \]  

(9)

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

Firms’ compliance behaviors can be grouped according to their abatement costs as follows: \textit{compliant} firms have very low abatement costs (i.e., \( c \leq c \leq \bar{c} \)) and always comply by reducing emissions; \textit{non-compliant} firms have medium and high costs (i.e., \( \bar{c} < c \leq \overline{c} \)). A non-compliant 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 by buying permits if its abatement cost is high (i.e., \( p^s < c \leq \overline{c} \)). 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., $c \leq c \leq p^*$) will reduce pollution all the time, and high-cost firms (i.e., $p^* < c \leq \bar{c}$) will always buy permits. Although $\phi F > p$ is possible for low inspection costs and high fines, in this paper we are interested in the case of partial compliance, or incomplete enforcement. Note also that if $\phi = 1$ and $F < p^*$, it is still possible to have a fraction of non-compliant firms.

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

$$q^*_e(x^*) = G(\bar{c}) + \gamma[G(p^*) - G(\bar{c})]$$

(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 non-compliant firms that came into compliance this period by reducing one unit of pollution (subscript “e” refers to effective amount).\textsuperscript{16,17}

\textsuperscript{16}To determine $\gamma$, denote by $K_t$ the number of non-compliant firms (i.e., those firms that follow $S_{NF}$) that are in compliance at date $t$, and by $N_t$ the number of non-compliant firms that are out of compliance at date $t$, and let $K_t + N_t = 1$. In other words, $K_t$ are non-compliant firms that were inspected in $t-1$ and brought under compliance at date $t$. The value of $K_t$ can then be obtained

$$K_t = \phi N_{t-1}$$

Note that in this multi-period model, at $t-1$ there are $N_{t-1}$ firms facing a probability $\phi$ of being inspected. Using $N_{t-1} = 1 - K_{t-1}$ and setting $K_t = K_{t-1}$ for steady state gives

$$K = \frac{\phi}{1 + \phi}$$

\textsuperscript{17}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 non-compliant firms under compliance) increases, $\gamma$’s upper limit approaches 1. That would be the case in our model, if we assume, for example, that the regulator is able to keep the non-compliant firm under compliance for more than one period.
Similarly, the effective control costs incurred by firms are expected to be

\[ C_s^e(x^s) = \int_{\bar{c}}^{e} cdG + \gamma \int_{\bar{c}}^{p^*} cdG \]  

(11)

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

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

\[ x^s = \gamma [1 - G(p^*)] \]  

(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 \( S_{NF} \) strategy that in this period come under compliance by buying permits instead of reducing pollution. Solving (12) gives

\[ p^*(x^s) = G^{-1}(1 - \frac{x^s}{\gamma}) \]  

(13)

where \( G^{-1}(1 - x/\gamma) \) can be viewed as the marginal cost \( c \) just after \( 1 - x^s/\gamma \) units of

---

18To see that grandfathered permits and auctioned permits are equivalent, let us write the market clearing condition under grandfathered permits (each firm receives \( x^s \) permits for free)

\[
x^s G(\bar{c}) + x^s \gamma (G(p^*) - G(\bar{c})) + x^s(1 - \gamma)(1 - G(\bar{c})) = (1 - x)\gamma(1 - G(p^*))
\]

On the left-hand side we have three types of sellers: compliant firms, a fraction \( \gamma \) of non-compliant firms that came into compliance this period by reducing emission so they can now sell their permits in the market, and a fraction \( (1 - \gamma) \) of non-compliant firms that are not in compliance today. On the right-hand side we have the buyers of permits: non-compliant firms that are in compliance this period by buying permits instead of reducing pollution. Developing the expression above yields (12).
pollution have been reduced.\textsuperscript{19} Since the equilibrium price of permits under full compliance would be $G^{-1}(1 - x^s)$, which occurs when firms are in compliance all the time (in this model, when $\gamma = 1$), it is immediate that incomplete enforcement lowers $p^s$. The reason for this result is simply that noncompliance and permits are (imperfect) substitutes, which depresses their demand and price. Note also that if $\gamma = 1$, $G(p^s) = 1 - x^s$ and $q^s_e(x) = 1 - x^s$.

2.2 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, $x^t_1$ and $x^t_2$ 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, $x^t_1 + \alpha x^t_2$, between the two markets. In other words, the regulator just need specify $x^t_{12} = x^t_1 + \alpha x^t_2$ and $\alpha$.

Firms’ compliance when markets work together is not much different from the previous analysis but for the market clearing conditions. If $p^t_1$ and $p^t_2$ are the clearing prices in markets 1 and 2, respectively, the “integrated markets” clearing condition is

$$x^t_{12} = \gamma_1[1 - G_1(p^t_1)] + \alpha \gamma_2[1 - G_2(p^t_2)]$$

\textsuperscript{19}Note that for a uniform distribution of $g(c) = 1/C'' = 1/(\tau - \zeta)$, we have

$$p^s = \tau - \frac{C''}{\gamma} x^s$$
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 is purchases from high-cost firms in market 1 (i.e., \( c_1 > p^t_1 \)) following \( S_{NF} \) strategy and the second term is purchases from high-cost firms in market 2 (i.e., \( c_2 > p^t_2 \)) following the same strategy.

When markets are fully integrated, a firm in either market is indifferent between buying one permit in market 1 at price \( p^t_1 \) than \( 1/\alpha \) permits in market 2 at price \( p^t_2 \), so we have that

\[
p^t_2 = \alpha p^t_1
\]

Thus, from (14) and (15) we can write both \( p^t_1 \) and \( p^t_2 \) as a function of \( x^t_{12} \) and \( \alpha \).

Finally, the effective amount of pollution reduction, \( q^{e}_t(x) \), and effective control costs, \( C^{e}_t(x) \), can be directly obtained, respectively, from eqs. (10) and (11). It only requires replacing \( p^* \) by either \( p^t_1 \) or \( p^t_2 \), where \( p^t_i = p^t_i(x^t_{12}, \alpha) \). Having understood firms’ compliance behavior under incomplete enforcement, we now turn to its effects on optimal instrument design and on instrument choice when the regulator is uncertain about costs and benefits.

### 3 Markets design

In the real world, regulators must choose policy goals and instruments in the presence of significant uncertainty concerning 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
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 certain benefit and cost curves in market \( i \) 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 \( C''_i \equiv \overline{c}_i - \underline{c}_i > 0 \) are all fixed coefficients.\(^{20}\)

Next, let the regulator’s prior belief for the marginal benefit curve be \( B'_i(q_i, \eta_i) = B'_i(q_i) + \eta_i \), 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 \( c_i(\theta_i) = c_i + \theta_i \), 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 \( c_i \in [\underline{c}_i, \overline{c}_i] \), which produces the desired “parallel” shift of the aggregate marginal cost curve, \( C'_i(q_i) \), by the amount \( \theta_i \), that is \( C'_i(q_i, \theta_i) = C'_i(q_i) + \theta_i \). 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.\(^{21}\)

\(^{20}\)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 \( c_i \) in the cost curve.

\(^{21}\)While it is true that the regulator may (imperfectly) deduce uncertainty with a lag from the aggregate behavior of firms, I am assuming that he adheres to the original regulatory design from the beginning of time. Alternatively, we can say that new sources of uncertainty arise continually, so the issue of...
3.1 Designing separated markets

The regulator needs to specify the number of permits $x_1^s$ and $x_2^s$ to be allocated in each market respectively. He considers each market separately and solves (I omit subscript $i$)

$$\max_{x^s} W^s = E [B(q^s_e(x^s, \theta), \eta) - C^s_e(x^s, \theta)]$$

where $q^s_e(x, \theta)$ and $C^s_e(x, \theta)$ can be derived from (10) and (11) as ($g(c) = 1/C'' = 1/(\tau - \zeta)$)

$$q^s_e(x^s, \theta) = \int_{\zeta + \theta}^{\bar{c}} \frac{1}{C''} dc + \gamma \int_{\zeta}^{\bar{c}} \frac{1}{C''} dc$$

$$C^s_e(x^s, \theta) = \int_{\zeta + \theta}^{\bar{c}} \frac{c}{C''} dc + \gamma \int_{\zeta}^{\bar{c}} \frac{c}{C''} dc$$

and where

$$p^s \equiv p^s(x^s, \theta) = p^s(x^s) + \theta = \frac{\gamma \bar{c} - x^s C''}{\gamma} + \theta$$

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

$$E [\bar{b} + \eta + B'' q^s_e(x^s, \theta) - p^s(x^s, \theta)] = 0$$

uncertainty is never resolved. For example, we can let $\theta$ and $\eta$ follow (independent or correlated) random walks. The computation of compliance strategies would be the same, but the computation of the welfare function would differ a bit because the variance of $\theta$ and $\eta$ would grow linearly with time.
where

\[ q_e^s(x^s, \theta) = q_s^e(x^s) - \frac{1 - \gamma}{C''} \theta = 1 - x^s - \frac{(1 - \gamma)(\tau - \phi F)}{C''} - \frac{1 - \gamma}{C''} \theta \]  

(21)

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 \( x^s \) is independent of \( \eta \) and \( \theta \). Eq. (21), on the other hand, shows that under incomplete enforcement (\( \gamma < 1 \)) the effective amount of reduction, \( q_e^s(x^s, \theta) \), 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 \( 1 - x^s \), which is simply baseline emissions minus the number of auctioned permits.

### 3.2 Designing integrated markets

If both markets are designed to work together, the regulator needs specify the total number of permits \( x_{12}^t = x_1^t + \alpha x_2^t \) to be allocated and the exchange rate \( \alpha \) at which permits can be traded. Then, he considers both markets simultaneously and solves

\[
\max_{x_{12}^t, \alpha} W^t = E \left[ \sum_{i=1,2} \left( B_i(q^t_{ei}(x_{12}^t, \alpha, \theta_i, \theta_{-i}), \eta_i) - C^t_{ei}(x_{12}^t, \alpha, \theta_i, \theta_{-i}) \right) \right]
\]

(22)

where \( q_{ei}^t \) and \( C_{ei}^t \) can be obtained, respectively, from (17) and (18) by simply replacing \( p^e \) by the corresponding price \( p^t_i \). From (14) and (15) and assuming \( C''_1 = C''_2 = C'' \), these
prices are

\[ p_1^t(x_{12}^t, \alpha, \theta_1, \theta_2) = \frac{\gamma(\tau_1 + \alpha \tau_2) - x_{12}^t C''}{\gamma(1 + \alpha^2)} + \frac{\theta_1 + \alpha \theta_2}{1 + \alpha^2} \quad (23) \]

\[ p_2^t(x_{12}^t, \alpha, \theta_1, \theta_2) = \alpha p_1^t(x_{12}^t, \alpha, \theta_1, \theta_2) \quad (24) \]

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 \( B_0^1 = B_0^2 = B \), \( C_0^1 = C_0^2 = C \), \( \phi_1 = \phi_2 \), and \( \theta_1 \) and \( \theta_2 \) are i.i.d. and not correlated with \( \eta \), the optimal design when markets are integrated is

\[ x_{12}^t = x_1^s + \alpha x_2^s \]

\[ \alpha = \frac{E[B_2'(q_{12}^t(x_2^s, \theta_2), \eta_2)]}{E[B_1'(q_{11}^t(x_1^s, \theta_1), \eta_1)]} = \frac{E[p_2^t(x_2^s, \theta_2)]}{E[p_1^t(x_1^s, \theta_1)]} \]

**Proof.** See the appendix. ■

Because of the symmetry of the problem the results under Proposition 1 are very intuitive. The first result indicates that the total number of permits is the same under either market design. The second result indicates the 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

\[ E[\eta_1 \theta_1] = E[\eta_2 \theta_2]. \]
of expected prices. Since we do not impose restrictions on the values of $b_i$ and $c_i$, the value of $\alpha$ can be equal, greater or lower than 1. In fact if $b_2 > b_1$ and $c_2 > c_1$, the optimal value of $\alpha$ will be greater than 1.

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

$$p^t_1 = p^s_1 - \frac{\alpha^2 \theta_1 - \alpha \theta_2}{1 + \alpha^2} \quad (25)$$

$$p^t_2 = p^s_2 + \frac{\alpha \theta_1 - \theta_2}{1 + \alpha^2} \quad (26)$$

where $p^s_i$ is given by (19). While expected prices do not vary with market design (i.e. $E[p^t_i] = E[p^s_i]$), actual prices generally do so. For example, if $\theta_1 > 0$ and $\theta_2 = 0$, $p^t_1 < p^s_1$. The equilibrium price $p^t_1$ do not go up as much because under integrated markets those firms with costs between $p^t_1$ and $p^s_1$ will find cheaper to buy permits from market 2 than reducing emissions themselves.

Quantities, on the other hand, are given by

$$q^t_{e1} = q^s_{e1} - \frac{(\alpha^2 \theta_1 - \alpha \theta_2) \gamma}{(1 + \alpha^2)C''} \quad (27)$$

$$q^t_{e2} = q^s_{e2} + \frac{(\alpha \theta_1 - \theta_2) \gamma}{(1 + \alpha^2)C''} \quad (28)$$

where $q^s_{e1}$ is given by (21). Again, while expected quantities do not vary with market
design (i.e. \( E[q_{ei}^t] = E[q_{ei}^s] \)), actual quantities generally do so. For example, if \( \theta_1 > 0 \) and \( \theta_2 = 0 \), \( q_{ei1}^t < q_{ei1}^s \) 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 \( p_i^t \) (or marginal cost) in each market less uncertain (this leads to cost savings in expected terms). At the same time, however, the actual reductions \( q_{ei1}^t \) and \( q_{ei2}^t \) that will take place in each market becomes more uncertain (this leads to benefit losses in expected terms). In deciding whether to have markets integrated and allow interpollutant trading, the regulator will inevitably face this trade-off between cost savings and benefit losses; trade-off that we study more formally in the next section.

4 Integrated vs separated markets

To find the optimal policy design, we start by writing the difference in expected welfare between the two market designs (integrated and separated markets)

\[
\Delta_{ts} = W_{12}^t(x_{12}, \alpha) - (W_1^s(x_1^s) + W_2^s(x_2^s)) \tag{29}
\]

where \( x_{12}^t, \alpha, x_1^s \) and \( x_2^s \) 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.

To explore under which conditions this is the case, we conveniently rewrite (29) as

\[
\Delta_{ts} = E \left[ \sum_{i=1,2} \left( \{ B_i(q_{ei1}^t, \eta_i) - B_i(q_{ei1}^s, \eta_i) \} - \{ C_{ei1}^t - C_{ei1}^s \} \right) \right] \tag{30}
\]

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

\[
\Delta_{ts} = E \left[ \sum_{i=1,2} \left( \left\{ (b_i + \eta_i)(q_{ei}^i - q_{ei}^s) + \frac{B''}{2} ((q_{ei}^i)^2 - (q_{ei}^s)^2) \right\} - \left\{ \gamma \int_{p^t_i}^{p^s_i} \frac{c}{C''} dc \right\} \right) \right] \tag{31}
\]

where \( p^t_i, p^s_i, q_{ei}^s \) and \( q_{ei}^t \) can be obtained from (25)–(28).

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

\[
\Delta_{ts} = \gamma (2 - \gamma) E[\theta^2]B'' + \gamma E[\theta^2]C'' \tag{32}
\]

where 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

\[
\Delta_{ts} = \gamma E[\theta^2] \frac{B''}{2(C'')^2} ((2 - \gamma)B'' + C'') \tag{33}
\]

where \( \gamma = \phi/(1 + \phi) < 1 \) is the fraction of non-compliant firms that are in compliance today, \( E[\theta^2] \) is the variance of the cost shocks in either market, \( B'' < 0 \) is the slope of the marginal benefit curves and \( C'' > 0 \) is the slope of the marginal cost curves. We summarize the above result in the following proposition

**Proposition 2** 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 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., tradeable 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., emission reduction), and therefore, he should use the price instrument.

The exact 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. Then, if the marginal cost curves are steeper than the marginal benefit curves, the regulator should pay more attention to cost of control rather than 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 incomplete enforcement ($\gamma < 1$) has important effects on the multipollutant markets design. Since $2 - \gamma > 1$, (33) indicates that incomplete 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 \(q_{ei}^t\) becomes relatively less uncertain than \(q_{ei}^s\). 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 cost shocks. 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 \((\text{i.e.}, \gamma = 0)\).

Finally, we can relax some of the assumptions regarding correlations between the different stochastic terms. If \(E[\eta_i \theta_i] > 0\),\(^{23}\) a negative term enters into (32) increasing the advantage of separated markets; otherwise, benefit uncertainty does not intervene.\(^{24}\) In addition, if \(E[\theta_i \theta_{-i}] > 0\),\(^{25}\) a term of opposite sign enters into (32) reducing the welfare difference between the two market designs.\(^{26}\)

\(^{23}\) A discussion on whether this correlation is more likely to be positive or negative is found in Stavins (1995).

\(^{24}\) If \(E[\eta_i \theta_i] = E[\eta_i \theta_2]\), the extra term is \(-\gamma E[\eta_i \theta_i]/C''\).

\(^{25}\) 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 NO\(_x\) and SO\(_2\).

\(^{26}\) The extra term is

\[-\frac{\alpha \gamma((2 - \gamma)B'' + C'')}{(1 + \alpha^2)(C'')^2} E[\theta_i \theta_2]\]

Note that because \(\alpha/(1 + \alpha^2) \leq 1/2\) for all \(\alpha > 0\) and \(E[\theta_i \theta_2] < E[\theta_1 \theta_1] = E[\theta_2 \theta_2]\), this extra term can never revert the policy choice prescribed by (33).


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 emission 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 flatter than the marginal benefit curves, which seems to be the case for some urban air pollutants, the regulator should have multipollutant markets integrated through optimal exchange rates, unless the level of enforcement is too weak, in which case, the regulator should keep markets separated.

The results of the paper are also relevant for a regulator that is implementing a multipollutant offset system where new firms must compensate 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 all its PM10 and NO\textsubscript{x} emissions with PM10 credits; question that also involves specifying that appropriate exchange rate between PM10 and NO\textsubscript{x} 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 our recommendations in Section 3.

I hope that the results of this paper 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. We leave an application of this model to the case of air pollution in Santiago-Chile for further research.

References


[16] South Coast Air Quality Management District (SCAQMD, 1993), Program Summary and Rules, Los Angeles, CA.


Appendix

To demonstrate Proposition 1, I proceed in two steps. Since I have already shown in the text that first-order conditions for \( x^s_1 \) and \( x^s_2 \) 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 \( x^t_{12} \) and \( \alpha \) are independent of \( \eta \) and \( \theta \).

Under certainty, the first-order conditions for \( x^t_{12} \) and \( \alpha \) are, respectively

\[
\left( b_1 + B'' q_{e1}^t - p_1^t \right) \frac{\partial p_1^t}{\partial x^t_{12}} + \left( b_2 + B'' q_{e2}^t - p_2^t \right) \frac{\partial p_2^t}{\partial x^t_{12}} = 0 \tag{34}
\]

\[
\left( b_1 + B'' q_{e1}^t - p_1^t \right) \frac{\partial p_1^t}{\partial \alpha} + \left( b_2 + B'' q_{e2}^t - p_2^t \right) \frac{\partial p_2^t}{\partial \alpha} = 0 \tag{35}
\]

Since both \( \frac{\partial p_i^t}{\partial x^t_{12}} \) and \( \frac{\partial p_i^t}{\partial \alpha} \) are different from zero, the solution of the above system of equations satisfies

\[
b_1 + B'' q_{e1}^t - p_1^t = 0 \tag{36}
\]

But (36) is first order condition for \( x^s_i \) (see (20)), which implies that under certainty we have

\[
p_i^t = p_i^s \tag{37}
\]

Using (37) and \( p_2^t = \alpha p_1^t \) (by an arbitrage condition of integrated markets), it follows
that \( \alpha = p_2^s/p_1^s \), where \( p_2^s \) is equal to the value of \( E[p_i^s] \) obtained from (19). Thus, we have that under certainty

\[
p_i^s = \frac{\gamma c_i - x_i^s C''}{\gamma}
\]

(38)

therefore

\[
p_1^s + \alpha p_2^s = (1 + \alpha^2)p_1^s = \frac{\gamma (c_1 + \alpha c_2) - (x_1^s + \alpha x_2^s) C''}{\gamma}
\]

(39)

Comparing (39) with the deterministic part of (23), it follows that \( x_{12}^t = x_1^s + \alpha x_2^s \) under certainty.

Proceeding with the second steep of the proof, the first-order conditions for \( x_{12}^t \) and \( \alpha \) under uncertainty are, respectively

\[
E \left[ \sum_{i=1,2} \left\{ (b_i + \eta_i + B'' q_{ei}(\theta_i, \theta_{-i}) - p_i^t(\theta_i, \theta_{-i})) \frac{\partial p_i^t(\theta_i, \theta_{-i})}{\partial x_{12}^t} \right\} \right] = 0
\]

(40)

\[
E \left[ \sum_{i=1,2} \left\{ (b_i + \eta_i + B'' q_{ei}(\theta_i, \theta_{-i}) - p_i^t(\theta_i, \theta_{-i})) \frac{\partial p_i^t(\theta_i, \theta_{-i})}{\partial \alpha} \right\} \right] = 0
\]

(41)

where

\[
q_{e1}^t(\theta_1, \theta_2) = A_1 + \frac{(\gamma - 1 - \alpha^2)\theta_1 + \alpha \gamma \theta_2}{(1 + \alpha^2)C''}
\]

(42)
\[ q_{t_2}^i(\theta_1, \theta_2) = A_2 + \frac{\alpha \gamma \theta_1 + (\alpha^2 \gamma - 1 - \alpha^2)\theta_2}{(1 + \alpha^2)C''} \] (43)

\[ p_1^i(\theta_1, \theta_2) = D_1 + \frac{\theta_1 + \alpha \theta_2}{1 + \alpha^2} \] (44)

\[ p_2^i(\theta_1, \theta_2) = D_2 + \frac{\alpha \theta_1 + \alpha^2 \theta_2}{1 + \alpha^2} \] (45)

\[ \frac{\partial p_1^i(\theta_1, \theta_2)}{\partial x_{12}^i} = -\frac{C''}{(1 + \alpha^2)\gamma} = \frac{1}{\alpha} \frac{\partial p_2^i(\theta_1, \theta_2)}{\partial x_{12}^i} \] (46)

\[ \frac{\partial p_1^i(\theta_1, \theta_2)}{\partial \alpha} = H_1 + \frac{-2\alpha \gamma^2 \theta_1 + \gamma^2(1 - \alpha^2)\theta_2}{[(1 + \alpha^2)\gamma]^2} \] (47)

\[ \frac{\partial p_2^i(\theta_1, \theta_2)}{\partial \alpha} = H_2 + \frac{\gamma^2(1 - \alpha^2)\theta_1 + 2\alpha \gamma^2 \theta_2}{[(1 + \alpha^2)\gamma]^2} \] (48)

where \( A_i, D_i \) and \( H_i \) 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 \( E[\theta_i^2] \) terms. On the one hand, the first-order condition (40) does not include any \( E[\theta_i^2] \) terms because \( \partial p_i^i/\partial x_{12}^i \) is independent of \( \theta_i \) and \( \theta_{-i} \), as shown by (46). The first-order condition (41), on the other hand, does include several \( E[\theta_i^2] \) terms because \( \partial p_i^i/\partial \alpha \) depends on \( \theta_i \) and \( \theta_{-i} \), as indicated by (47) and (48). However, the multiplicative interaction of \( \partial p_i^i/\partial \alpha \) with \( q_{ei}^i \) and \( \partial p_i^i/\partial \alpha \) with \( p_i^i \) results in a total of eight \( E[\theta_i^2] \) and
$E[\theta_2^2]$ terms that cancel out when $E[\theta_1^2] = E[\theta_2^2]$. This demonstrates that (41) 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 (41) that cancel out when $E[\theta_1 \eta_1] = E[\theta_2 \eta_2]$).