2005

# Enforcing emissions trading when emissions permits are bankable

JK Stranlund

C Costello

CA Chavez

## Recommended Citation

Stranlund, JK; Costello, C; and Chavez, CA, "Enforcing emissions trading when emissions permits are bankable" (2005). *Journal of Regulatory Economics*. 187.
Retrieved from https://scholarworks.umass.edu/resec_faculty_pubs/187

# Enforcing Emissions Trading when Emissions Permits are Bankable*

JOHN K. STRANLUND
University of Massachusetts-Amherst
*Department of Resource Economics*
*214 Stockbridge Hall Amherst, MA 01003, USA*
*E-mail:stranlund@resecon.umass.edu*

CHRISTOPHER COSTELLO
University of California
*Donald Bren School of Environmental Science and Management*
*Donald Bren Hall 4410*
*Santa Barbara, CA 93106-5131, USA*
*E-mail: costello@bren.ucsb.edu*

CARLOS A. CHÁVEZ
Universidad de Concepción
*Departamento de Economía*
*Casilla 1987, Concepción, Chile*
*E-mail: cchavez@udec.cl*

**Abstract**

We propose enforcement strategies for emissions trading programs with bankable emissions permits that guarantee complete compliance with minimal enforcement costs. Our strategies emphasize imperfect monitoring supported by a high unit penalty for reporting violations, and tying this penalty directly to equilibrium permit prices. This approach is quite different from several existing enforcement strategies that emphasize high unit penalties for emissions in excess of permit holdings. Our analysis suggests that a high penalty for excess emissions cannot be used to conserve monitoring effort, and that it may actually increase the amount of monitoring necessary to maintain compliance.

**Key words:** compliance, enforcement, emissions trading, permit banking

**JEL Classification:** L51, Q28

---

[*] The authors are grateful to Anthony Heyes and three anonymous referees for their valuable comments. Partial support for this research was provided by the Cooperative State Research Extension, Education Service, U. S. Department of Agriculture, Massachusetts Agricultural Experiment Station under Project No. MAS00871.

## 1. Introduction

Whether pollution sources should be allowed to bank or borrow transferable emissions permits, and what restrictions should be placed on this activity are fundamental design choices for market-based pollution control. We examine theoretically the compliance incentives and enforcement of an emissions trading program when emissions permits can be banked for future use or borrowed against future permit holdings. There are several emissions trading programs that allow some form of permit banking and borrowing; the most well known is the Sulfur Dioxide ($SO_2$) Allowance Trading program, which allows unrestricted banking of permits, but not borrowing. Pollution sources have made good use of this feature of the $SO_2$ program, banking just over 30% of the total allocation of $SO_2$ allowances during Phase I of the program (1995–1999).[1] At the same time, all sources in the $SO_2$ program were perfectly compliant (US EPA 2000). There have been a few minor incidences of noncompliance since then, but by the measure of maintaining compliance the enforcement apparatus of the $SO_2$ program has been very successful (US EPA 2003a).

Policy analysts usually point to two features of the $SO_2$ program to explain its success in inducing and maintaining near-perfect compliance (e.g., Swift 2001; Stranlund et al. 2002; US EPA 2003b). First, should a source fail to hold sufficient allowances to cover its emissions in a compliance period it is automatically assessed a financial penalty, the unit value of which has always been many times higher than going allowance prices. Second, all facilities in the program are required to install a continuous emissions monitoring system which generates and submits quarterly emissions reports to the EPA. These systems are fully automated, thereby minimizing the opportunities for submitting false emissions data. Two additional program features may be partly responsible for the high rate of compliance. In addition to facing a high financial penalty, a firm's excess emissions in one period are offset by a one-to-one deduction from its allocation of permits in the next period. Finally, misreporting of emissions is a separate violation in the $SO_2$ program that is distinct from the failure to hold sufficient permits, although there is not an explicit penalty for reporting violations.[2]

Although the $SO_2$ program has been successful in achieving near-perfect compliance, the relative contribution of each of these enforcement features to its success has not been examined. For that reason, it is not at all clear to what extent similar measures could effectively be applied in other pollution control situations; in particular, those situations in which perfect monitoring is not possible or simply too

---

[1] For an empirical examination of banking behavior in the $SO_2$ program see Ellerman and Montero (2002).

[2] See section 412(e) of the Title IV ($SO_2$ Allowance) regulations; http://www.epa.gov/airmarkt/arp/regs/index.html. Stranlund et al. (2002) provide a detailed account of the enforcement strategy of the $SO_2$ program and a comparison to the enforcement strategy of the Regional Clean Air Incentives Market (RECLAIM) program.

expensive. In this paper, we construct a dynamic model of compliance in emissions trading programs to examine the design of enforcement when emissions permits are bankable and when monitoring of emissions need not be perfect.[3] We focus on enforcement strategies that induce perfect compliance with minimal enforcement costs.[4] Like most analysts we assume that setting penalties is costless, but monitoring is not. Since we focus on inducing perfect compliance, no costs of collecting penalties are incurred: therefore, minimizing enforcement costs means minimizing monitoring effort.

The main contribution of our analysis is to uncover the importance of detecting and punishing under-reported emissions. It is clear that compliance in trading programs with bankable permits and imperfect monitoring requires self-reported emissions with a separate penalty for misreporting. This is distinct from most of the literature on self-reporting in law enforcement, which assumes that self-reporting is a voluntary activity that can be encouraged by offering a lower penalty for self-reported violations. In fact, in most of the models in this literature self-reporting is not necessary to achieve compliance.[5] However, for an emissions trading program with bankable permits and imperfect monitoring, self-reporting of emissions is required simply because, if a firm is not monitored in a particular period, its report is the only available information on its emissions to determine how many permits are used for current compliance and how many are carried into the next period. Moreover, the misreporting violation must be distinct from a permit violation, in which a firm fails to hold sufficient permits to cover its emissions. When emissions permits can be banked but not borrowed, an important type of noncompliance occurs when a firm that holds enough permits to cover its emissions in a period still under-reports its emissions to generate additional permits for its permit bank. Offering a lower penalty for truthful self-revelation of permit violations cannot deter this sort of noncompliance.

---

[3] Following Kling and Rubin (1997) we use the term "banking" to mean saving permits for future use and "borrowing" to mean borrowing permits from future permit holdings. "Bankable permits" refers to policies that allow banking only, or both banking and borrowing.

[4] Our focus on inducing perfect compliance is motivated by the high rates of compliance in well-known emissions trading programs like the $SO_2$ Allowance Trading, RECLAIM, and the $NO_X$ Budget Trading programs. It is clear that the enforcement strategies for these programs were designed to achieve high rates of compliance. Since achieving high compliance rates appears to be an important objective for policymakers, it is worth analyzing how to accomplish this objective with minimal enforcement costs. Moreover, given certainty about abatement costs, as we assume in our model, and the freedom to choose penalties and permit allocations without restrictions, designing an efficient emissions trading program that involves perfect compliance is always possible. This is not to say that analyzing imperfect compliance in dynamic emissions trading programs is unimportant; however, we leave that to future investigations.

[5] However, analysts have identified several reasons why self-reporting can conserve enforcement and compliance costs. See Malik (1993), Kaplow and Shavell (1994), Livernois and McKenna (1999), and Innes (1999, 2000, 2001). Each of these works models a lower penalty for reported violations. Like our model, Harford (1987), Stranlund and Chavez (2000), and Chavez and Stranlund (2003) make misreporting of emissions a separate violation with its own penalty.

Furthermore, our analysis suggests that emphasizing a high unit penalty for permit violations, as in the $SO_2$ program, is not warranted. When permit borrowing is not allowed, a permit violation penalty has only limited deterrence value; increasing this penalty does not reduce the amount of monitoring necessary to maintain compliance. In fact, we show that setting this penalty at a high level may actually increase the amount of monitoring necessary to maintain compliance, because it increases the incentive firms have to under-report their emissions. In any period but the final period of a program without permit borrowing, the permit violation penalty should be set to just make up the difference between the current period permit price and the present value of next period's permit price. Of course, when both permit banking and borrowing are allowed, there should be no permit violation penalty in periods before the last, because emissions in excess of permit holdings do not constitute a violation in these periods. In contrast, whether or not permit borrowing is allowed, a penalty for under-reported emissions allows regulators to maintain compliance with imperfect monitoring, and setting this penalty as high as is practicable conserves monitoring costs. Our analysis suggests that it is possible to maintain compliance in a dynamic emissions trading program with imperfect monitoring, but doing so requires focusing on punishing reporting violations rather than on punishing permit violations.

The $SO_2$ program is not the only program with bankable permits that imposes a high penalty for permit violations. The EPA's $NO_X$ Budget Trading Program (US EPA 2004) and its recently proposed Clear Skies Initiative (US EPA 2003c) are similar to the $SO_2$ program in their banking provisions and most elements of their enforcement provisions. However, the permit violation penalty in the $NO_x$ program is a three-to-one offset from future permit allocations, while the Clear Skies permit violation penalty is one to three times the clearing price in the last auction of permits. We examine both of these penalty schemes and conclude that, from the perspective of minimizing enforcement costs, these penalties are also probably excessive.

Several authors have demonstrated that when emissions permits trade across time on a one-to-one basis and firms actually bank or borrow permits, the price of permits increases through time at the rate of discount (Cronshaw and Kruse 1996; Rubin 1996; Kling and Rubin 1997). From the compliance and enforcement perspective, this implies that firms' incentives toward noncompliance also increase, which requires monitoring effort to increase through time if the reporting penalty is fixed. Tying the reporting penalty directly to current permit prices eliminates the need to incur increasing enforcement costs over time. In fact, a constant level of monitoring can maintain compliance with minimal discounted enforcement costs over the life of a program if penalties are structured so that: (1) the reporting violation penalty is a constant multiple of the current permit price with a multiplication factor that is as high as is practicable, and (2) the permit violation penalty is set to just make up the difference between the current price and the present value of next period's price when borrowing is not allowed, and that is set to zero when borrowing is allowed.

This paper contributes to and draws upon three lines of theoretical research. The first is the literature on self-reporting in law enforcement that we have already mentioned. The second is the theoretical literature on intertemporal trading of emissions permits, which has focused on banking behavior, the welfare consequences of allowing banking, and the optimal design of banking rules (Cronshaw and Kruse 1996; Rubin 1996; Kling and Rubin 1997; Schennach 2000; Yates and Cronshaw 2001; Leiby and Rubin 2001; Phaneuf and Requate 2002). Finally the literature on enforcing emissions trading programs has provided insights into the determinants of compliance and the design of enforcement strategies for these programs (Malik 1990, 1992; Keeler 1991; van Egteren and Weber 1996; Stranlund and Dhanda 1999; Stranlund and Chavez 2000). With one exception, however, the models of permit banking assume away the fact that emissions trading programs must be enforced, and the models of compliance and enforcement are static models that do not allow permit banking.

Innes (2003) is the only other analysis that we are aware of that considers enforcement with intertemporal permit trading. He argues that giving sources the ability to bank and borrow permits eliminates the need to impose costly sanctions to maintain compliance when emissions are stochastic. However, Innes simplifies his analysis by assuming that an enforcement strategy only consists of a costly penalty for permit violations. Because he admittedly ignores the reporting and monitoring functions of enforcement altogether, he does not address the issues that are the main focus of this paper—the importance of self-reported emissions under imperfect monitoring and the separate roles played by penalties for permit violations and penalties for reporting violations.

The rest of the paper proceeds as follows. In the next section we develop a dynamic model of compliance when permits are bankable. In section 3, which is the heart of the paper, we derive enforcement strategies for maintaining compliance over the life of an emissions trading program with minimal monitoring effort. In this section we make our main case that a dynamic enforcement strategy for bankable emissions permits should focus on inducing truthful reporting of emissions, rather than on punishing permit violations. In section 4 we draw further conclusions about how enforcement strategies should evolve over time to minimize the present value of enforcement costs. In section 5 we modify our model to incorporate alternative trading restrictions on intertemporal permit trading, and alternative offset penalties for permit violations. In addition we discuss how enforcing a dynamic emissions trading program differs from enforcing a static program. Section 6 concludes.

## 2. A Model of Compliance in a Dynamic Emissions Trading Program

The analysis of this paper is based largely on a model of compliance by a risk-neutral firm in a dynamic emissions trading program that lasts $T$ periods. Let $x_t$ be the number of emissions permits the firm holds at the beginning of period $t$. Each permit confers the legal right to release one unit of emissions. During period

$t$ the firm chooses how many permits $l_t$ to purchase ($l_t > 0$) or sell ($l_t < 0$). Permits trade in period $t$ at a competitive price $p_t$. A system is in place to track emissions permits so that at any point in time the regulator has perfect information about the number of permits held by each firm. During period $t$ the firm also chooses its emissions $e_t$. The firm has an abatement cost function, $c(e_t)$, which is strictly decreasing and convex and does not vary over the life of the program.

The firm is required to submit a report, $r_t$, of its emissions in $t$. At this point it is important to distinguish between the two types of violations the firm can commit. A *reporting violation* occurs in period $t$ if the firm under-reports it emissions; that is, $e_t > r_t$. A *permit violation* occurs when the firm holds insufficient permits to cover its emissions; that is, $e_t > (x_t + l_t)$. Under a trading program in which borrowing against future permit allocations is not allowed, a permit violation can occur in any period. However, when borrowing is allowed, emissions in excess of permit holdings (in all periods but the last) do not constitute a violation. Excess emissions in a period are simply paid back in a later period. In the last period no permit borrowing is possible so excess emissions in this period constitute a permit violation.

Implicit in the emissions report is a report of the firm's compliance status and whether it is banking or borrowing permits. If a firm's reported emissions exceed its permit holdings, $r_t > (x_t + l_t)$, then the firm is reporting that it is borrowing permits (when allowed), and is reporting a permit violation if borrowing is not allowed. If $r_t < (x_t + l_t)$, then the firm is reporting that it is banking permits. When permit borrowing is not allowed, $r_t \leq (x_t + l_t)$ indicates that the firm is reporting that it is permit compliant. Of course, we must distinguish reported permit violations, permit compliance, and permit banking and borrowing from their actual values. If actual emissions exceed permit holdings, $e_t > (x_t + l_t)$, then there is an actual permit violation when borrowing is not allowed, or the firm is borrowing permits when this is allowed. If $e_t < (x_t + l_t)$ the firm has excess permits to bank, and if $e_t \leq (x_t + l_t)$, then the firm is permit compliant when borrowing is not allowed.

Monitoring for compliance by authorities is potentially imperfect in the sense that the probability that the authority is able to make a determination of a firm's compliance status is $\pi_t \in [0, 1]$. We assume that the costs of monitoring are increasing in this probability.

In this paper we consider enforcement strategies that are structured to maintain compliance without perfect emissions monitoring. It is worthwhile then to be clear about the sorts of monitoring and reporting technologies that allow this, and how they are different from the perfect (or near-perfect) strategies that employ the continuous emissions monitoring technologies that are featured in several programs, including the $SO_2$ program. Without continuous emissions monitoring devices it is unlikely that an emissions trading program can be based on firms' actual emissions. Rather, a trading program would be based on estimates of emissions that are determined by combining such information as periodic measurements of emissions flow, types and performance of abatement equipment, as well as production data. A firm's report then is a report of the variables that are used to calculate estimated emissions. Monitoring by authorities produces an independent estimate of emissions with site visits to measure emissions flow and to inspect abatement equipment, and other checks to assess the accuracy of the firm's reported information. If the monitor finds a discrepancy with the firm's report, and consequently a different estimate of emissions, the firm is violating its reporting responsibilities, and perhaps its emissions permits as well. Since this monitoring is costly, authorities are likely to monitor only a subset of firms.[6]

While our setup is stylized, it conforms closely to several real-world policies. For example, in the Emissions Compensation Program of Santiago, Chile, firms report daily emissions based on a once a year measurement of emissions flow parameters. Based on a source's reports, regulators approximate its "actual" emissions by the maximum amount of emissions that it could potentially emit in a given year. Regulators monitor a limited number of firms per year to perform their own measurement of the flow parameters. A violation occurs if the reported flow parameters are different from the measured emissions obtained in the inspection.[7] A second example concerns major stationary sources in the United States that operate under Title V of the Clean Air Act Amendments of 1990. These sources are required to report their aggregate annual emissions to their state air quality authorities and pay fees based, at least in part, on this report of total emissions. Usually, total emissions are estimated with periodic emissions flow tests, information on raw materials use, production schedules, and the performance of abatement equipment, although there appears to be significant variation in the reporting required by the states (US GAO 2001). Agency monitoring could, in principle, involve generating an independent estimate of emissions with the agency's own source tests and other checks to assess the accuracy of a firm's report. If this strategy can be used to support the collection of emissions fees, it can be used to support an emissions trading program.

Returning to our model development, under a system of permit trading with no permit borrowing, permit violations in period $t$ (whether they are revealed in a firm's emissions report or discovered by the authorities) are penalized at $\phi_t$ per unit. This penalty corresponds to the permit violation penalty in the $SO_2$ program. When borrowing is allowed, emissions in excess of permit holdings in periods $t < T$ are not violations, so $\phi_t = 0$ for $t = 0, \ldots, T-1$. In the last period excess emissions do represent a permit violation so $\phi_T > 0$. Whether or not borrowing is allowed, reporting violations that are discovered through an audit are penalized at $\gamma_t$ per unit. Both $\phi_t$ and $\gamma_t$ are scalars known by all parties.[8] As mentioned

---

[6] We henceforth assume, like most other analysts, that monitoring produces a measure of emissions that is accurate enough to determine a firm's compliance status without errors.

[7] For detailed accounts of the enforcement strategy and compliance results of the Emissions Compensation Program see Palacios and Chavez (2005, 2002).

[8] Our assumption of linear penalties (i.e., constant marginal penalties) is not common in the literature on compliance and enforcement of emissions trading programs. However, linear penalties are much more common for actual and proposed emissions trading programs than nonlinear penalties (see Boemare and Quirion 2002 for examples). Our use of linear penalties is motivated, in part, by this fact.

in the introduction, there is no penalty like $\gamma_t$ in the $SO_2$ regulations, although there appears to be an avenue for punishing misreporting.

We can now define the expected penalty under a program without permit borrowing for a firm that is violating its permits and under-reporting its emissions:

$$\begin{aligned} f(e_t, l_t, r_t) &= \phi_t(r_t - (x_t + l_t)) + \pi_t \{\gamma_t(e_t - r_t) + \phi_t[(e_t - (x_t + l_t)) - (r_t - (x_t + l_t))]\} \\ &= \phi_t(r_t - (x_t + l_t)) + \pi_t(\gamma_t + \phi_t)(e_t - r_t). \end{aligned} \tag{1}$$

To understand how $f(e_t, l_t, r_t)$ is constructed, note that a firm that reports a part of its permit violation faces an automatic penalty of $\phi_t(r_t - (x_t + l_t))$. If the firm is audited so that its reporting violation is discovered (this occurs with probability $\pi_t$), the penalty for this violation, $\gamma_t(e_t - r_t)$, is assessed. Of course, if a firm does not hold enough permits to cover its emissions and also under-reports its emissions, it has not reported the full extent of its permit violation. If this is discovered the firm is liable for the incremental penalty for its unreported permit violation, $\phi_t(e_t - (x_t + l_t)) - \phi_t(r_t - (x_t + l_t) = \phi_t(e_t - r_t)$. When borrowing is allowed, excess emissions in periods $t < T$ are not violations, so we modify (1) by simply setting $\phi_t = 0$, for $t = 0, \ldots, T-1$. In the last period excess emissions do represent a permit violation so the expected penalty in $t = T$ is given by (1).

Combining the elements defined thus far yields the firm's single-period expected costs,

$$v(e_t, l_t, r_t, x_t) = c(e_t) + p_t l_t + f(e_t, l_t, r_t). \tag{2}$$

We now turn to characterizing the evolution of the firm's stock of permits. Assume that under a program without permit borrowing that a firm's reporting or permit violation in a period is offset with a one-to-one deduction from its allocation of permits in the next period. This is a characteristic of the $SO_2$ and other programs. The offset implies that the number of permits held by the firm at the beginning of period $t+1$, $x_{t+1}$, includes the firm's predetermined endowment of permits that period, $\bar{l}_{t+1}$, plus permits saved from the previous period, or less permits deducted because of a violation discovered or reported in the previous period. When both banking and borrowing are allowed, emissions in excess of permit holdings in a period do not represent a violation, but are nevertheless deducted from the firm's permit endowment in the next period. For now we assume that the intertemporal trading ratio for permits is one-to-one, but will briefly analyze the impact of alternative trading ratios on enforcement in section 5. Assuming that permit violations are offset on a one-for-one basis when borrowing is not allowed and that the intertemporal trading ratio for permits is one-to-one when borrowing is allowed implies that a firm's stock of permits evolves in the same way whether borrowing is allowed or not.[9]

---

[9] In a sense, permit and reporting violations when borrowing is not allowed are simply a stochastic form of borrowing against future allocations with a potential price (penalty) for the privilege of doing so.

From the perspective of period $t$ choices, $x_{t+1}$ is potentially a random variable because of incomplete monitoring and the possibility of under-reporting of emissions in $t$. If an audit is conducted in $t$ a firm's actual permit shortfall, $e_t - (x_t + l_t) > 0$, or bank, $e_t - (x_t + l_t) < 0$, is carried into the next period. If an audit is not conducted the firm's reported permit shortfall, $r_t - (x_t + l_t) > 0$, or bank, $r_t - (x_t + l_t) < 0$, is carried forward. Thus, from the perspective of period $t$ choices, the expected number of permits with which the firm will begin period $t+1$ is

$$E_t(x_{t+1}) = \bar{l}_{t+1} + \pi_t(x_t + l_t - e_t) + (1 - \pi_t)(x_t + l_t - r_t), \tag{3}$$

where the subscript on the expectation operator indicates that the expectation is from the perspective of period $t$.

We can now specify a firm's decision problem in its entirety. Its objective is to choose a time path of emissions, permit transactions, and emissions reports to minimize its discounted sum of expected costs, subject to (3), and nonnegativity constraints for emissions, reported emissions, and permit holdings in every time period. In the final period the firm will never find it advantageous to hold excess permits or report that it holds excess permits, because excess permits at the end of $T$ have no value. Therefore, we will impose the constraint that $e_T - (x_t + l_T) \geq 0$. Formally, the firm's problem is to choose $\{e_t, l_t, r_t\}$, $t = 0, \ldots, T$, to solve

$$\begin{aligned}
&\min E\left[\sum_{t=0}^{T} \beta^t \left(c(e_t) + p_t l_t + f(e_t, l_t, r_t)\right)\right] \\
&s.t.\ x_{t+1} = \begin{cases} \bar{l}_{t+1} + x_t + l_t - e_t & \text{with probability} \quad \pi_t \\ \bar{l}_{t+1} + x_t + l_t - r_t & \text{with probability} \quad 1 - \pi_t, \end{cases} \qquad t = 0, 1, \ldots, T-1, \\
&\qquad e_t \geq 0, \quad r_t \geq 0, \quad x_t + l_t \geq 0, \quad t = 0, 1, \ldots, T, \\
&\qquad e_T - (x_T + l_T) \geq 0, \\
&\qquad x_0 = \bar{l}_0.
\end{aligned} \tag{4}$$

In the objective function $\beta$ is the discount factor, which is assumed to be constant over the life of the program.

The objective function and constraints specified in (4) define a discrete-time stochastic dynamic programming problem, which will be solved by developing and analyzing the dynamic programming equation. The uncertainty in our problem stems from incomplete monitoring: there are no other stochastic elements in the problem. In particular, we assume that each firm can accurately forecast equilibrium permit prices over the life of the policy. Define $J_t(x_t)$ as minimum expected discounted costs from period $t$ on through the last period, given that the firm has $x_t$ permits at the beginning of $t$. The stochastic dynamic programming equation is then

$$J_t(x_t) = \min_{e_t, l_t, r_t} c(e_t) + p_t l_t + f(e_t, l_t, r_t) + \beta E_t[J_{t+1}(x_{t+1})], \tag{5}$$

subject to the constraints specified in (4). The dynamic programming equation has the usual interpretation: it balances the effects of decisions $e_t$, $l_t$, and $r_t$ on period

$t$ expected costs against the effects of these decisions through the state equation (3) on minimum discounted expected costs from $t+1$ on through the last period.

The regulator's objective is to induce full compliance in every period with minimal enforcement effort. Therefore, from the firm's period $t$ choices we will derive an enforcement strategy to motivate it to be compliant in this period, given full compliance in all future periods. By induction, we will derive an enforcement strategy that will induce full compliance in every period. This approach is equivalent to deriving the enforcement strategy by backward induction. Using this approach we can specify $E_t[J_{t+1}(x_{t+1})]$ up to a constant. The following lemma is proved in Appendix A.

**Lemma 1:** *Given full compliance in periods* $j=t+1,\dots,T$,

$$E_t[J_{t+1}(x_{t+1})] = -p_{t+1}[\pi_t(x_t+l_t-e_t)+(1-\pi_t)(x_t+l_t-r_t)]+\tilde{C}_{t+1}, \tag{6}$$

*for some constant* $\tilde{C}_{t+1}$.

Equation (6) reveals that from the perspective of possibly noncompliant choices in period $t$, $\beta E_t[J_{t+1}(x_{t+1})]$, is the discounted expected value of the number of permits the firm will carry into period $t+1$ plus a constant term. Since $\beta E_t[J_{t+1}(x_{t+1})]$ is minimum discounted expected costs from $t+1$ on through the last period, we interpret the marginal effects of period $t$ choices on $\beta E_t[J_{t+1}(x_{t+1})]$ as the marginal future costs of beginning $t+1$ with fewer permits. A firm's compliance choices in a period will depend in large measure on these marginal future costs. Likewise, these costs will be important determinants of the enforcement strategy we propose.

Combining (5) and (6), the value function in any period $t=0,\dots,T-1$ is

$$\begin{aligned} J_t(x_t) = {} & \min_{e_t,l_t,r_t} c(e_t)+p_t l_t + f(e_t,l_t,r_t) - \beta p_{t+1}[\pi_t(x_t+l_t-e_t)+(1-\pi_t)(x_t+l_t-r_t)]+\beta\tilde{C}_{t+1} \\ & s.t.\ e_t \geq 0,\ \ r_t \geq 0,\ \ x_t+l_t \geq 0. \end{aligned} \tag{7}$$

In the last period, $J_T(x_T) = \min\limits_{e_T,l_T,r_T} c(e_T)+p_T l_T + f(e_T,l_T,r_T)$, subject to $e_T \geq 0$, $r_T \geq 0, x_t + l_T \geq 0$, and $e_T - (x_t + l_T) \geq 0$.

## 3. Enforcing Emissions Trading with Bankable Permits

Having derived the value function for each period $t = 0,\dots,T$ puts us in a position to derive enforcement strategies that achieve perfect compliance for each period of an emissions trading program with minimal enforcement costs. For the derivation of the appropriate enforcement strategy, it turns out that we do not need to specify the optimal choice of emissions, other than to assume throughout that $e_t > 0$. We need only focus on the emissions report, $r_t$, and permit transactions, $l_t$, given positive emissions in $t$.[10]

Consider the objective function of (7). Let emissions, permit transactions and the emissions report in period $t$ vary from their optimal values and use (1) to substitute for $f(e_t, l_t, r_t)$. We can then rewrite the objective function of (7) as

$$\begin{aligned} J(x_t; e_t, l_t, r_t) = c(e_t) + p_t l_t + \phi_t(r_t - (x_t + l_t)) + \pi_t(\gamma_t + \phi_t)(e_t - r_t) \\ -\beta p_{t+1}[\pi_t(x_t + l_t - e_t) + (1 - \pi_t)(x_t + l_t - r_t)] + \beta\tilde{C}_{t+1}, \quad (8) \end{aligned}$$

where recall that when both banking and borrowing of permits is allowed, $\phi_t = 0$ for $t = 0, \dots, T-1$. We have dropped the time subscript on $J$ to simplify the notation; from now on a subscript on $J$ denotes a first derivative. We have also written the value function explicitly as a function of the variables $e_t, l_t$, and $r_t$. Because of the linear penalties, (8) is linear in a firm's emissions report and its permit transactions. This is convenient because it allows us a more intuitive derivation of the appropriate enforcement strategies by focusing solely on $J$, rather than deriving enforcement strategies from the first-order conditions for (7).

### 3.1. The Evolution of Permit Prices

With bankable emissions permits, any enforcement strategy must account for the evolution of permit prices over time. Therefore, to guide our development of enforcement strategies we need some fundamental results about the evolution of permit prices.

Suppose at first that permit borrowing is not allowed. Previous work has shown that intertemporal equilibrium in a permit market under certainty requires that real permit prices be nonincreasing across time periods, and that firms will bank permits only when real permit prices are expected to remain constant (Cronshaw and Kruse 1996; Rubin 1996). This is true when you allow for noncompliance in a period as well.[11] To see why, note that if a firm is in permit violation in $t$, the derivative of (8) with respect to $l_t$ is $J_l = p_t - \beta p_{t+1} - \phi_t$. If the firm is permit compliant the firm does not face the penalty for permit violations, so $J_l = p_t - \beta\ p_{t+1}$. Whether a firm is permit compliant or not, $p_t < \beta p_{t+1}$ implies $J_l < 0$ for every firm. This inequality implies that all firms would demand an unbounded number of permits in $t$. Since this cannot be true in equilibrium, we must have $p_t \geq \beta p_{t+1}$. Now suppose that $p_t > \beta p_{t+1}$, but that a firm is banking permits in $t$. Since the firm is banking permits it is permit compliant and $J_l = p_t - \beta p_{t+1} > 0$. However, the inequality implies that the firm could reduce it expected compliance costs from

---

[10] Given the achievement of perfect compliance and permit market equilibrium in every period, it is easy to demonstrate the standard result that a firm's emissions choice in any period $t$ is determined by $c'(e_t) + p_t = 0$. The proof of Lemma 1 in Appendix A derives and uses this result.

[11] It is important to recall that all period-$t$ results are derived under the assumption of full compliance in all future periods. The characteristics of permit price paths that we present are not guaranteed to hold if there is noncompliance in future periods.

$t$ on by reducing its permit holdings so that it is not holding excess permits. The contradiction implies that firms will bank permits in period $t$ only if $p_t = \beta p_{t+1}$.

Now suppose that both permit borrowing and banking are allowed. Then intertemporal permit market equilibrium requires that real permit prices are constant throughout (Kling and Rubin 1997). The reasoning is simple. If $p_t > \beta p_{t+1}$, then all firms will demand zero permits in $t$, and if $p_t < \beta p_{t+1}$, then all firms will demand an unbounded number of permits. Since neither of these cases are consistent with permit market equilibrium, we must have $p_t = \beta p_{t+1}$.

### 3.2. Enforcement when Permit Borrowing is not Allowed

We now turn to deriving enforcement strategies, and first consider the case in which permit borrowing is not allowed. Making the common assumption that if a firm is indifferent between compliance and violation that it chooses compliance, from (8), a firm truthfully reports its emissions in period $t$ if and only if enforcement is designed so that

$$J_r = \phi_t - \pi_t(\gamma_t + \phi_t) + (1 - \pi_t)\beta p_{t+1} \leq 0. \tag{9}$$

That is, reporting violations are deterred if and only if a firm's expected compliance costs from period $t$ on are not increasing in its emissions report in $t$.[12] Similarly, permit violations in period $t$ are deterred if and only if the enforcement strategy induces

$$J_l = p_t - \beta p_{t+1} - \phi_t \leq 0. \tag{10}$$

To interpret (9) rearrange it to obtain $\pi_t(\gamma_t + \phi_t) \geq (1 - \pi_t)\beta p_{t+1} + \phi_t$. There are two reasons a firm may choose to under-report its emissions. One is to cover up a permit violation while the other is to carry additional permits into the next period. On the left-hand side of the relation above, $\pi_t(\gamma_t + \phi_t)$, is the expected marginal penalty for a reporting violation and the undisclosed part of a permit violation. On the right-hand side term, $(1 - \pi_t)\beta p_{t+1} + \phi_t$, is the expected marginal benefit of under-reporting emissions: $(1 - \pi_t)\beta p_{t+1}$ is the expected discounted marginal benefit of carrying additional permits into the next period because emission are under-reported, while $\phi_t$ is the certain unit penalty for the part of the permit violation that the firm avoids by under-reporting its emissions.

The intuition of (10) is as follows. Suppose that a firm contemplates a permit violation in period $t$. We can interpret the current permit price as the marginal benefit of a current permit violation, because it is the unit cost of purchasing additional permits to come into compliance. On the other hand, the discounted permit

---

[12] A firm will never report that its emissions are higher than they really are. There is no benefit to the firm of over-reporting its emissions. However, if a firm is violating its permits, over-reporting increases the certain penalty for reported permit violations without changing the reporting penalty it faces, which would be zero anyway. Moreover, whether the firm is violating its permits or not, over-reporting of emissions reduces the number of permits it begins the next period with.

price in the next period is the discounted marginal increase in future expected costs from carrying fewer permits into period $t+1$. Thus, the discounted permit price in the next period is part of the marginal cost of a current permit violation. The other part is, of course, the unit penalty $\phi_t$.

Achieving perfect compliance with minimal enforcement costs in periods $t=0,\ldots,T-1$ requires satisfying (9) and (10) with minimal monitoring. From (9), this implies that monitoring in each period $t$ should be set so that $\pi_t = (\phi_t + \beta p_{t+1})\big/(\gamma_t + \phi_t + \beta p_{t+1})$. Note that the required probability of detection is monotonically increasing in the permit violation penalty, $\phi_t$. This is due to the fact that increasing the permit violation penalty increases the incentive firms have to under-report their emissions. To offset this increased incentive, monitoring must be increased. To conserve monitoring costs, therefore, $\phi_t$ should be set as low as possible while maintaining (10); that is, $\phi_t = p_t - \beta p_{t+1}$. Note also that the required probability of detection is decreasing in the penalty from a reporting violation. If no such penalty exists ($\gamma_t = 0$), the probability of detection must be set to one to ensure full compliance.

These conclusions hold in the last period as well. The objective function in the last period is $J_T(x_T,\ e_T,\ l_T,\ r_T) = c(e_T) + p_T l_T + \phi_T(r_T - (x_T + l_T)) + \pi_T(\gamma_T + \phi_T)(e_T - r_T)$. Truthful emission reporting is achieved if and only if $J_r = \phi_T - \pi_T(\gamma_T + \phi_T) \leq 0$, and permit compliance is achieved if and only if $J_l = p_T - \phi_T \leq 0$. Achieving $J_r \leq 0$ with minimal monitoring, given $\phi_T$, requires $\pi_T = \phi_T\big/(\gamma_T + \phi_T)$. As in all previous periods, however, $\pi_T$ is increasing in $\phi_T$, implying that $\phi_T$ should be set as low as possible; that is, $\phi_T = p_T$.

Thus achieving and maintaining compliance with minimal enforcement costs when permit borrowing is not allowed requires:

$$\pi_t = (\phi_t + \beta p_{t+1})\big/(\gamma_t + \phi_t + \beta p_{t+1}) \quad \text{for } t = 0, 1, \ldots, T-1, \tag{11}$$

$$\pi_T = \phi_T\big/(\gamma_T + \phi_T), \tag{12}$$

$$\phi_t = p_t - \beta p_{t+1} \quad \text{for } t = 0, 1, \ldots, T-1, \tag{13}$$

$$\phi_T = p_T. \tag{14}$$

This enforcement strategy along with the conditions for the evolution of prices we presented earlier have important implications for the design of enforcement in an dynamic emission trading system when permit borrowing is not allowed. For all time periods before the last, (13) indicates that the permit violation penalty should be set to exactly make up the difference between the current permit price and the discounted price in the next period. Therefore in periods in which firms are banking permits so that real permit prices are constant across periods, this penalty should be set to zero. In other words, *the permit violation penalty serves no deterrence role when real permit prices are constant*. A positive permit violation penalty is needed to deter permit violations in periods in which firms are not banking permits and real permit prices are falling, but again it should just cover the difference between the current price and the discounted price in the next period. In the last period the permit violation penalty should be set at the equilibrium permit price. At no point should the permit violation penalty exceed the current permit price.

These conclusions contrast sharply with the permit violation penalty in the $SO_2$ Allowance program, which has always been many times higher than prevailing permit prices. Our analysis suggests that there is no gain to setting such a high penalty. To deter permit violations, given truthful reporting, the penalty only needs to cover the difference between the current period price and the present value of next period's price. Setting the penalty higher than this increases enforcement costs when monitoring is imperfect because it increases the amount of monitoring necessary to induce truthful emissions reports. Therefore, it cannot be increased to conserve monitoring effort. In this sense the permit violation penalty has only limited deterrence value.

From the monitoring conditions (11) and (12), it is clear that the reporting violation penalty, $\gamma_t$, plays a crucial role when monitoring is imperfect. In fact, compliance can be maintained with imperfect monitoring if and only if the reporting violation penalty is positive. Moreover, the monitoring requirement is monotonically decreasing in this penalty, suggesting that setting it as high as is practicable will conserve monitoring costs.

### 3.3. Enforcement when Both Permit Banking and Borrowing are Allowed

Allowing permit borrowing changes the required enforcement strategy in a simple way. Since emissions in excess of permit holdings do not constitute a violation in each period but the last, we modify (8) by setting the permit violation penalty, $\phi_t$, equal to zero in each period but the last, and focus on inducing truthful emissions reporting in these periods. Thus when borrowing is allowed a firm truthfully reports its emissions if and only if $J_r = -\pi_t\gamma_t + (1-\pi_t)\beta p_{t+1} \leq 0$. Minimal monitoring to induce truthful reporting requires $\pi_t = \beta p_{t+1} \big/ (\gamma_t + \beta p_{t+1})$. In the last period, emissions in excess of permit holdings would constitute a violation. Thus the enforcement strategy to induce both reporting and permit compliance in the last period of a program with permit borrowing is the same as the last period enforcement strategy when borrowing is not allowed. Putting all the elements together, the following strategy guarantees complete compliance with minimum enforcement costs in an emissions trading program with both banking and borrowing:

$$\pi_t = \beta p_{t+1} \big/ (\gamma_t + \beta p_{t+1}) \qquad \text{for } t = 0, 1, \ldots, T-1, \tag{15}$$

$$\pi_T = \phi_T \big/ (\gamma_T + \phi_T), \tag{16}$$

$$\phi_T = p_T. \tag{17}$$

As when permit borrowing is not allowed, compliance can be maintained with imperfect monitoring if and only if the reporting violation penalty is positive, and setting it as high as possible conserves monitoring costs.

## 4. The Evolution of Enforcement

While the enforcement strategies we propose suggests how a policy of emissions trading with bankable permits should be enforced in every period, the particular evolution of enforcement and the present value of enforcement costs over the life of a program will depend on the dynamics of permit prices. In this section we draw further conclusions about the evolution of enforcement over time. Because the enforcement strategies we propose induce perfect compliance in every period, we can simply combine our analysis with existing models of the evolution of permit prices that implicitly assume perfect and costless enforcement.[13]

As before let us begin with a program that precludes permit borrowing. Schennach (2000) has provided a theoretical analysis of the time paths of aggregate emissions and permit prices in a model with bankable permits that she constructed to mimic the $SO_2$ regulations. In Schennach's analysis, an industry with a fixed number of firms faces an aggregate emissions standard that is relatively high in the first phase of an emissions trading program, but then is reduced in the second phase. Given her implicit assumption that enforcement induces complete compliance, as well as her explicit assumptions that abatement costs and the discount rate do not change over time and that there is no uncertainty about future permit prices, she shows that firms will save permits during the first phase of the program to smooth out the decrease in the aggregate standard. During this banking period the price of permits rises at the rate of discount (that is, $p_t = \beta p_{t+1}$). At some point during the second phase of the program, say $\bar{t}$, the permit bank is exhausted. Thereafter, facilities in aggregate will hold their emissions to the aggregate standard imposed in every period of the second phase, and the permit price remains constant. In sum, then, the emissions permit price is at its minimum of $p_0$ in the initial period of the policy, it rises at the rate of discount through the banking phase until it reaches a maximum at $p_{\bar{t}}$, and remains at that level until the termination of the program.

What would the price path derived by Schennach imply for the evolution of the enforcement strategy (11)–(14)? Combine the conditions on monitoring, (11) and (12), with the conditions on the permit violation penalty, (13) and (14), to obtain $\pi_t = p_t \big/ (\gamma_t + p_t)$ for each $t = 0, \ldots, T$. Since permit prices rise while firms are banking permits, to ensure perfect compliance either the level of monitoring or the unit penalty for reporting violations must also rise during this phase. This is so because higher permit prices increase the incentive for firms to under-report their emissions. Since increasing monitoring is costly but increasing penalties is not, it is clear that to minimize the present value of monitoring costs over the life of the program it is the reporting violation penalty that should respond as the permit price increases, not the level of monitoring. A simple way to accomplish this is by

---

[13] It would not be possible to use existing results about equilibrium price paths if we were to analyze imperfect compliance. Weak enforcement and significant noncompliance will likely change equilibrium price paths in ways that have not yet been examined. We feel that this would be an interesting area for future research.

tying the reporting violation penalty directly to the current permit price; that is, by setting $\gamma_t = gp_t$, where $g$ is a fixed positive constant. With this approach the required level of monitoring is a constant, $\pi_t = p_t \big/ (gp_t + p_t) = 1 \big/ (g+1)$, throughout the entire program. More importantly, with $g$ set as high as is practicable, the present value of enforcement costs over the life of the program is minimized.

The permit violation penalty should also increase as the market evolves, but only in three discrete jumps. During the banking period this penalty should be set to zero. Given that the permit price is equal to $p_{\bar{t}}$ when the banking period is over, the permit violation penalty should be set at $\phi_t = p_t - \beta p_{t+1} = (1-\beta)p_{\bar{t}}$, until the last period when it should be increased to $\phi_T = p_{\bar{t}}$.

Now let us allow for permit borrowing. Recall that when permits can be banked and borrowed, permit prices increase at the rate of discount. Combining (15), (16), and (17), with $p_t = \beta p_{t+1}, t = 0, \ldots, T-1$, yields $\pi_t = p_t \big/ (\gamma_t + p_t)$ for each $t = 0, \ldots, T$. Once again if the reporting violation penalty is fixed through time, then monitoring must increase as prices increase. Again, tying this penalty directly to the permit price by setting $\gamma_t = gp_t$ yields a constant probability of detection, $\pi_t = 1 \big/ (g+1)$, and minimizes the present value of enforcement costs over the life of the program.

There is a precedent for tying penalties to current permit prices. The EPA's recently proposed Clear Skies Initiative calls for a unit penalty for permit violations that is tied to the clearing price in the most recent auction of permits (US EPA 2003c). Within 30 days after the end of a compliance period, a firm with excess emissions for the period must offset these excess emissions with an equal number of permits and pay a financial penalty that is equal to the clearing price in the latest EPA auction. After 30 days the penalty increases to three times the latest auction price. In contrast to our suggestion to tie the reporting violation penalty to going permits prices, Clear Skies ties the permit violation penalty to prices.

Since the Clear Skies Initiative allows unrestricted permit banking (but not borrowing), we can easily incorporate its permit violation penalty into our analysis. Suppose that auctioned permits and traded permits carry the same price in a period, and let us set $\phi_t = hp_t$, where $h$ is a positive constant. Then, from (11), reporting violations are deterred with minimal monitoring for periods $t = 0, 1, \ldots, T-1$, by $\pi_t = (hp_t + \beta p_{t+1}) \big/ (\gamma_t + hp_t + \beta p_{t+1})$. From (10), permit violations are deterred if and only if $hp_t \geq p_t - \beta p_{t+1}$. Notice that the required monitoring is again increasing in the permit violation penalty, so enforcement with minimal monitoring requires that $h$ be set to satisfy $hp_t = p_t - \beta p_{t+1}$, or rather $(h-1)p_t = -\beta p_{t+1}$. It is clear then that $h$ should be less than one in all periods. Therefore, from the perspective of minimizing the enforcement costs of inducing compliance, the financial penalties for permit violations in the Clear Skies Initiative appear to be excessive.

Furthermore, even though the Clear Skies permit violation penalty is tied to going permit prices, it cannot by itself eliminate the need to increase monitoring as permit prices increase over time. To see why, suppose that firms are banking permits so that $p_t = \beta p_{t+1}$ across time periods. Substituting this into the monitoring requirement yields $\pi_t = (h+1)p_t \big/ (\gamma_t + (h+1)p_t)$. Clearly, without also tying the reporting violation penalty to the current permit price, more monitoring effort must be applied as permit prices increase over time.

## 5. Further Analysis and Discussion

### 5.1. Intertemporal Trading Ratios and Offset Penalties

Thus far we have assumed that permits trade across periods on a one-to-one basis when both banking and borrowing are allowed. When borrowing is not allowed we have assumed that permit and reporting violations in one period are offset from the next period permit allocation, also on a one-to-one basis. Both of these features are modified in proposed and actual emissions market designs. For example, several authors have noted that unrestricted banking and borrowing of permits on a one-for-one basis is not likely to be efficient. Kling and Rubin (1997) show that with stationary abatement costs and pollution damages from a flow pollutant, the optimal path of aggregate emissions is constant through time. However, unrestricted banking and borrowing will motivate firms to shift emissions toward the present in order to push abatement costs off into the future. To motivate firms to choose the optimal path of emissions, they propose a system of bankable permits in which permits trade across time at a ratio that is the inverse of the discount factor. That is, banked permits grow at the rate $1/\beta$, while borrowed permits must be paid back at that rate. This intertemporal trading ratio discourages the inefficient banking that occurs if permits trade across time on a one-to-one basis.[14]

Furthermore, at least one existing program with permit banking, but not borrowing, penalizes permit violations with an offset from future permit allocations instead of a financial penalty. In the EPA's $NO_X$ Budget Trading Program permit violations are penalized with an offset from a future permit allocation on a three-to-one basis.[15] One-to-one banking of permits is allowed in this program, except for an aggregate restriction.[16]

To incorporate alternative intertemporal trading ratios or alternative offset penalties into our analysis, we first generalize the state equation (3) in the following

---

[14] The efficient trading ratio must be modified if abatement costs and/or pollution damages are not stationary. Other authors have examined the optimal intertemporal trading ratio under a variety of circumstances: Leiby and Rubin (2001) in the case of a pollutant that causes both stock and flow damages; Yates and Cronshaw (2001) when regulators are uncertain about firms' abatement costs, and Innes (2003) when emissions are stochastic.

[15] Similarly, the penalty for excess emissions by Annex 1 countries under the Kyoto Protocol is an offset of 1.3 for each unit of excess emissions (Barrett 2003, pp. 384–385).

[16] If the aggregate bank of permits exceeds 10% of the total allocation of permits in a year, this triggers what is called "progressive flow control", which in effect implies a very heavy discounting of banked permits (US EPA 2004).

way:

$$\begin{aligned} &E_t(x_{t+1}) = \bar{l}_{t+1} + \mu[\pi_t(x_t + l_t - e_t) + (1 - \pi_t)(x_t + l_t - r_t)], \\ &\mu = \begin{cases} \mu^0 & \text{if } x_t + l_t - e_t \geq 0 \text{ or } x_t + l_t - r_t \geq 0, \\ \mu^1 & \text{if } x_t + l_t - e_t < 0 \text{ or } x_t + l_t - r_t < 0. \end{cases} \end{aligned} \tag{18}$$

Alternative intertemporal trading ratios when permits can be banked and borrowed can be modeled by setting $\mu = \mu^0 = \mu^1$. Modeling a program with unrestricted banking, but no borrowing, that penalizes permit violations with an offset instead of a financial penalty implies $\phi_t = 0$, $\mu^0 = 1$, and $\mu^1 > 1$.

With (18), Lemma 1 in section 2 is the same except that equation (6) becomes $E_t[J_{t+1}(x_{t+1})] = -\mu p_{t+1}[\pi_t(x_t + l_t - e_t) + (1 - \pi_t)(x_t + l_t - r_t)] + \tilde{C}_{t+1}$. (The proof of Lemma 1 using the state equation (18) proceeds in the very same fashion). The value function (8) is similarly modified: $J(x_t; e_t, l_t, r_t) = c(e_t) + p_t l_t + \phi_t(r_t - (x_t + l_t)) + \pi_t(\gamma_t + \phi_t)(e_t - r_t) \quad - \beta\mu p_{t+1}[\pi_t(x_t + l_t - e_t) + (1 - \pi_t)(x_t + l_t - r_t)] + \beta\tilde{C}_{t+1}$. As before, all reporting violations in $t < T$ are deterred if and only if $J_r \leq 0$. Doing so with minimal monitoring requires $\pi_t = (\phi_t + \mu\beta p_{t+1})\big/(\gamma_t + \phi_t + \mu\beta p_{t+1})$. When permit borrowing is not allowed, permit violations are deterred if and only if $J_l = p_t - \mu\beta p_{t+1} - \phi_t \leq 0$.

Consider a program with both banking and borrowing and an intertemporal trading ratio $\mu = \mu^0 = \mu^1$. In this case the equilibrium price path is partially characterized by $p_t = \mu\beta p_{t+1}$. Since permits can be borrowed, there is no permit violation penalty for all periods except the final period. Therefore, deterring reporting violations with minimal monitoring requires $\pi_t = \mu\beta p_{t+1}\big/(\gamma_t + \mu\beta p_{t+1}) = p_t\big/(\gamma_t + p_t)$. This is exactly the same monitoring strategy that is required when permits trade across time on a one-to-one basis. However, since the trading ratio affects the equilibrium path of permit prices, it also affects the evolution of firms' incentives to under-report their emissions. As before, tying the reporting violation penalty to the equilibrium permit price by setting $\gamma_t = gp_t$ implies a constant monitoring probability, $\pi_t = 1\big/(g + 1)$, for all periods. In general this strategy makes monitoring independent of the evolution of permit prices, and consequently, independent of the intertemporal trading ratio. Clearly, then, the same monitoring strategy can be applied for all alternative trading ratios.

Now consider a program with one-for-one permit banking, but no permit borrowing, that penalizes permit violations with only an offset penalty; that is, $\mu^0 = 1$, $\mu^1 > 1$, and $\phi_t = 0$. The required level of monitoring in this case is $\pi_t = \mu^1\beta p_{t+1}\big/(\gamma_t + \mu^1\beta p_{t+1})$, which is clearly increasing in the offset penalty. As with a financial penalty for permit violations, a higher offset penalty increases the incentive firms have to under-report their emissions. Thus, the offset should be set as low as possible while making sure that all permit violations are deterred; that is, $\mu^1$ should be chosen to satisfy $p_t - \mu^1\beta p_{t+1} = 0$.

With violations deterred and no permit borrowing, the equilibrium price path evolves in the same fashion as when permit violations are deterred by a financial penalty; that is, market equilibrium requires $p_t \geq \beta p_{t+1}$, and firms bank permits only if $p_t = \beta p_{t+1}$. Thus, if firms are banking permits, $p_t - \mu^1\beta p_{t+1} = p_t(1 - \mu^1) = 0$, indicating that the offset should be set to one. If firms are not banking permits and $p_t > \beta p_{t+1}$, then the offset is $\mu^1 = p_t/\beta p_{t+1}$. As we are discovering with high permit violation penalties of all forms, from the perspective of minimizing enforcement costs, the three-to-one offset penalty in the $NO_X$ Budget Trading Program appears to be too high. Setting the offset at such a high level only makes sense if firms are not banking permits because real permit prices are continually falling by two-thirds across time periods.

### 5.2. Comparison to Static Enforcement

An important conclusion of our analysis is that a permit violation penalty appears to play only a minor role in maintaining compliance in an emissions trading program with bankable permits. The existing literature on compliance and enforcement of emissions trading programs reaches the opposite conclusion. This literature examines only static models, and only a few papers in this literature examine enforcement with self-reported emissions. Without self-reporting (for example the models of Malik 1990, 1992; Keeler 1991; van Egteren and Weber 1996; Stranlund and Dhanda 1999), a permit violation penalty plays an important role in maintaining compliance with imperfect monitoring and increasing this penalty reduces monitoring costs. Thus, at least one reason for our finding that a high permit violation penalty serves no purpose is the presence of the self-reporting requirement. Recall that we argued in the introduction, however, that requiring self-reported emissions is a necessary component of any enforcement strategy for bankable permits when monitoring is imperfect. Therefore, one cannot simply eliminate the self-reporting requirement to restore the importance of permit violation penalties.

It is important to note a clear distinction between enforcing static and dynamic emissions trading programs that is revealed in the preceding discussion: self-reporting in not necessary to maintain compliance in static settings, but it is critical in dynamic settings.

While the self-reporting requirement is an important factor in our finding that the permit violation penalty has only limited deterrence value in enforcing dynamic emissions trading programs, it cannot be the sole factor. Stranlund and Chavez (2000) included a self-reporting requirement in a static model of compliance in an emissions trading program, and also concluded that permit violation penalties can be increased to conserve monitoring costs. The difference between our results and those in Stranlund and Chavez (2000) is not due to the fact that one is dynamic while the other is static. The difference is due to the fact that Stranlund and Chavez use strictly convex penalty functions, while we use linear penalties to mimic the penalty structure in real-world enforcement scenarios. Using linear penalties in the Stranlund and Chavez model would yield a result identical to the one we derive for the last period of our dynamic model in which we conclude that the permit violation penalty cannot be used to conserve monitoring costs. Overall then, it is the self-reporting requirement together with linear penalties for reporting and permit violations that drives the contrast between our finding that a permit violation penalty provides only limited deterrence and the opposite conclusion one would obtain from the rest of the related literature.

Usually, however, the permit violation penalty should be much lower in the dynamic case that in a static setting. In the static setting with self-reported emissions, the permit violation penalty should be set at the going permit price. This is true in the last period of a trading program with bankable permits as well. In all previous periods, however, when permit borrowing is not allowed, the permit violation penalty should be set to make up the difference between real permit prices across periods; that is, $\phi_t = p_t - \beta p_{t+1}$. When real permit prices are constant across periods the permit violation penalty is set to zero. On the other hand, suppose that firms are not banking permits in a period because the permit price in the next period is expected to be the same as this period's price. Then, $\phi_t = p_t(1-\beta)$, which normally implies a penalty that is a small fraction of the current permit price. For example, if firms discount future expected compliance costs with a discount rate of 6%, then the permit violation penalty is set to only about 5.5% of the current permit price.

Perhaps the most important insight Stranlund and Chavez's (2000) static analysis holds in the dynamic context as well. The required enforcement strategies for bankable emissions permits do not depend on information about the firms' abatement costs: they only depend on how permit prices evolve. Thus, if the penalty schedules are applied uniformly to all firms and permits are traded competitively so that all firms face the same permit price, no firm-specific information is required to set the appropriate enforcement strategy. Information about their production or emissions-control technologies, or their abatement costs more generally, is simply not useful to an enforcer. This is due to the fact that firms' compliance incentives are independent of their abatement costs. In contrast to the static case, however, firms' compliance incentives in the dynamic case may differ if they use different discount rates or form different beliefs about how permit prices will evolve.

## 6. Conclusion

We have proposed enforcement strategies for emissions trading programs with bankable emissions permits that guarantee complete compliance with minimal enforcement costs. Our strategies emphasize imperfect monitoring supported primarily by a unit penalty for reporting violations, and tying this penalty directly to equilibrium permit prices. This approach is quite different from the enforcement strategies of several EPA programs—the $SO_2$ Allowance Trading, $NO_X$ Budget Trading, and the Clear Skies programs—that emphasize high unit penalties for permit violations. We find that a permit violation penalty plays only a limited deterrence role and, in particular, that this penalty cannot be used to conserve monitoring effort. In fact, setting this penalty at a high level appears to increase the amount of monitoring necessary to maintain compliance, because it increases the incentive firms have to under-report their emissions. In contrast, a unit penalty for reporting violations allows regulators to achieve full compliance with imperfect monitoring, and increasing this penalty reduces the level of monitoring necessary to maintain compliance. Moreover, tying this penalty directly to going permit prices allows regulators to monitor firms at a constant level in every compliance period over the life of an emissions trading program.

Our analysis has important policy implications because it suggests that compliance in emissions trading programs with bankable permits can be maintained with imperfect monitoring, but only if regulators focus on penalizing reporting violations rather than permit violations. In particular, this conclusion might broaden the possible application of emissions trading programs to include contexts in which perfect monitoring is impractical or too costly.

Our analysis can be extended in several directions. Perhaps the most obvious extension is to examine how our results would change under uncertainty. The strategies we have derived make enforcement dependent on the evolution of permit prices; thus, uncertainty about future permit prices undoubtedly implies that the enforcement strategy would be modified to account for this uncertainty. It seems unlikely, however, that the main qualitative conclusions of our work would be invalidated. However, designing an optimal policy with uncertainty about aggregate abatement costs may cause one to consider a different role for the permit violation penalty. In this case, the permit violation penalty may be employed as a "safety-valve" tax that would allow firms to pay a price to escape the burden of unexpectedly high abatement costs (Roberts and Spence 1976). Even in this case, though, it is likely that enforcing self-reporting of emissions with a high reporting violation penalty that is tied to the current permit price would remain an important characteristic of the optimal policy.

It would also be useful to modify our model to examine the consequences of noncompliance in an emissions trading program with bankable permits. We have focused on strategies that maintain perfect compliance, but to have a more complete understanding of compliance and enforcement of programs with bankable permits it would be worthwhile to have some understanding of the consequences of not enforcing these programs well. In particular, it would be interesting to investigate how noncompliance, either because of permit or reporting violations, affects banking behavior and equilibrium paths of prices and emissions.

In general, continuing to extend the theoretical foundations of designing and enforcing emissions trading programs in dynamic environments will yield important insights into how regulatory authorities can implement and manage these programs more efficiently.

## Appendix A

**Proof of Lemma 1:** To prove the desired result, we require the following Lemmas 1A and 1B

**Lemma 1A:** *For all periods s for which the firm is perfectly compliant, if* $J_{s+1}(x_{s+1}) = -p_{s+1}x_{s+1} + C_{s+1}$*, then* $J_s(x_s) = -p_s x_s + C_s$*, for some constants* $C_{s+1}$ *and* $C_s$*.*

**Proof of Lemma 1A:** Full compliance in period $s$ implies that no penalties are collected in this period. Therefore, using (2),

$$v(e_s, l_s, r_s, x_s) = c(e_s) + p_s l_s. \tag{A.1}$$

Furthermore, truthful reporting in $s$ ($r_s = e_s$) implies from the state equation (3) that

$$x_{s+1} = \bar{l}_{s+1} + x_s + l_s - e_s. \tag{A.2}$$

Note that $x_{s+1}$ is deterministic, which implies that $J_{s+1}(x_{s+1})$ is also deterministic. Using this fact, the assumption that $J_{s+1}(x_{s+1}) = -p_{s+1}x_{s+1} + C_{s+1}$, (A.1), and (A.2), the period $s$ value function is

$$\begin{aligned} J_s(x_s) &= \min_{e_s, l_s} c(e_s) + p_s l_s + \beta J_{s+1}(x_{s+1}) \\ &= \min_{e_s, l_s} c(e_s) + p_s l_s - \beta p_{s+1}(\bar{l}_{s+1} + x_s + l_s - e_s) + C_{s+1}. \end{aligned} \tag{A.3}$$

Note that $J_s(x_s)$ is linear in $l_s$. Differentiate (A.3) with respect to $l_s$ to obtain $p_s - \beta p_{s+1}$. Note first that if $p_s < \beta p_{s+1}$, then all firms demand an unbounded number of permits. Since this is inconsistent with permit market equilibrium, we can ignore this case.

When permit borrowing is allowed $p_s > \beta p_{s+1}$ is also inconsistent with permit market equilibrium, because this motivates all firms to divest themselves of all their permits in $s$. Therefore, permit market equilibrium requires $p_s = \beta p_{s+1}$ (Kling and Rubin 1997). Substitute this into (A.3) to obtain

$$J_s(x_s) = \min_{e_s, l_s} c(e_s) - p_s(x_s - e_s) - \beta p_{s+1}\bar{l}_{s+1} + C_{s+1}. \tag{A.4}$$

The optimal choice of emissions, $\bar{e}_s$,is uniquely determined by $c'(e_s) + p_s = 0$. Substitute $\bar{e}_s$ into (A.4) to obtain

$$J_s(x_s) = -p_s x_s + C_s, \tag{A.5}$$

where $C_s = c(\bar{e}_s) + p_s\bar{e}_s - \beta p_{s+1}\bar{l}_{s+1} + C_{s+1}$ is a constant.

When permit borrowing is not allowed, $p_s > \beta p_{s+1}$ could be consistent with permit market equilibrium if firms are not banking permits (Rubin 1996). Again all firms have the incentive to divest themselves of all their permits in $s$, but then all firms would be violating their permits. However, the assumption of full compliance in $s$ implies that enforcement prevents this from occurring, and a firm holds just enough permits to cover its emissions; that is, $x_s + l_s = e_s$. Substitute this into (A.3) to obtain

$$J_s(x_s) = \min_{e_s, l_s} c(e_s) + p_s(e_s - x_s) - \beta p_{s+1}\bar{l}_{s+1} + C_{s+1}.$$

Since this is identical to (A.4), we have $J_s(x_s) = -p_s x_s + C_s$, where $C_s$ is a constant.

When borrowing is not allowed, but firms are banking permits, permit market equilibrium requires $p_s = \beta p_{s+1}$. Then, as when permit borrowing is allowed, $J_s(x_s) = -p_s x_s + C_s$, where $C_s$ is a constant. Therefore, we have the desired result. ■

**Lemma 1B:** *The period-j value function is $J_j(x_j) = -p_j x_j + C_j$, for some constants $C_j$, $j = t+1, \ldots, T$.*

**Proof of Lemma 1B:** Given perfect compliance in $T$ and $J_{T+1}(x_{T+1}) = 0$, we have $J_T(x_T) = \min\limits_{e_T, l_T} c(e_T) + p_T l_T$. Permit compliance in $T$ implies $l_T = e_T - x_T$ and $J_T(x_T) = \min\limits_{e_T} c(e_T) + p_T(e_T - x_T)$. Upon substitution of the optimal $e_T$, derived from $c'(e_T) + p_T = 0$, we have $J_T(x_T) = -p_T x_t + C_T$. Using Lemma 1A, by induction the period $T-, T-2, \ldots, t+1$ value functions take the same form, which is the desired result. ■

From Lemma 1B, $J_{t+1}(x_{t+1}) = -p_{t+1}x_{t+1} + C_{t+1}$, for some constant $C_{t+1}$. From the perspective of possibly non-compliant period $t$ choices, $E_t[J_{t+1}(x_{t+1})] = -p_{t+1}E_t(x_{t+1}) + C_{t+1}$. Upon substitution of the state equation (3) we have

$$\begin{aligned} E_t[J_{t+1}(x_{t+1})] &= -p_{t+1}[\bar{l}_{t+1} + \pi_t(x_t + l_t - e_t) + (1-\pi_t)(x_t + l_t - r_t)] + C_{t+1}. \\ &= -p_{t+1}[\pi_t(x_t + l_t - e_t) + (1-\pi_t)(x_t + l_t - r_t)] + \tilde{C}_{t+1}, \end{aligned}$$

where $\tilde{C}_{t+1} = -p_{t+1}\bar{l}_{t+1} + C_{t+1}$ is a constant. The proof of Lemma 1 is complete. ■

## References

Barrett, S. 2003. *Environment and Statecraft*. New York, NY: Oxford University Press.

Boemare, C., and P. Quirion. 2002. “Implementing Greenhouse Gas Trading in Europe: Lessons from Economic Theory and International Experiences.” *Ecological Economics* 43: 213–230.

Cronshaw, M., and J. B. Kruse. 1996. “Regulated Firms in Pollution Permit Markets With Banking.” *Journal of Regulatory Economics* 9: 179–189.

Chavez, C. A., and J. K. Stranlund. 2003 “Enforcing Transferable Permit Systems in the Presence of Market Power.” *Environmental and Resource Economics* 25(1): 65–78.

Ellerman, A. D., and J-P Montero. 2002. “The Temporal Efficiency of SO2 Emissions Trading.” MIT Center for Energy and Environmental Policy Research Working Paper No. 02–003.

Harford, J. 1987. “Self-Reporting of Pollution and the Firm’s Behavior under Imperfectly Enforceable Regulations.” *Journal of Environmental Economics and Management* 14: 293–303.

Innes, R. 1999. “Remediation and Self-Reporting in Optimal Law Enforcement.” *Journal of Public Economics* 72: 379–393.

Innes, R. 2000. “Self-Reporting in Optimal Law Enforcement When Violators Have Heterogeneous Probabilities of Apprehension.” *Journal of Legal Studies* 29: 287–300.

Innes, R. 2001. “Violator Avoidance Activities and Self-Reporting in Optimal Law Enforcement.” *Journal of Law, Economics, and Organization* 17(1): 239–256.

Innes, R. 2003. “Stochastic Pollution, Costly Sanctions, and Optimality of Emission Permit Banking.” *Journal of Environmental Economics and Management* 45(3): 546–568.

Kaplow, L., and S. Shavell. 1994. “Optimal Law Enforcement with Self-Reporting of Behavior.” *Journal of Political Economy* 103(3): 583–606.

Keeler, A. 1991. "Noncompliant Firms in Transferable Discharge Permit Markets: Some Extensions." *Journal of Environmental Economics and Management* 21:180–189.

Kling, C., and J. Rubin. 1997. "Bankable Permits For the Control of Environmental Pollution." *Journal of Public Economics* 64: 101–115.

Leiby, P., and J. Rubin. 2001. "Bankable Permits for the Control of Stock and Flow Pollutants: Optimal Intertemporal Greenhouse Gas Emission Trading." *Environmental and Resource Economics* 19: 229–256.

Livernois, J., and C. J. McKenna. 1999. "Truth or Consequences: Enforcing Pollution Standards with Self-Reporting." *Journal of Public Economics* 73(3): 415–440.

Malik, A. 1990. "Markets for Pollution Control when Firms are Noncompliant." *Journal of Environmental Economics and Management* 18: 97–106.

Malik, A. 1992. "Enforcement Costs and the Choice of Policy Instruments for Pollution Control." *Economic Inquiry* 30: 714–721.

Malik, A. 1993. "Self-Reporting and the Design of Policies for Regulating Stochastic Pollution." *Journal of Environmental Economics and Management* 24: 241–257.

Palacios, M., and C. Chavez. 2005. "Determinants of Compliance in the Emissions Compensation Program, Santiago, Chile." Forthcoming in *Environment and Development Economics* Vol. 10(4): 453–483.

Palacios, M., and C. Chavez. 2002. "Programa de Compensación de Emisiones: Evaluación del Diseño de Fiscalización y su Cumplimiento." *Estudios Públicos* 88 (Primavera): 97–126.

Phaneuf, D. J., and T. Requate. 2002. "Incentives for Investment in Advanced Pollution Abatement Technology in Emission Permit Markets with Banking." *Environmental and Resource* 22: 369–390.

Roberts, M. J., and A. M. Spence. 1976. "Effluent Charges and Licenses Under Uncertainty," *Journal of Public Economics* 5: 193–208.

Rubin, J. 1996. "A Model of Intertemporal Emissions Trading, Banking, and Borrowing." *Journal of Environmental Economics and Management* 31: 269–286.

Schennach, S. 2000. "The Economics of Pollution Permit Banking in the Context of Title IV of the 1990 Clean Air Act Amendments." *Journal of Environmental Economics and Management* 40(3): 189–210.

Stranlund, J. K., C. A. Chavez, and B. Field. 2002. "Enforcing Emissions Trading Programs: Theory, Practice and Performance." *Policy Studies Journal* 30(3): 343–361.

Stranlund, J. K., and C. Chavez. 2000. "Effective Enforcement of a Transferable Emissions Permit System with a Self-Reporting Requirement." *Journal of Regulatory Economics* 18(2): 113–131.

Stranlund, J. K., and K. K. Dhanda. 1999. "Endogenous Monitoring and Enforcement of a Transferable Emissions Permit System." *Journal of Environmental Economics and Management* 38(3): 267–282.

Swift, B. 2001. "How Environmental Laws Work: An Analysis of the Utility Sector's Response to Regulation of Nitrogen Oxides and Sulfur Dioxide Under the Clean Air Act." *Tulane Environmental Law Journal* 14(2): 309–423.

U. S. Environmental Protection Agency. 2000. *Acid Rain Program 1999 Compliance Report.* Washington DC: U.S. EPA Acid Rain Program.

U. S. Environmental Protection Agency. 2003a. *Acid Rain Program 2002 Compliance Report*. Washington DC: U.S. EPA Acid Rain Program.

U. S. Environmental Protection Agency. 2003b. *Tools of the Trade: A Guide to Designing and Operating a Cap and Trade Program for Pollution Control*. Washington DC: U.S. EPA Office of Air and Radiation.

U. S. Environmental Protection Agency. 2003c. *Section-by-Section Summary of the Clear Skies Act of 2003*. Washington DC: U.S. EPA Office of Air and Radiation.

U. S. Environmental Protection Agency. 2004. *$NO_X$ Budget Trading Program: 2003 Progress and Compliance Report*. Washington DC: U.S. EPA Office of Air and Radiation, Clean Air Market Programs. Available at http://www.epa.gov/airmarkets/fednox/index.html.

U. S. General Accounting Office. 2001. "EPA Should Improve Oversight of Emissions Reporting by Large Facilities." GAO-01–46.

Egteren van, H., and M. Weber. 1996. "Marketable Permits, Market Power, and Cheating." *Journal of Environmental Economics and Management* 30: 161–173.

Yates, A. J., and M. B. Cronshaw. 2001. "Pollution Permit Markets with Intertemporal Trading and Asymmetric Information." *Journal of Environmental Economics and Management* 42: 104–118.