Optimal Design of Permit Markets with an *Ex Ante* Pollution Target

Sergey S. Rabotyagov, Hongli Feng, and Catherine L. Kling

Working Paper 06-WP 430
August 2006

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

Sergey Rabotyagov is a graduate assistant, Hongli Feng is an associate scientist and adjunct assistant professor of economics, and Catherine Kling is professor of economics and head of the Resource and Environmental Policy Division in the Department of Economics and Center for Agricultural and Rural Development at Iowa State University.

The authors appreciate helpful comments from Kerry Smith and from participants of the following workshops and paper sessions: Camp Resources XIV, the 2006 American Agricultural Economics Association Selected Papers Session, and the Iowa State University Environmental Economics workshop. Rabotyagov acknowledges support from the U.S. Environmental Protection Agency Science to Achieve Results Graduate Fellowship #91659501-0.

Questions or comments about the contents of this paper should be directed to Sergey Rabotyagov, 260 Heady Hall, Iowa State University, Ames, IA 50011-1070; Ph: (515) 294-8612; Fax: (515) 294-0221; E-mail: rabotyag@iastate.edu.

Abstract

In this paper, we examine the design of permit trading programs when the objective is to minimize the cost of achieving an *ex ante* pollution target, that is, one that is defined in expectation rather than an *ex pos*t deterministic value. We consider two potential sources of uncertainty, the presence of either of which can make our model appropriate: incomplete information on abatement costs and uncertain delivery coefficients. In such a setting, we find three distinct features that depart from the well-established results on permit trading: (1) the regulator's information on firms' abatement costs can matter; (2) the optimal permit cap is not necessarily equal to the *ex ante* pollution target; and (3) the optimal trading ratio is not necessarily equal to the delivery coefficient even when it is known with certainty. Intuitively, since the regulator is only required to meet a pollution target on average, she can set the trading ratio and total permit cap such that there will be more pollution when abatement costs are high and less pollution when abatement costs are low. Information on firms' abatement costs is important in order for the regulator to induce the optimal alignment between pollution level and abatement costs.

Keywords: delivery coefficient, *ex ante* pollution target, *ex post* pollution target, permit trading, total permit cap, trading ratio.

JEL codes: Q58, D02

1. Introduction

A highly celebrated property of emissions trading markets is that decentralized decisions made by firms will achieve a preset emissions target at the least possible cost and no information on the firm's abatement costs is required to achieve this outcome (Baumol and Oates, 1988; Montgomery, 1972).[1] Montgomery (1972) demonstrates that this property extends to the class of non-uniformly mixed pollutants, pollutants whose damages differ based on their location. He shows that if the regulatory authority allows firms to trade emissions according to the ratio of delivery coefficients (the effect that a source's emissions have on resulting pollution loadings) and sets the pollution cap equal to the desired pollution standard, the least-cost property is retained.

The basic model underlying these findings assumes that the regulator is interested in minimizing the cost of meeting an *ex post* environmental standard. While *ex ante* uncertainty regarding a firm's abatement costs is commonly used to motivate the attractiveness of a permit system, the pollution constraint is typically specified in *ex post* terms—the environmental target is invariant with respect to realizations of any sources of uncertainty. As has long been recognized, characterization of the objective function in this way requires that the pollution control level is independent of the actual realization of costs—no trade-off between abatement costs and benefits (pollution levels) is permitted.

In this paper, we study the optimal design of a permit trading system when the regulator is uncertain about the firms' abatement costs and specifies her objective function based on minimizing expected costs subject to meeting an expected pollution level. We call this an *ex ante* target. Our model is applicable to cases where the regulator possesses some information about

[1] The total permit quantity can be set at the socially efficient level, a legally mandated requirement, or any other level deemed appropriate by the regulator.

the firms' abatement costs, but is uncertain about their magnitude either because of the existence of genuine aleatory uncertainty in abatement costs or for reasons of asymmetric information, where regulator's uncertainty is epistemic in nature.

Several striking findings emerge from our model. First, the optimal total permit cap does not necessarily equal the regulator's pollution target. One is an *ex ante* concept (the desired pollution level) while the other is an *ex post* construct (the emissions cap). This can be viewed as a two-stage decision where in the first time period the regulator settles on a desired pollution target and then, based on the firms' expected emissions decisions, sets the number of permits and trading ratio to implement the market.

Second, the optimal trading ratio depends on the moments of the uncertain costs as well as the delivery coefficients. Surprisingly, even when the delivery coefficients are assumed to be known with certainty, it is not optimal to set trading ratios equal to the simple ratio of delivery coefficients—the basic Montgomery (1972) solution. Instead, the regulator can lower expected costs by including some information on the uncertain abatement costs in the formation of the trading ratio.[2]

These somewhat surprising findings come directly from the fact that our regulator's objective function is specified in *ex ante* terms: she minimizes expected costs subject to an expected pollution level. This allows the regulator flexibility that is not present when emission levels must be met with certainty.[3] In essence, this allows the regulator to anticipate, at least to some degree, the actual cost realizations of firms: if costs are unexpectedly high (low), the

[2] That the optimal trading ratio depends on both the regulator's information about costs and the delivery coefficients is consistent with the findings of Horan and Shortle (2005) and Malik et al. (1993), although we do not assume perfect information on costs.

[3] In this way, our model and findings are in the spirit of Roberts and Spence (1976) and Montero (2001) who each recognize that rigidity of a quantity mechanism may be socially costly. Roberts and Spence propose a penalty for exceeding the pollution cap, while Montero models incomplete enforcement to provide a softening of the quantity constraint.

resulting pollution levels will be higher (lower) than they would be without this flexibility. Intuitively, once the regulator is interested in both costs and benefits, it becomes optimal for the regulator to design the system so that if costs are unexpectedly high (a big positive stochastic shock), higher-than-expected pollution levels are permitted. In considering this trade-off, the regulator recognizes that the ultimate abatement levels chosen by firms will depend upon their cost realization, and therefore the ultimate emission levels become stochastic from the regulator's perspective. By choosing the parameters of the trading program to be a function of the moments of the distribution of costs, the regulator can lower total expected abatement costs, while ensuring that the environmental goal is still being met on the average.[4]

Whether the regulator has (or should have) the freedom to design a permit market that allows the aforementioned flexibility is a policy question that will have a case-by-case answer. However, there are many real-world examples where averages over time or space define standards. Examples include carbon monoxide (with both an 8-hour and 1-hour average standard), nitrogen dioxide (an annual arithmetic mean), ozone (1- and 8-hour averages), lead (quarterly average during the phase-out), and sulfur dioxide (annual means, a 24- hour average, and a 2-hour average) (USEPA, 2006a). Examples from water pollution abound as well: values for arsenic, cadmium, cyanide, and selenium emissions in storm water under the National Pollutant Discharge Elimination System (a key regulatory program that regulates point sources of water effluents) trigger need for action only when the annual average exceeds the benchmark (USEPA, 2006b). Indeed, when the values of the delivery coefficients are uncertain, expected values is the only meaningful way to form pollution constraints.

[4] Note that we do not consider the important problem of information extraction from firms but assume that the regulator has some independent source of cost information. See Montero (2000) or Lewis (1996) for careful discussions of asymmetric information problems.

In the next section of the paper, we present the basic model of firms' behavior under a tradable emissions program and the regulator's problem. In section 3, we examine the optimal permit market design under two different assumptions. First, we consider the case in which the delivery coefficient is known. This provides results that contrast with the *ex post* standards studied in Baumol and Oates (1988) and Montgomery (1972), highlighting the implication of using *ex ante* targets and objective functions. Second, we consider the important case in which the delivery coefficient is uncertain. While this latter feature is typically viewed as a characteristic of nonpoint sources, there are likely many point sources for which the true impact of emissions from the source are known with less-than-perfect certainty such as air sheds where dispersion of particulates may depend on stochastic weather conditions (Foster and Hahn, 1995). Final remarks and conclusions complete the paper in section 4.

2. The model

Suppose there are two firms acting as sources of emissions and the environmental impacts of the two firms' emissions are not identical. Specifically, we assume that the impact of the first firm on the resulting pollution level is such that one unit of Firm 1's emissions increases the resulting pollution level by one unit. The impact of Firm 2 is described by the delivery coefficient d, that is, one unit of Firm 2's emissions increases the resulting pollution level by d units. The delivery coefficient can be thought of as describing the relative environmental impact of the two firms' emissions. Specifically, the total resulting pollution level is $e_1 + de_2$, where e_i for $i = 1,2$ represents Firm *i*'s emissions. We model both the situation in which the delivery coefficient is fixed and known by the regulator, as well as a more realistic case wherein the delivery coefficient is random. In the latter case the regulator, however, knows the distribution of the delivery coefficient: its mean, $E(d) = \mu$, and its variance, $Var(d) = \sigma_d^2$. The model lends itself to multiple interpretations, including (1) two firms located spatially apart whose emissions contribute differentially to loadings at the receptor (Baumol and Oates, 1988); (2) two firms whose emissions contribute differentially to loadings for reasons other than spatial location, such as production process or concentration of emissions released; or (3) two firms, of which one is a point source and the other is a nonpoint source with an uncertain delivery coefficient.[5]

The abatement cost function for Firm *i* is $C_i(e_i^0 - e_i; \theta_i)$, where, for $i = 1, 2$, e_i^0 represents the initial (unregulated) emissions level for firm *i* and $e_i^0 - e_i$ represents the abatement of Firm *i* after the implementation of a permit trading program. The abatement cost function is assumed to be increasing and convex in abatement, that is, $C_i' > 0$ and $C_i'' \geq 0$. The parameter (θ_i) in the cost function captures the information uncertainty regarding the costs of pollution abatement on the regulator's side. We assume that the regulator has some, albeit incomplete, information on abatement costs. While throughout our model the formal depiction of the regulator's uncertainty is unchanged, we can endow our formal modeling of uncertainty with two different interpretations: asymmetric information or stochastic information not revealed at the design stage of the permit market but revealed at the time permit trading decisions are made. Formally, when making decisions, firms know θ_1 and θ_2 while the regulator knows only their distribution: the means (zero), variances (σ_1^2 and σ_2^2), and covariance, ($\text{cov}(\theta_1, \theta_2)$). Furthermore, the regulator is assumed to know the covariances, if any, between the delivery coefficient and the cost parameters: $\text{cov}(d, \theta_1)$ and $\text{cov}(d, \theta_2)$. Such correlations may arise, for example, when weather affects the efficacy and cost of abatement as well as its spatial impacts.

[5] Because of the inherent unobservability of nonpoint source pollution, the focus has been on the trading in expected emissions in the nonpoint-source literature (e.g., Horan and Shortle, 2005). See footnote 9 for a related discussion.

2.1. Ex ante and ex post pollution targets, total permit cap, and actual pollution level

Since cost-minimizing firms equate marginal abatement costs with permit prices in order to choose their emission levels, once uncertainty is introduced into the cost functions, there is uncertainty in emission levels and it is necessary to clearly differentiate between *ex ante* and *ex post* measures of pollution as well as other constraints that relate to the design of an emissions trading system. Only one of the two constraints will be relevant for a particular policy. The two constraints can be written as

(1) (Ex ante pollution constraint) $E[e_1] + E[de_2] \le \bar{P}_{ante}$,
(Ex post pollution constraint) $e_1 + de_2 \le \bar{P}_{post}$.

If the pollution target is specified in an *ex ante* manner, the first equation in (1) describes the constraint and indicates that the expected pollution has to be less than or equal to a pre-fixed target ($\bar{P}_{ante}$). Under this constraint, the *ex post* realization of the pollution level can be greater or less than the target. In contrast, if the constraint is specified as *ex post*, the realized *ex post* pollution levels must be less than a pollution target ($\bar{P}_{post}$) in each realization.

A third relevant constraint defines the restriction faced by the permit market:

(2) (Permit market constraint) $e_1 + te_2 \le \bar{P}_{permit}$.

Here, t is the trading ratio for the emissions of the two firms—1 unit of Firm 2's emissions is equivalent to t units of Firm 1's emissions—and $\bar{P}_{permit}$ is the total permit cap, denominated in terms of Firm 1's emissions. Thus, this constraint requires that total emissions (weighted by the trading ratio) be less than or equal to the total permit cap. Note that the firms are only concerned with the permit market constraint while the regulator will care predominately about the pollution constraint (either the *ex ante* or *ex post* version). Finally, $e_1 + de_2 = P_{actual}$ specifies the actual realization of pollution given firms' emissions decisions and the realization of the delivery coefficient.

With perfect information, there is no distinction between *ex ante* and *ex post* and we know from Montgomery (1972) that efficiency dictates that we set $t = d$, resulting in $\bar{P}_{ante} = \bar{P}_{\mathrm{post}} = \bar{P}_{permit}$. That is, the three constraints are essentially the same. However, when there is incomplete information, either $\bar{P}_{post}$ or $\bar{P}_{ante}$ may be used as a target in pollution reduction policies, resulting, as we will show, in very different efficient designs for a permit program. If it is legally stipulated, or the damage function dictates, that pollution not exceed a deterministic, prefixed standard, then $\bar{P}_{post}$ is the relevant constraint for cost minimization. This is the commonly analyzed case when total pollution is limited to a prefixed cap, regardless of firms' abatement costs. As discussed in the introduction, there are many examples of standards that are framed in terms of averages, suggesting that such an inflexible target may not be appropriate or necessary in many cases.

Given an *ex ante* target, the regulator potentially has the flexibility to issue permits, $\bar{P}_{permit}$, and set the trading ratio, t, to achieve the expected pollution target at least cost. Figure 1 illustrates the decision process and the occurrence of events:

This sequential timing process makes clear that the actual pollution, P_{actual}, varies with the realization of firms' abatement costs and/or the delivery coefficient, whereas $\bar{P}_{ante}$ (or $\bar{P}_{\mathrm{post}}$, or $\bar{P}_{permit}$) must be set before the realization of these uncertainties and will not change when the uncertainties are resolved.

2.2. *Firms' emission decisions in a permit trading market*

Should an emissions trading program be introduced, the firms will face the permit market constraint in (2). Suppose the initial permit endowments allocated to Firm *i* (and denominated in Firm *i*'s emissions) are $\bar{e}_i$ for $i=1,2$; and $\bar{e}_1 + t\bar{e}_2 = \bar{P}_{permit}$. Through trading, both firms can hold the permits denominated in terms of another firm's emissions, and the trading ratio is used to convert between the two types of permits. The trading program requires that each firm's actual emissions do not exceed its holding of permits. Let y_i, denominated in terms of Firm *i*'s emissions, denote the equilibrium quantity of permits traded. Specifically, y_i is the permit quantity sold by Firm *i* and purchased by the other firm. Assuming that each firm takes permit prices as given, then Firm 1's problem would be as follows:

$$\begin{aligned} &\min_{e_1, y_1, y_2} C_1(e_1^0 - e_1) - p_1 y_1 + p_2 y_2 \\ &\textit{subject to} \;\; e_1 + y_1 - t y_2 \leq \bar{e}_1. \end{aligned} \tag{3}$$

Firm 2's problem is similar. Solving for the firms' problems, it is well known that market equilibrium requires that $MC_i \equiv C_i'(e_i^0 - e_i^*) = p_i$, for $i=1,2$; and $p_1/p_2 = 1/t$. This implies that the ratio of permit prices must be equal to the trading ratio. Otherwise, costless arbitrage opportunities would be available to firms. Then, we have

$$\frac{MC_2}{MC_1} = t\,. \tag{4}$$

From (4) and the permit market constraint in (2), we can solve for firms' optimal emissions as a function of t and $\bar{P}_{permit}$, that is, $e_i^*(t, \bar{P}_{permit}; \theta_1, \theta_2)$ for $i=1,2.$ When emission decisions are made in the permit trading market, firms have complete information about their costs, i.e., θ_1 and θ_2 are known with certainty. Equation (4) indicates that the results of the permit trading market are such that the ratio of marginal costs equals the trading ratio. However, with complete information on θ_1 and θ_2, we know from Montgomery (1972) that social efficiency requires $t = d$, resulting in

(5) $$\frac{MC_2}{MC_1} = d\ .$$

Any gains in setting t at a level other than d in an *ex ante* targeting program would need to be weighed against the efficiency costs of not attaining the equality in (5). This is an issue we will return to in the next section.

2.3. *The regulator's problem*

Our paper focuses on the design of permit trading programs where the goal is to reach an environmental target at the lowest cost when the target is set as an *ex ante* pollution level, rather than an *ex post* standard.[6] When damage is linear in pollution, the solution to our problem coincides with the solution to the problem of minimizing the sum of damage and abatement costs. While we believe the conditions of uncertainty we model are representative of a broad variety of environmental pollutants, water quality provides a strong motivating example. Imagine there are two sources of effluent that enter a river: source 1 is a large "point" source that is located at the river's edge and source 2 is a "nonpoint" source that is located some distance from the river. Given the proximity of source 1 to the river, its delivery coefficient is known with certainty to be unity whereas the nonpoint nature of source 2 means that the delivery coefficient is uncertain because of weather variability.

For the situation analyzed most in the permit trading literature where the delivery coefficient is known and an *ex post* pollution target is used, the regulator must set the trading

[6] We focus on the design of permit trading programs in the context of cost-effectiveness for the same reason as typically provided in the literature. Pollution targets are often set by political processes as in the case of sulfur permit trading program or water quality trading programs (Horan and Shortle, 2005) and in practice the social damage of pollutants is often unknown making cost minimization the most relevant policy approach.

ratio equal to d and $\bar{P}_{permit} = \bar{P}_{post}$ if she does not have complete information on firms' abatement costs. Otherwise, there is no guarantee that the target will be met. This is because, from (1), we know that

(6) $$\bar{P}_{permit} - \bar{P}_{post} = (t - d)e_2 .$$

If $t = d$, then $\bar{P}_{post} = \bar{P}_{permit}$, regardless of the value of e_2. However, if the regulator is to set $t \neq d$, then she needs to adjust $\bar{P}_{permit}$ as well so that the *ex post* pollution target will be met. However, any adjustment will depend on the magnitude of e_2, which is assumed unknown to the regulator when designing the permit market (because of uncertain abatement costs).

Interestingly, as mentioned in the introduction, it is not even feasible to use an *ex post* pollution constraint if the delivery coefficient is uncertain. This is because, for any given $(t,\ \bar{P}_{permit})$, the value of $\bar{P}_{post}$ will vary with d and e_2. While the realization of d is affected by weather conditions, the decision regarding e_2 depends on $(t,\ \bar{P}_{permit})$ and the parameters of the abatement cost function. Thus, there may be different realizations of d for the same value of e_2. It is then obvious that (6) will not hold for all possible values of d and e_2 in a permit trading program. In this case, an *ex ante* constraint is the only meaningful policy option.

When an *ex ante* pollution target is used, the realization of total pollution can be higher or lower than the target. Even though the regulator cannot directly control the realization of total pollution, she may be able to set the parameters of the permit system (t and $\bar{P}_{permit}$) in conjunction with her (incomplete) knowledge of the firms' abatement costs to generate higher-than-average emission levels when firms' abatement costs turn out to be high and vice versa. Formally, we can set up the regulator's problem as follows:

(7) $$\min_{t,\bar{P}_{permit}} \quad E[TC] \equiv E\Big[C_1(e_1^*(t,\bar{P}_{permit};\theta_1,\theta_2)+C_2(e_2^*(t,\bar{P}_{permit};\theta_1,\theta_2)\Big],$$

subject to *Ex ante pollution constraint in (1).*

Note that firms' emission decisions, $e_i^*(t,\bar{P}_{permit};\theta_1,\theta_2)$ for $i=1,2$, are incorporated into the regulator's program. We next explore the optimal trading ratio and total permit cap.

3. Optimal permit trading ratio and total permits

For tractability, we assume that one firm faces a linear abatement cost function while the other faces an increasing convex abatement cost function, as specified below:[7]

(8) $$C_1(e_1^0-e_1,\theta_1)=(a+\theta_1)(e_1^0-e_1),$$

(9) $$C_2(e_2^0-e_2,\theta_2)=(b+\theta_2)(e_2^0-e_2)+c(e_2^0-e_2)^2.$$

In (8), we assume that $a^2-\sigma_1^2>0$, that is, the mean of the marginal abatement cost (which represents the deterministic part) dominates the variance (which represents the stochastic part). This assumption also ensures that the second-order condition for the problem in (7) is satisfied. With the above cost functions, we can derive firms' optimal emissions from equation (4) and the permit market constraint in (2):

(10) $$e_1^*(t,\bar{P}_{permit};\theta_1,\theta_2)=\frac{2c(\bar{P}_{permit}-e_2^0t)-(b+\theta_2)t+t^2(a+\theta_1)}{2c}$$

(11) $$e_2^*(t,\bar{P}_{permit};\theta_1,\theta_2)=\frac{2ce_2^0+(b+\theta_2)-t(a+\theta_1)}{2c}.$$

[7] As will be clear later, the linear-quadratic setup used is sufficiently rich while remaining simple enough for the intuitive and graphical discussion that follows.

Clearly, the amount of emissions generated by the firms depends on regulatory decisions on t and $\bar{P}_{permit}$ as well as the values of the parameters θ_1 and θ_2. Note that e_2^* does not vary with $\bar{P}_{permit}$, implying that when $\bar{P}_{permit}$ is altered, e_1^* will absorb all the changes in $\bar{P}_{permit}$, i.e.,

(12) $$\frac{\partial e_1^*}{\partial \bar{P}_{permit}} = 1 \qquad \text{and} \qquad \frac{\partial e_2^*}{\partial \bar{P}_{permit}} = 0 .$$

This feature comes directly from the linearity in (8), and while not likely typical of real world situations, it makes the analysis tractable with no obvious loss of generality. We also note that

(13) $$\text{(i) } \frac{\partial e_1^*}{\partial t} \le 0 \text{ iff } -1 \le \varepsilon_{e_2,t} \equiv \frac{t}{e_2^*}\frac{\partial e_2^*}{\partial t}, \text{ and (ii) } \frac{\partial e_2^*}{\partial t} < 0 .$$

Part (ii), derived by differentiating (11) with respect to *t*, implies that e_2^* decreases as t increases. Since $e_1^* + te_2^* = \bar{P}_{permit}$, the sign of $\partial e_1^* / \partial t$ is the opposite of $\partial\left(te_2^*\right)/\partial t$. Thus, for any given $\bar{P}_{permit}$, $\partial e_1^* / \partial t \le 0$ if and only if the elasticity of e_2^* with respect to t is greater than or equal to -1. Given that e_i^* is adjusted when t and $\bar{P}_{permit}$ change, the regulator can adjust the total actual pollution level by changing these policy variables. This important fact will be discussed in detail below.

With analytical solutions in (10) and (11) for the firms' choice of emission levels, it is straightforward to solve the *ex ante* optimization problem (7)[8] and derive the optimal trading ratio and permit cap. First, we obtain the optimal trading ratio as a function of the regulator's prior information on the covariance structure of abatement cost uncertainties and the delivery coefficient, or specifically,

[8] The problem is a standard optimization problem with one constraint and so the details on the derivation of the solutions are not presented. To simplify our discussions, interior solutions are assumed throughout the paper, unless otherwise noted.

$$(14) \qquad t^* = \mu + \frac{1}{a^2 - \sigma_1^2}\left(\mu\sigma_1^2 + a\,\mathrm{cov}(d,\theta_1) - \mathrm{cov}(\theta_1,\theta_2)\right).$$

To derive the optimal permit cap, we first note that as long as the program is intended to reduce emissions, both the *ex ante* pollution constraint in (1) and the market permit constraint in (2) will be binding. Then, we can derive the following:

$$(15) \qquad \overline{P}^*_{permit} - \overline{P}_{ante} = E[e_2^*](t^* - \mu) + \frac{Cov(d,\theta_2) - t^*Cov(d,\theta_1)}{2c}.$$

The details are provided in the appendix. Equations (14) and (15) imply that with known values of θ_1, θ_2, and d (with $d = \mu$), the optimal trading ratio would be set equal to the delivery coefficient, i.e., $t^* = d(= \mu)$, and the total permit quantity allocated to firms would equal the pollution target, i.e., $\overline{P}^*_{permit} = \overline{P}_{ante}$. However, in general, $t^* \neq d$ and $\overline{P}^*_{permit} \neq \overline{P}_{ante}$. This, of course, differs starkly from most permit trading programs in which the trading ratio is the same as the delivery coefficient and the total permit cap is set the same as the pollution target that the regulator sets out to achieve.

3.1 *The total pollution effect and the deadweight loss effect*

To see the effects of setting $t \neq d$ and $\overline{P}_{permit} \neq \overline{P}_{ante}$, we use a benchmark permit trading program where $t = d$ and $\overline{P}_{permit} = \overline{P}_{ante}$. The total abatement costs as a result of implementing the benchmark trading program and as a result of any other permit trading program are denoted as $TC\left(d, \overline{P}_{ante}\right)$ and $TC\left(t\ , \overline{P}_{permit}\right)$, respectively. The difference between these total costs can be broken down as follows:

$$TC\left(d,\bar{P}_{ante}\right)-TC\left(t,\bar{P}_{permit}\right)$$

(16)
$$=\underbrace{TC\left(d,\bar{P}_{ante}\right)-TC\left(d,\bar{P}_{actual}\right)}_{total\ pollution\ effect}+\underbrace{TC\left(d,\bar{P}_{actual}\right)-TC\left(t,\bar{P}_{permit}\right)}_{deadweight\ loss\ effect},$$

where $\bar{P}_{actual}$ is the actual amount of pollution resulting from a trading program with $\left(t,\bar{P}_{permit}\right)$, that is,

(17) $\bar{P}_{actual}=e_1^*(t,\bar{P}_{permit})+de_2^*(t,\bar{P}_{permit})$ and $\bar{P}_{permit}=e_1^*(t,\bar{P}_{permit})+te_2^*(t,\bar{P}_{permit})$.

The total cost of each trading program is derived by using firms' emissions decisions under the program, for example, $TC\left(d,\bar{P}_{ante}\right)=C_1(e_1^*(d,\bar{P}_{ante}))+C_2(e_2^*(d,\bar{P}_{ante}))$.

The *total pollution effect* represents the cost difference that is due to the deviation of the total pollution level from the benchmark program. This deviation can occur because the total permit cap is not equal to that of the benchmark program and/or because $t\neq d$. The latter causes a divergence because when t is used in the permit market constraint instead of d, one unit of permit is no longer necessarily the same as one unit of pollution. The total pollution level will equal the permit cap in the benchmark program. However, this is not necessarily true in a program with a trading ratio not equal to *d*.

The *deadweight loss effect* is directly linked to the use of t as the trading ratio, instead of the delivery coefficient d, which leads to a suboptimal allocation of emissions as we pointed out in section 2.2. We refer to this effect as the deadweight loss effect since it represents the extra cost incurred by using t to achieve the same amount of total pollution ($\bar{P}_{actual}$). Given an *ex ante* pollution target, the regulator would like to induce a high pollution level when abatement cost turns out to be high and vice versa, that is, to exploit the total pollution effect by setting $t\neq d$ and $\bar{P}_{permit}\neq\bar{P}_{ante}$. However, doing so incurs a cost in the form of a deadweight loss. In designing an optimal program, the regulator will seek to achieve a balance between these two effects in order to minimize abatement costs to achieve the pollution target on average.

3.2 *The case of a known delivery coefficient*

To isolate the role of uncertainty with regard to the abatement costs, we next examine the optimal trading ratio and permit quantity in the absence of uncertainty in the delivery coefficient. We have the following from (14):

(18) $$t^* = d + \frac{d\sigma_1^2 - \text{cov}(\theta_1, \theta_2)}{a^2 - \sigma_1^2} \text{ when } d \text{ is known.}$$

If θ_1 is known with certainty, i.e., $\sigma_1^2 = 0$ and $\text{cov}(\theta_1, \theta_2) = 0$, then the optimal trading ratio should equal the delivery coefficient. However, in general, even if the delivery coefficient is known, the optimal trading ratio in our model is not necessarily equal to the delivery coefficient. We explore the intuitive rationale behind this result in the rest of this section.

3.2.1 *Differences between pollution target, total permits, and actual pollution*

When the delivery coefficient is a known constant, we know from (1) and (2) that the gap between the total permits allocated and the pollution target is

(19) $$\bar{P}_{permit} - \bar{P}_{ante} = E[e_2](t - d).$$

Thus, if $t \neq d$, then the total permit quantity will also deviate from the *ex ante* target so that the *ex ante* pollution constraint will be met. Similarly, we can derive

(20) $$P_{actual} - \bar{P}_{permit} = -e_2(t - d).$$

That is, if $t > d$, then the actual pollution will be less than the permit allocated. This occurs because 1 unit of Firm 2's emissions contributes d units to total actual pollution, but 1 unit of

Firm 2's emissions requires t units of permits in the market constraint. Adding up the previous two equations, we have

(21) $$P_{actual} - \overline{P}_{ante} = (t - d)\left(E[e_2] - e_2\right) = (t - d)\frac{t\theta_1 - \theta_2}{2c}.$$

To derive the second equality in (21), equation (11) is used. For any given θ_2, the higher θ_1 is, the higher the actual pollution will be if $t > d$.

From (18), we know that the optimal trading ratio is greater than *d* if $\operatorname{cov}(\theta_1, \theta_2) = 0$. The intuition is as follows. As we described earlier, for any realization of θ_1 and θ_2, the emissions that would result in the least abatement cost for any actual *ex post* pollution level would satisfy (5); that is,

(22) $$(b + \theta_2) + 2c(e_2^0 - e_2) = d(a + \theta_1).$$

In other words, the marginal cost of controlling total pollution would be determined by θ_1 regardless of the allocation of emissions from the two firms—a higher θ_1 would imply a higher marginal abatement cost. For a regulator who is required to meet a pollution level in expectation (i.e., on average), therefore, it makes sense to design policies that require lower abatement (a higher pollution level) when the marginal abatement cost (θ_1) turns out to be high and vise versa. Setting a trading ratio that is greater than the delivery coefficient accomplishes this, since equation (21) implies that $\frac{\partial P_{actual}}{\partial \theta_1} = (t - d)\frac{t}{2c} > 0$ if $t - d > 0$.

3.2.2. The trade-off of the total pollution effect and the deadweight loss effect

With firms' emissions decisions, $e_i^*(d, \bar{P}_{ante})$, $e_i^*(d, \bar{P}_{actual})$ and $e_i^*(t, \bar{P}_{permit})$, we can obtain an expression for the two effects (assuming d is known):

$$\text{(23)} \qquad total\ pollution\ effect = \frac{1}{2c}(t-d)(a+\theta_1)(t\theta_1-\theta_2),$$

$$\text{(24)} \qquad deadweight\ loss\ effect = -\frac{1}{4c}(a+\theta_1)^2(t-d)^2 \ .$$

Equation (23) indicates that the magnitude of the total pollution effect depends on the parameters in the cost functions and how much the trading ratio differs from *d*. As is expected, (24) implies that the deadweight loss effect is never positive. For any given θ_1, the larger the difference between the trading ratio and the delivery coefficient, the larger the deadweight loss effect.

Figure 2 and Figure 3 illustrate the intuition and magnitude of the two effects. For simplicity, the delivery coefficient in the figures is set to one, which is assumed known by the regulator. In both figures, the total length of the horizontal axis represents the total permits available and the solid downward-sloping line is the marginal abatement cost curve of Firm 2 as emissions are increased (i.e., abatement is decreased) for the case in which $\theta_2 < 0$. In Figure 2, the marginal abatement cost curve of Firm 1 (for $\theta_1 = 0$, i.e., $MC_1 = a$) is represented by the horizontal line that intersects with Firm 2's marginal cost curve at B^0. When $t = d = 1$, $\bar{P}_{permit}^{t=1}$ is set equal to $\bar{P}_{ante}$ by (19). Since $MC_1 = MC_2$ at B^0, B^0 represents the permit market equilibrium, indicating the split of the emissions by the two firms with Firm 1's emissions reading from the right (O_1) and Firm 2's emissions reading from the left (O_2). As (5) is satisfied at B^0, the *ex post* abatement cost is minimized to reach a total pollution level of $\bar{P}_{permit}^{t=1}$ (i.e., with complete information on θ_1 and θ_2, B^0 represents the least-cost solution).

When the trading ratio is set greater than the known delivery coefficient several changes occur in Figure 2. First, the optimal total permit cap increases to $\overline{P}_{permit}^{t>1}$ by (19), which is reflected by the shifting out of the right boundary of Figure 2 from O_1 to O_1'. Second, the new permit market equilibrium is represented by point B', indicating a reduction in e_2. Third, we can no longer obtain e_1 from the right (O_1') to the equilibrium point (B'), since the permit market constraint now requires that the total permits be greater than or equal to the weighted sum of emissions (with the weight on e_2 equal to t), not to the simple sum of emissions from the two firms. To reflect the weighting, it would be necessary to adjust the MC curve as in the dotted downward-sloping curve to represent $t*e_2$ for every e_2 on MC_2. Then, Firm 1's emissions can be obtained by reading from the right (O_1') to B''.

The two effects of setting $t>d$ on the total abatement cost of meeting the *ex ante* pollution target are illustrated by the shaded areas in Figure 2. As for the deadweight loss effect, note that the marginal abatement cost curve is still the horizontal line a, not the horizontal line ta. However, firms make their decisions based on the latter, which leads to too few emissions (i.e., too much abatement) by Firm 2, resulting in deadweight loss as reflected by the shaded triangle. The area of the triangle is equal to (24). For the case illustrated in Figure 2 (with $\theta_1 = 0$ and $\theta_2 < 0$), we know from (21) that the actual total pollution is greater than the *ex ante* pollution target. The savings in abatement cost are represented by the area of the shaded rectangle.

An optimally designed permit market will try to achieve a balance between the total pollution effect and the deadweight loss effect. To show how the regulator can reduce total *ex ante* expected abatement costs by setting $t>d$, we use the illustration in Figure 3, which is the same as Figure 2 except that it illustrates a case in which θ_1 can take on two values ($+\hat{\theta}_1$>0 and $-\hat{\theta}_1$) with equal probability. For simplicity we assume $\text{cov}(\theta_1, \theta_2) = 0$. Consistent with (24), the figure shows that there is a deadweight loss regardless of whether marginal abatement cost is high or low. The larger (smaller) shaded triangle represents the higher (lower) distortion when the realization of Firm 1's marginal abatement cost is high, i.e., $\theta_1 = +\hat{\theta}_1$ (low, i.e., $\theta_1 = -\hat{\theta}_1$).

The total pollution difference between setting $t > d$ and $t = d$ is given by (21) and is represented by the width of the large shaded rectangle for $\theta_1 = +\hat{\theta}_1$ and by the width of the small shaded rectangle for $\theta_1 = -\hat{\theta}_1$. When marginal cost is high (i.e., $\theta_1 = +\hat{\theta}_1$), setting $t > d$ will result in a cost saving from less abatement (or higher than expected pollution level) which is represented by the area of the large shaded rectangle. Similarly, when marginal cost is low (i.e., $\theta_1 = -\hat{\theta}_1$), setting $t > d$ will result in an extra cost from more abatement which is represented by the area of the small shaded rectangle. When the difference between the cost savings and the extra cost is positive, and when the difference is greater than the deadweight loss (the sum of the two shaded triangles), the regulator reduces total abatement cost with $t > d$. As illustrated in Figure 3, the area of the larger rectangle is larger than the sum of the areas of the smaller rectangle and the two shaded triangles, resulting in a welfare gain from setting $t > d$.

3.2.3 Effects of the covariance structure on the optimal permit trading program

As noted earlier, if $\sigma_i^2 = 0$ and $\text{cov}(\theta_1, \theta_2) = 0$, then the benchmark program will also be the optimal program. However, as long as $\sigma_1^2 > 0$, in general (14) implies that $t^* \neq d$ even when d is known for certain. In this section, we examine the effects of the covariance structure. Since an *ex ante* design minimizes expected abatement costs, we begin by taking the expectation of the total pollution effect and the deadweight loss effect (when d is known),

(25) $$E[total\ pollution\ effect] = \frac{(t-d)}{2c}[t\sigma_1^2 - \text{cov}(\theta_1, \theta_2)] \text{, and}$$

(26) $$E[deadweight\ loss\ effect] = -\frac{(t-d)^2}{4c}(a^2 + \sigma_1^2)\,.$$

These two expected effects will help us understand how the covariance structure will affect the optimal trading ratio, which will then determine the optimal permit cap through (19).

First note that if $\text{cov}(\theta_1, \theta_2) \le 0$, then clearly $t^* \ge d$ from (18). However, it is possible to have $t^* < d$, if $\text{cov}(\theta_1, \theta_2)$ is positive and large enough. The intuition is as follows. Equation (21) implies that the actual pollution level, compared to the *ex ante* target, depends on $(t-d)(t\theta_1 - \theta_2)$. If θ_1 is high when θ_2 tends to be low, then P_{actual} is high if $t - d > 0$. Since it is desirable to have a high pollution level when marginal cost is high (i.e., θ_1 is high), it is optimal for $t^* - d > 0$. On the other hand, if θ_1 is high when θ_2 tends to be high, and if θ_2 tends to be so high that $(t\theta_1 - \theta_2) \le 0$, then P_{actual} will be relatively high only if $t - d \le 0$. In such a situation, $t^* - d \le 0$ is optimal. In the following, most of our discussion will focus on the case in which $t^* - d > 0$. The other case can be analyzed similarly.

The covariance and the trading ratio move in opposite directions since (18) implies that $\frac{\partial t^*}{\partial\, \text{cov}(\theta_1, \theta_2)} = \frac{-1}{a^2 - \sigma_1^2} < 0$. This is because the expected total pollution effect will be larger when the cost shocks move in opposite directions than when they move in the same direction, according to (25). On the other hand, the deadweight loss effect does not depend on $\text{cov}(\theta_1, \theta_2)$. Thus, a smaller positive or more negative correlation increases the trading ratio because a larger total pollution effect can overcome the effect of a larger deadweight loss effect resulting from a higher trading ratio.

Regarding the impacts of variance on the optimal trading ratio, from (18) we have $\frac{\partial t^*}{\partial \sigma_1^2} = \frac{t^*}{a^2 - \sigma_1^2} > 0$; that is, as the abatement cost of Firm 1 becomes more variable, the optimal trading ratio increases. Mathematically, (25) and (26) imply that, for $t > d$, as σ_1^2 increases, the total pollution effect increases faster than the deadweight loss effect. Intuitively, as discussed earlier, when θ_1 is high, there will be savings of abatement cost due to extra pollution and when θ_1 is low there will be extra costs due to a lower pollution level that has to be achieved. As θ_1 becomes more variable, P_{actual} also becomes more variable and the total pollution effect will be larger because the cost savings become larger while the extra costs become smaller. A larger total pollution effect can outweigh a higher deadweight loss and so the optimal trading ratio can be set higher.

3.3 *Case of an uncertain delivery coefficient*

The delivery coefficient is likely to be known for some pollutants (e.g., carbon dioxide), but there are many pollutants for which delivery coefficients will be uncertain. While uncertain delivery coefficients clearly characterize nonpoint source pollution, many point sources can also have uncertain delivery coefficients; for example, wind and weather uncertainty can affect air pollution deposition rates. Many water pollutants exemplify this notion well. The fate and transport of water pollutants is subject to both stochastic elements related to weather as well as scientific uncertainty concerning the physical diffusion process.[9] This is true for both point and nonpoint water pollution sources.

The impact of an uncertain delivery coefficient is reflected in (14) by $\text{cov}(d,\theta_1)$ and the use of the expected value of d. The optimal trading ratio moves in the same direction as $\text{cov}(d,\theta_1)$: $\frac{\partial t^*}{\partial \text{cov}(d,\theta_1)} = \frac{a}{a^2 - \sigma_1^2} > 0$. Suppose $\text{cov}(d,\theta_1) > 0$, that is, if the delivery coefficient is expected to be high, the marginal cost of abatement by Firm 1 is also expected to be high. For given emissions, a high d means more total pollution in the absence of any abatement. In order to reduce pollution to a fixed target, more abatement has to be undertaken. To ameliorate the pressure for more abatement, equation (21) implies that the trading ratio is increased and so more emissions will be allowed when the delivery coefficient is high and the abatement cost is also expected to be high. By the same logic, when the delivery coefficient is low and abatement cost also tends to be low (e.g., negative), equation (21) then implies that a higher trading ratio will restrict the amount of emissions that are allowed. However, the cost savings from extra pollution are higher than the increased cost from more abatement and so total abatement costs are reduced. As noted before, a larger total pollution effect can outweigh a larger deadweight loss and so the optimal trading ratio can be set higher.

The optimal permit allocation gap with an uncertain delivery coefficient is given by equation (15). Compared to the case with a known delivery coefficient as given in equation (19), there are two additional covariance terms, which represent the covariance between e_2 and d

[9] In the nonpoint source pollution literature, where one of the defining features of nonpoint source pollution is its inherent unobservability (Segerson, 1988), the focus has been on the trading in expected, as opposed to actual, emissions from a nonpoint source (e.g., Horan and Shortle, 2005). In this case, basically, another layer of uncertainty would be added to the design of the permit market: both firms and the regulator only know the distribution of emissions given any action taken by the firms. We can show that, like the uncertainty on firms' abatement costs and the delivery coefficient, this uncertainty will also be reflected in the optimal trading ratio and the optimal total number of permits.

(see the appendix). The terms indicate that if e_2 and d are positively correlated, then the optimal total permit cap should be even higher and vice versa. Thus, with an uncertain delivery coefficient, there is an additional reason that the optimal total permit cap might differ from the *ex ante* pollution target.

4. Conclusions

In this paper, we have investigated the optimal design of permit trading programs in a setup that incorporates three key features: (1) the regulator's objective is to minimize the expected abatement costs of meeting an *ex ante* pollution target (i.e., the pollution standard or target is represented as an expectation); (2) the regulator does not have complete information on firms' abatement costs; and (3) the delivery coefficient of emissions can be uncertain. It is well known that the regulator does not have to have any information on firms' abatement costs for a permit trading program in order to minimize the cost of achieving an *ex post* pollution target. However, we found that such information is useful in designing a trading program that meets an *ex ante* target at the lowest abatement costs.

In addition to the result that the optimal total permit cap is in general not equal to the *ex ante* pollution target, we found that the optimal trading ratio is not equal to the delivery coefficient even if the regulator has complete information on the delivery coefficient *ex ante*. The latter result arises from the dual roles that the trading ratio plays in a permit trading program. First, the trading ratio determines the substitution rate among emissions of different sources. Some studies have examined thoroughly the optimal trading ratio in situations in which the regulator, with complete information on firms' abatement costs, seeks to minimize the sum of abatement costs and damages from pollution (e.g., Kling and Rubin, 1997). Second, and equally importantly, the trading ratio affects the actual amount of pollution resulting from a trading program. This is because when the trading ratio is not equal to the delivery coefficient, the total permit cap is no longer the same as the total pollution that will result from a trading program. When designing a program, the regulator can use the trading ratio to induce the desirable pollution level.

Our findings indicate that it is important that the nature of a pollution target be clarified prior to the design of a trading program, given the stark difference between the optimal trading programs with an *ex ante* pollution target and the optimal trading programs with an *ex post* target. Under an *ex ante* target, not surprisingly, the actual pollution level as the result of implementing an optimal trading program would fluctuate around the target. In contrast, when an *ex post* target is used, the actual pollution level exactly equals the target by construction.

Appendix: Proof of equation (15)

Since the permit market constraint must hold for every level of firms' emissions, it also must hold for expected emissions levels, that is, $E[e_1^*] + tE[e_2^*] = \bar{P}^*_{permit}$. Taking the difference of this equation and the *ex ante* pollution constraint in (1), we obtain

(A1) $$\bar{P}^*_{permit} - \bar{P}_{ante} = E[e_2^*]t^* - E[de_2^*].$$

Note that $E[de_2^*] = E[d]E[e_2^*] + Cov(d, e_2^*)$, $E[d] = \mu$, and $Cov(d, e_2^*) = \frac{Cov(d, \theta_2) - t^* Cov(d, \theta_1)}{2c}$.

Rearranging equation (A1) with these relationships, we obtain (15).

References

Baumol W., Oates W. *The Theory of Environmental Policy*, Cambridge University Press, second edition: 1988.

Foster V., Hahn R. Designing More Efficient Markets: Lessons from Los Angeles Smog Control. *Journal of Law and Economics* 1995; 38(1);19-48.

Horan R.D., Shortle J.S. When Two Wrongs Make a Right: Second-Best Point-Nonpoint Trading Ratios. *American Journal of Agricultural Economics* 2005; 87(2); 340-352.

Kling C.L., Rubin J. Bankable Permits for the Control of Environmental Pollution. *Journal of Public Economics* 1997; 64; 101-115.

Lewis T.R. Protecting the Environment When Costs and Benefits are Privately Known. *RAND Journal of Economics* 1996; 27(4); 819-847.

Malik A., Letson D., Crutchfield S. Point/Nonpoint Source Trading of Pollution Abatement: Choosing the Right Trading Ratio. *American Journal of Agricultural Economics* 1993; 75(4); 959-967.

Montero J-P. Multipollutant Markets. *RAND Journal of Economics* 2001; 32(4); 762-774.

Montero J-P. Optimal Design of a Phase-in Emissions Trading Program. *Journal of Public Economics* 2000; 75; 273-291.

Montgomery W.D. Markets in Licenses and Efficient Pollution Control Programs. *Journal of Economic Theory* 1972; 5; 395-418.

Roberts M.J. Spence AM. Effluent Charges and Licences Under Uncertainty. *Journal of Public Economics* 1976; 5; 193-208.

Segerson K. Uncertainty and Incentives for Nonpoint Pollution Control. *Journal of Environmental Economics and Management* 1988; 15; 87-98.

U.S. Environmental Protection Agency (USEPA). National Ambient Air Quality Standards (NAAQS). 2006a. http://www.epa.gov/air/criteria.html.

U.S. Environmental Protection Agency (USEPA). Fact Sheet—Proposed Reissuance of National Pollutant Discharge Elimination System (NPDES) Stormwater Multi-Sector General Permit for Industrial Activities. 2006b. http://www.epa.gov/npdes/pubs/msgp2006_factsheet-proposed.pdf

Figure 1. Decision process and sequence of events in emissions trading

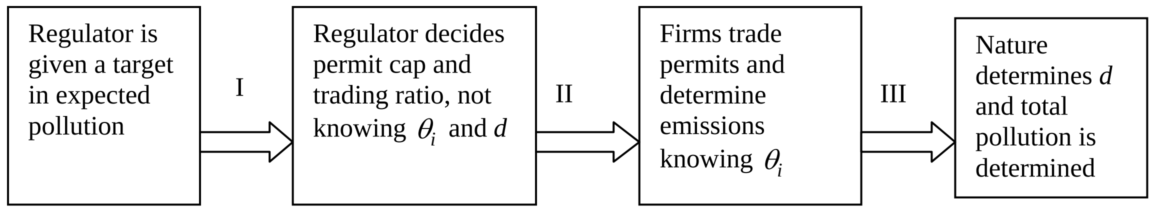

Figure 2. The effects of setting $t^* > d = 1$ under the *ex ante* pollution constraint $e_1 + de_2 = \overline{P}_{ante}$ and the permit market constraint $e_1 + te_2 = \overline{P}_{permit}$ (for $\theta_1 = 0,\ \theta_2 < 0$).

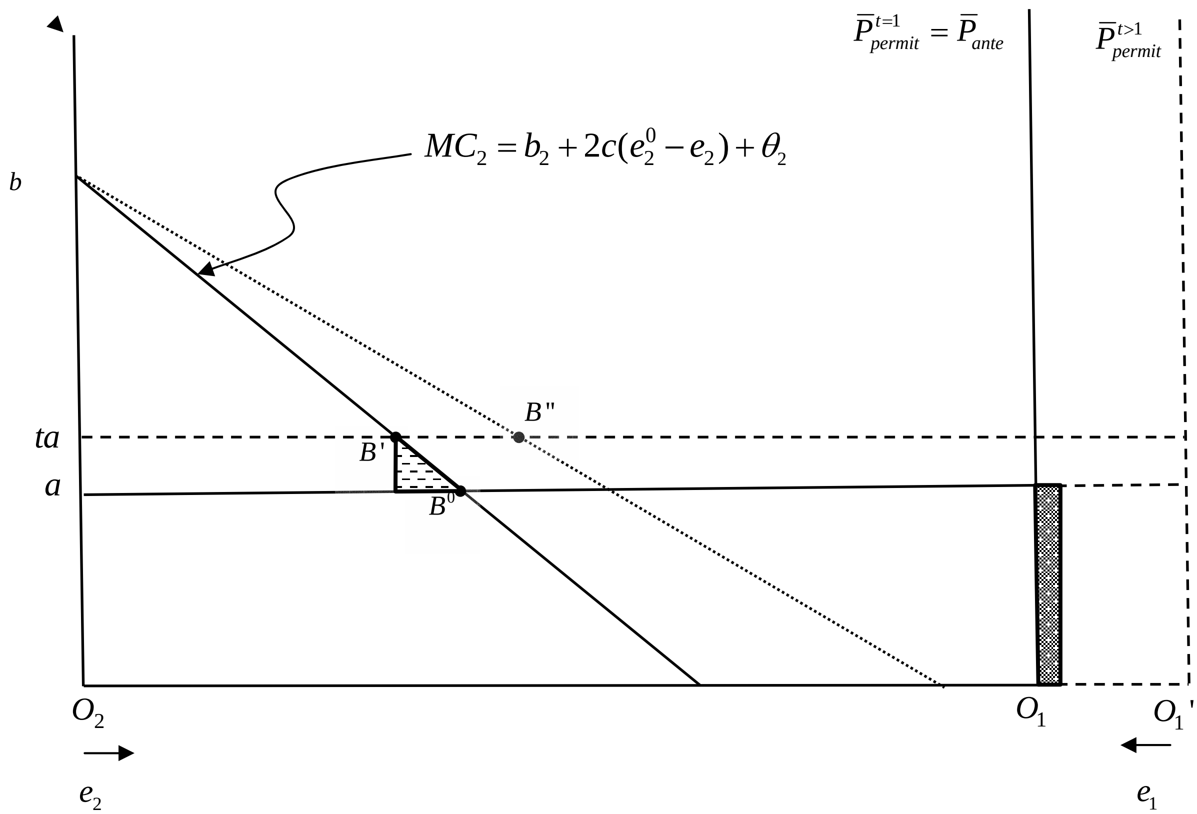

Figure 3. A comparison of the welfare effects when θ_1 is high versus when θ_1 is low for a given value of θ_2 ($\theta_2 = 0$). (In the figure, θ_1 is assumed to take two values, $+\hat{\theta}_1$>0 and $-\hat{\theta}_1$, with equal probability.)