CEEPR **Center for Energy and Environmental Policy Research**

The Economics of Pollution Permit Banking in the Context of Title IV of the 1990 Clean Air Act Amendments

by

Susanne M. Schennach

98-007 WP **May 1998**

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

The Economics of Pollution Permit Banking in the Context of Title IV of the 1990 Clean Air Act Amendments

by

Susanne M. Schennach

ABSTRACT

Tradable pollution permits are the basis of a new market-based approach to environmental control. The Acid Rain Program, established under Title IV of the Clean Air Act Amendments of 1990, and aimed at drastically reducing the SO_2 emissions of electricity generating units in the US, is the world's first large-scale implementation of such a program.

An important feature of this program is that pollution permits, called allowances under Title IV, can be banked for future use. This thesis introduces a model of the collective banking behavior of affected units in the context of Title IV. The present theoretical investigation differs from previous work by its rigorous treatment of the constraint that allowances can only be banked, but never borrowed from future allocations, a consideration which has important consequences. The model presented captures the effects of the changes in electricity demand, the number of affected units, environmental regulations and technological innovations on the utilities' banking behavior and on the allowance price. The effect of uncertainty on the banking behavior is explored, and an analysis of how the allowance market would react in a world of uncertainty to various circumstances is then presented.

The author would like to thank Denny Ellerman, Paul Joskow, Dick Schmalensee and Liz Bailey for their helpful comments, as well as the Massachusetts Institute of Technology for financial support.

TABLE OF CONTENTS

I. INTRODUCTION

The Acid Rain Program, established under Title IV of the Clean Air Act Amendments of 1990, is aimed at drastically reducing the SO_2 emissions by electric utilities.1 Accordingly, it sets a permanent cap to SO_2 emissions by utilities at about half of their annual emissions in 1980. These reductions were to be implemented in two stages: Phase I, covering the years 1995 through 1999, applies to the 263 dirtiest electricity-generating units with a capacity above 100 MW, as well as to any units that voluntarily opt into the program early. Phase II, which is to start in the year 2000, will not only restrict emissions by these Phase I units even further, but also impose restrictions upon cleaner and smaller units, so as to ultimately cover more than 2000 units.

Title IV is the world's first major market-based approaches to environmental regulation. At its core is the virtually unrestricted trading of emission allowances by electric utilities. An allowance gives the permission to emit one ton of sulfur dioxide. Each unit has been allocated in advance a certain annual number of allowances mainly based upon its average heat input in the baseline period.[2] Due to the trading component of the Acid Rain Program, a unit can reduce its emissions below the allocated number of allowances, transferring the unused permits to other units within the same plant or holding company or selling them to other units and brokers. Or it can decide not to abate its emissions, and purchase allowances from other units, to cover the emissions above its allocation of allowances. In addition to this spatial trading, Title IV allows for intertemporal trading, so that units can save their allowances for use in the future with lower permit allocations. The only limitations the EPA, that is, the US Environmental Protection Agency, imposes on the trading program are that a unit cannot borrow allowances from its allocations for future years, and that at the end of each year, every unit

---

[1] Another concern of Title IV are NOx emissions. In the following, I will focus on the SO_2 aspect of the program.

[2] The actual allocation mechanism is somewhat more complex. For a more complete description, see, for example, Joskow and Schmalensee (1998).

has to have enough allowances in its 'account' to cover its SO_2 emissions for the year, these allowances are then deducted from the account.[3]

In contrast to the traditional 'command-and-control' approach to environmental regulation, Title IV of the 1990 Clean Air Act accords a lot more freedom to the generating units, as long as they can provide enough allowances to account for their annual emissions: No specific technology is required, and no uniform emissions rate or percentage reduction is imposed. Thus, each unit can decide whether to install a scrubber, for instance, switch to coal with a lower sulfur content, or simply purchase allowances. The trading program gives units with low cost methods of abatement an incentive to overcomply and sell the remaining permits to other sources that would otherwise have to face very costly emissions reductions. Unlike the previous programs, Title IV relies on the market to achieve the cumulative emissions limits at least cost.

Several empirical studies of this new market-based approach have included investigations of the spatial and intertemporal trading behavior by the units subject to SO_2 regulation.[4] However, despite the large interest in this and similar trading programs, important theoretical issues underlying any trading program remain unexplored. In particular, while the spatial component of trading has been investigated in depth since the early 70's,[5] theoretical analyses of the dynamics of allowance banking have only recently received interest.[6]

Previous theoretical investigations of the banking dynamics (Cronshaw and Kruse (1996), Rubin (1996), Kling and Rubin (1997)) are very general in many respects. No particular functional form for the costs functions is assumed and the time-dependence of the imposed emission cap is left unrestricted. Cronshaw and Kruse (1996) focus on the effect of profit regulations on the firms' pollution permit trading and saving behavior. Rubin (1996) presents an analysis of emissions trading, banking and borrowing in a continuous-time model. In Kling and Rubin (1997), the case of emission trading when

---

[3] To give credibility to this emissions trading program, each unit has had to install a continuous emissions monitoring system, or CEMS, to ensure that the SO_2 emissions are measured accurately.
[4] See, for example, Ellerman et al. (1997), EIA (1997), or EPA's Acid Rain Program Homepage.
[5] See, for instance, Mongomery (1972) and Tietenberg (1985).
[6] Cronshaw and Kruse (1996), Rubin (1996) and Kling and Rubin (1997).

both banking and borrowing are allowed is explored and a suggestion is made about how borrowing can be regulated so as to minimize negative environmental impacts.

However, none of these analyses adequately explore constraints on the intertemporal trading of permits, an important feature of many pollution permit programs. Title IV, for instance, does not allow borrowing of emissions permits issued in the future for present use. This requires the bank of allowances to remain nonnegative at all times, a constraint that is not trivial to take into account. For this reason, previous attempts to include this constraint (Cronshaw and Kruse (1996), Rubin (1996)) have been limited to deriving first-order conditions for optimality where the Lagrange multiplier of the non-negativity constraint is left unevaluated. While providing some insight into the effect of the inability to borrow from the future, this approach does not provide an explicit expression for the optimal paths of emissions and the allowance price, and fails to answer even simple questions such as "for how long will the units be willing to bank allowances?" or "if a new cost-effective abatement technology becomes available in the future, how will the path of emissions be modified?".

We explicitly take into account the non-negativity constraint of the allowance bank. Admittedly, to keep our analysis tractable, we need to make more restrictive assumptions about the functional forms of the cost functions and the time-dependence of the emissions cap. While previous analyses, limited to first-order conditions, explicitly model only the dependence of current emissions on the *current* abatement cost and environmental standards, our analysis clearly couples the present optimal level of emissions not only to current, but also to *future* abatement costs, electricity demand, or environmental regulations.

Another aspect of tradable permit programs that has received little attention in the theoretical literature is the effect of uncertainty on the units' banking behavior. Deviations from the results obtained under certainty have essentially been attributed to transaction costs.[7] Yet it has been shown in the commodity market literature[8] that these deviations can occur even in the absence of transaction costs. Along these lines, we show that, in the

---

[7] Bailey (1996a).
[8] See, for example, Williams and Wright (1991)

absence of transaction costs, the non-negativity constraint on the allowance bank introduces deviations from the paths obtained under certainty.

Previous theoretical studies have so far also ignored the impact of changes in expectations on the banking dynamics. We provide various examples describing how the paths of the allowance price and of emissions are updated when new information about the future becomes available.

This paper thus seeks to address various unexplored theoretical aspects of the economics of allowance banking in the context of Title IV of the Acid Rain Program. Throughout the paper, a special effort will be made to place the pollution permits banking problem in a setting that is rich enough to model the important features of the allowance market and yet simple enough to enable an approach that goes beyond simple first-order conditions.

Section II presents the units' behavior in a world of perfect foresight. After recalling the previously established first order condition required for optimality, we show that, under mild conditions, the banking behavior can be divided into two periods, namely a period when allowances are banked for future use followed by a period when they are not. We then outline the general framework which enables us to determine the length of the banking period and the absolute level of the emissions and the allowance price, in addition to the changes in emissions and price as a function of time, as given by the first-order condition. In order to gain insight about the factors influencing the banking behavior, we first examine the simple case of constant demand for electricity. Various useful extensions of this simple model are then presented: The effect of growth in demand, the implications of the increase in the number of affected units in Phase II and the consequences of technological innovations are analyzed.

In Section III, we then describe how uncertainty affects the results derived under certainty, explaining the origin and the direction of the biases caused by neglecting uncertainties. Based on our global approach derived under certainty, we introduce a simple way to take into account changes in the units' expectations as time progresses.

II. THE ALLOWANCE MARKET UNDER CERTAINTY

2.1 Background Information

In this section we will briefly discuss previously established results upon which our model is built, define the most important variables and introduce our main assumptions.

The electricity generating units face an intertemporal optimization problem. In each period of time, they have to decide by how much they want to reduce their SO_2 emissions, considering the current and future costs of abatement, as well as the fact that they have to account for the remaining tons of SO_2 with allowances taken from their currently available allowances, instead of saving them for future use. The crucial constraint under the set-up of Title IV is that the bank of allowances has to be non-negative at any time; that is, the units cannot borrow allowances from their future allocations. Thus we arrive at the following optimization problem:

$$\min_{\{e_0, e_1, \ldots\}} \left\{ \sum_{t=0}^{\infty} (1+r)^{-t} c_t(a_t) \right\} \qquad \text{subject to } \begin{cases} S_{t+1} = S_t + Y_t - e_t \\ S_{t+1} \geq 0 \end{cases},$$

where

a_t is the number of tons of SO_2 abated by all units at time t; $a_t = \varepsilon_t - e_t$,

e_t the actual SO_2 emissions at time t, after any abatement has taken place,

ε_t the SO_2 emissions that would be needed to satisfy the demand for electricity from Title IV units at time t without any SO_2 abatement requirements (counterfactual emissions),

$c_t(a_t)$ the cost of abating a_t tons of SO_2 at time t,

Y_t the total number of allowances with vintage time t issued to all affected units,

S_t the stock of allowances (bank) available at the beginning of period t, and

r the riskless rate of interest, assumed to be constant over time for notational convenience.

The first order condition for this problem is well known and fully derived for instance in Cronshaw and Kruse (1996). We will simply state this condition and give it a

simple interpretation. The optimum path of emissions e_t is achieved when the following equality holds:

$$m_{t+1}(a_{t+1}) = (1+r)(m_t(a_t) - \lambda_t) \qquad (1)$$

where $m_t(a_t)$ is the marginal cost $c_t'(a_t)$, which is increasing in a_t.[9] λ_t is the Lagrange multiplier associated with the constraint $S_{t+1} \geq 0$.

When the constraint is not binding, λ_t is zero and the marginal cost simply increases at the rate of interest. If the marginal cost of abatement in the future, discounted to the present, is higher than the present marginal cost, units will be willing to save more allowances for future use, because by doing so, they decrease their discounted future cost by more than they increase their present cost. On the contrary, a discounted future marginal cost lower than the present marginal cost will tend to decrease savings. Incentives to save more or less will persist until the discounted marginal cost is equalized across all times during the banking period.[10]

When the constraint is binding, that is, when no allowances are being carried over from period t to t+1, then $\lambda_t > 0$ and the marginal cost increases at a rate less than the rate of interest. The units would have an incentive to borrow allowances from the future, but the constraint that $S_{t+1} \geq 0$ makes this impossible.

Whether the marginal cost rises at the rate of interest or not over a specific period of time depends, of course, on the factors that determine when banking should occur. To understand these factors, let us consider the marginal cost $m_t(\varepsilon_t - Y_t)$ the units would face in

---

[9] We should remind the reader that our marginal cost function $m_t(a_t)$ is the marginal cost of abatement of all units required to abate SO_2 emissions. To focus on the intertemporal aspect of allowance trading, we take optimal spatial trading as given; that is, at each point in time, the units trades with each other to equalize their marginal cost of abatement. The marginal cost resulting from this spatial trading is the marginal cost of the whole industry at that point in time. To construct the marginal cost function used in the following analysis, we plot the marginal cost of the industry for each level of abatement by all the units. Note that optimal spatial trading requires that units make optimal abatement decisions and that the allowance market be efficient. While the optimality of the firms' abatement decisions needs to be assumed, the existence of a unique allowance price supports the efficiency of the allowance market. And indeed, the $S0_2$ allowance market has experienced a convergence of various price indices to a uniform allowance price (Ellerman et al. (1997), p.27-28).

[10] This is simply Hotelling's rule.

each time period if banking were not allowed. Every time this marginal cost increases faster than the rate of interest, there is an incentive for units to save allowances for future use. Such savings would increase the current marginal cost and reduce the future marginal cost, so as to prevent the marginal cost from increasing at a rate greater than the rate of interest. More formally, we can state that if the marginal cost obtained when no banking is allowed, $m_t(\varepsilon_t - Y_t)$, rises at a rate faster than r in an interval $[t_0,t_1]$, then the allowance price obtained when banking is allowed, $m_t(\varepsilon_t - e_t)$, will rise at rate r in an interval $[T_0,T_1]$ such that $[t_0,t_1] \subseteq [T_0,T_1]$.

The main reason for the banking of allowances under Title IV is undoubtedly the phase-in aspect of the trading program. To capture the idea of a trading program with more stringent regulations following an initial, less restrictive compliance period in our model, we define a Phase I, extending from 0 to time T, where Y_H allowances are issued per unit time and a phase II, starting at time T, where a smaller number Y_L of allowances are issued per unit time. As is certainly the case in the SO_2 Program, the units are assumed to know their allowance allocations in advance. We first assume that the number of units affected in the two Phases remains constant. As we will see later, relaxing this assumption does not change our results significantly.

Without the possibility to carry over allowances from one period to the next, there would be a substantial jump in the marginal cost at the onset of Phase II, when the number of allowances drops from Y_H to Y_L. Given the possibility to bank, the units will, of course, smooth out this jump in their marginal cost of abatement.

In addition to the reduction in the number of allowances issued, the units face other fluctuations in their marginal cost due to, for example, fluctuations in the electricity demand ε_t or fluctuations in the marginal cost function itself, $m_t(\cdot)$. In what follows, we will assume that the fluctuations in the electricity demand and in the marginal cost function never cause the marginal cost $m_t(\varepsilon_t - Y_t)$ to increase faster than the rate of interest. Under this assumption, the only incentive to start banking allowances is to prevent the jump in

the marginal cost at the beginning of phase II, at time T.[11] Thus, our assumptions imply that the banking period is unique: It starts in Phase I and ends at some time τ in Phase II.[12]

Under the same assumption, we can further show that τ is actually finite. There is an incentive to save in period t for a period t'>T in the distant future only if the marginal cost in period t', discounted to time t, is higher than the current marginal cost – that is, if $(1+r)^{-(t'-t)}\, m_{t'}(\varepsilon_{t'}-Y_{t'}) > m_t(\varepsilon_t-Y_t)$. But if the marginal cost function $m_t(\varepsilon_t-Y_t)$ increases strictly less than the rate of interest beyond the jump at T, then there exists a t'>T large enough so to make the left hand side smaller than any positive marginal cost today. Thus, for t' large enough, the incentive to save does not exist anymore and banking has to stop at some finite time τ>T. This banking period can, in principle, start anytime in Phase I, but in the context of Title IV, units are willing to start saving from the beginning of Phase I. We thus simply identify time 0 as the beginning of the banking period as well as the beginning of Phase I.[13]

2.2 General framework

Equation (1) only gives us the relationship of the marginal cost between different time periods. However, it fails to provide us with clear information about the length of the banking period as well as the shape and level of the path of SO_2 emissions. In order to find these quantities and analyze their dependence on changes in the various parameters of the model, it is convenient to switch from a discrete time optimization problem to a continuous time approach. The identities used so far translate very easily from the discrete approach to the continuous one. The equalization of the discounted marginal cost during the banking period is simply written as:

$$m_t(\varepsilon_t - e_t) = e^{rt} m_0(\varepsilon_0 - e_0) \quad \text{for } t \le \tau$$

[11] Even though these fluctuations are not large enough to create a banking period by themselves, they do influence the number of allowances saved during the banking period (which is caused by the drop in the number of allowances issued).

[12] In principle, the framework of our model can be extended to allow for multiple banking periods. However, we will not explore this here.

[13] This amounts to assuming that the constraint $S_t \ge 0$ is never binding in Phase I.

while after the end of the banking period, we have:

$$m_t\left(\varepsilon_t - e_t\right) = m_t\left(\varepsilon_t - Y_L\right) \text{ for } t \geq \tau,$$

since the banking period has to end in Phase II, when Y_L allowances are distributed.

The path of the allowance price and the path of emissions are then given by:

$$P_t = \begin{cases} e^{rt} m_0\left(\varepsilon_0 - e_0\right) & \text{for } t \leq \tau \\ m_t\left(\varepsilon_t - Y_L\right) & \text{for } t > \tau \end{cases} \qquad (2)$$

$$e_t = \begin{cases} \varepsilon_t - m_t^{-1}\left(e^{rt} m_0\left(\varepsilon_0 - e_0\right)\right) & \text{for } t \leq \tau \\ Y_L & \text{for } t > \tau \end{cases} \qquad (3)$$

In equations (2) and (3), two unknowns, namely e_0 and τ , still have to be determined before the allowance price and emissions path can be fully identified. To find these unknowns, we first note that at $t=\tau$, the path of emissions must be continuous:

$$e_\tau = \varepsilon_\tau - m_\tau^{-1}\left(e^{r\tau} m_0\left(\varepsilon_0 - e_0\right)\right) = Y_L . \qquad (4)$$

The constraint of continuity arises from two factors. On the one hand, a downward jump in the level of emissions at $t=\tau$ is prohibited by the first order condition, as it would result in a jump upward in the path of price. On the other hand, an upward jump in the emissions would imply that allowances given out at $t=\tau$ are not used immediately, indicating that τ is not truly the end of the banking period.

We also know that all the allowances issued over the banking period have to be used in that same period:

$$\int_0^\tau e_t dt = \int_0^\tau Y_t dt , \qquad (5)$$

Combining equations (3) and (5) with

$$Y_t = \begin{cases} Y_H & \text{if } t \leq T \\ Y_L & \text{if } t > T \end{cases},$$

we obtain:

$$\int_0^{\tau}\left(\varepsilon_t - m_t^{-1}\left(e^{rt} m_0\left(\varepsilon_0 - e_0\right)\right)\right)dt = T\left(Y_H - Y_L\right) + \tau Y_L \ . \qquad (6)$$

We now have two equations, (4) and (6), in two unknowns, e_0 and τ. Once these unknowns are found, equations (2) and (3) fully define the path of the allowance price and the path of emissions.

It is interesting to remark that the quantities Y_H and T come into play only through the term $T(Y_H-Y_L)$ which represents the total number of "extra" allowances distributed in Phase I compared to what would have been distributed if Phase II had started immediately. The exact time-dependence of Y_t over Phase I is therefore irrelevant, only the quantity $\int_0^T \left(Y_t - Y_L\right)dt$ matters.[14] This is important in the context of Title IV, since the annual allowance allocations do indeed differ slightly, due to various special legislative provisions discussed in more detail in Joskow and Schmalensee (1998).

Note that the analysis so far is completely general, making no assumptions regarding the specification of the marginal cost function and the growth in demand. In order to make all subsequent derivations analytically tractable, we will from now on model the marginal cost $m_t(a_t)$ by a linear function, $m_t(a_t) = A_t + B_t a_t$.

For notational convenience, we rewrite the marginal cost function in the following way:

$$m_t(a_t) = A_t + B_t a_t = A_t + B_t (\varepsilon_t - e_t) = B_t (\tilde{\varepsilon}_t - e_t) \qquad (7)$$

where the quantity $\tilde{\varepsilon}_t = A_t / B_t + \varepsilon_t$ embodies the effect of changes in both the electricity demand ε_t and changes in the marginal cost of abatement.

[14] This is the case, provided that the fluctuations in Y_t in Phase I are not large enough to cause a temporary depletion of the bank in Phase I. This problem does not apply to the SO_2 trading program, where the fluctuations in Phase I are small relative to the total number of permits issued.

2.3 Constant Demand for Electricity

To gain some insight about the units' possible reaction to the Acid Rain Program, it is instructive to consider first a simple case. In particular, if counterfactual aggregate emissions were constant over time, such as might occur with unchanging electricity demand and no technological change, then $\tilde{\varepsilon}_t = \tilde{\varepsilon}$, and $B_t = B$, the marginal cost function simplifies to $m_t(a_t) = m(a_t) = B(\tilde{\varepsilon} - e_t)$.

We can now solve equations (4) and (6), to express τ and e_0 as a function of known parameters:

$$\tau = \frac{1}{r} f\left(r \frac{T(Y_H - Y_L)}{(\tilde{\varepsilon} - Y_L)} \right) \qquad (8).$$

$$e_0 = \tilde{\varepsilon} - (\tilde{\varepsilon} - Y_L) e^{-r\tau} \qquad (9), \text{ with } \tau \text{ given by (8).}$$

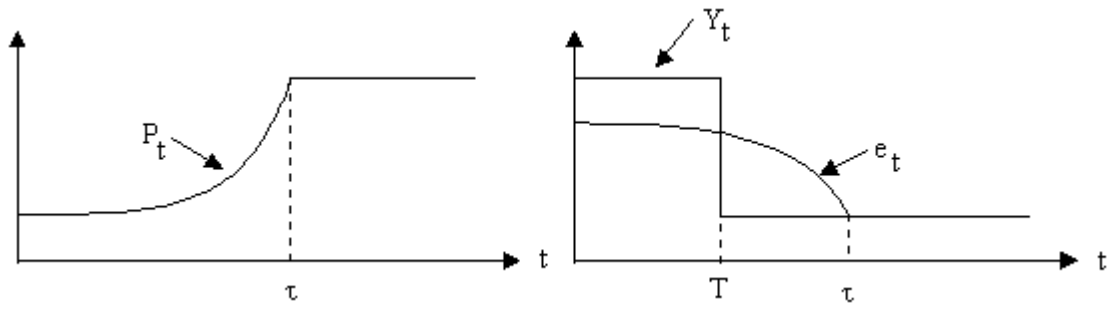
In equation (8), f(s) is a strictly increasing function defined as the solution to the following transcendental equation: $(1 - e^{-f}) = f - s$. For a derivation of this result see Appendix B. This expression illustrates clearly which factors influence the length of the banking period. On the one hand, τ is increasing in $T(Y_H - Y_L)$, the total amount of 'extra' allowances given out in phase I. On the other hand, τ is decreasing in $\left(\frac{A}{B} + \varepsilon - Y_L\right)$. The term $\varepsilon - Y_L$ is a measure of the stringency of phase II: It is the difference between the number of allowances that would be needed to cover the emissions in Phase II in the absence of abatement and the actual number of allowances issued per unit time.

In this case, equations (2) and (3) which give the path price and the path of emissions, can be rewritten by replacing e_0 by its value given by equation (9):

$$P_t = \begin{cases} B(\tilde{\varepsilon} - Y_L) e^{r(t-\tau)} & \text{for } t \le \tau \\ B(\tilde{\varepsilon} - Y_L) & \text{for } t > \tau \end{cases} \qquad (10)$$

$$e_t = \begin{cases} \tilde{\varepsilon} - (\tilde{\varepsilon} - Y_L) e^{r(t-\tau)} & \text{for } t \le \tau \\ Y_L & \text{for } t > \tau \end{cases}. \qquad (11)$$

A graphical representation of these two equations is straightforward:

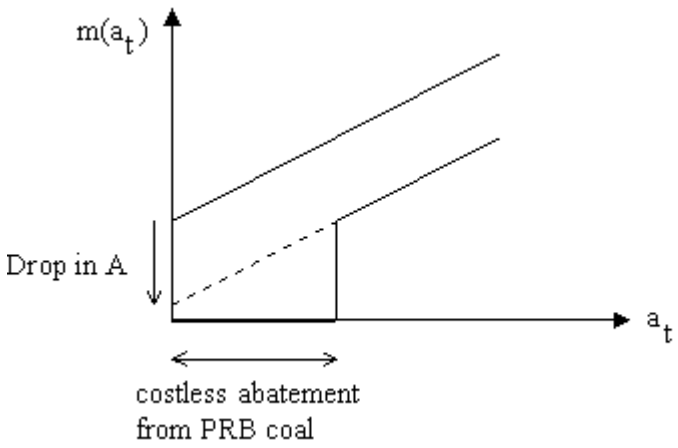


As can be seen in the graph for the emissions path, SO_2 emissions by electricity generating units decrease throughout Phase I as well as in the beginning of Phase II, until they reach the level of allowances issued per unit time in Phase II.

Even this simplified version of the model encompasses not only a derivation of the paths of emissions and price, but also an analysis of the effects of changes in various parameters on the units' banking behavior. As mentioned above, equation (8) clearly illustrates how τ is affected by a change in any of the parameters of the model. In order to analyze the corresponding changes in the whole path of emissions, the interdependence among the variables has to be taken into account, for instance by rewriting equation (11). That is, if a given parameter z affects τ, the emissions path needs to be expressed entirely in terms of z or entirely in terms of τ. The latter approach, which we will use here, turns out to be simpler to carry out. We then simply study the effect of changes in τ on e_t, keeping in mind that the changes in τ are caused by a change in z.

As an example, let us consider the effect on SO_2 emissions of the greater availability of the extremely low-sulfur PRB coal that is due to the lower rail rates resulting from the deregulation of the railroads. As explained in Ellerman et al. (1997), this is equivalent to a permanent costless reduction in SO_2 emissions under Title IV. In our model, this essentially amounts to a decrease in A[15], and thus in $\tilde{\varepsilon}$.

[15] This is true since the abatement made possible by PRB coal is not sufficient to satisfy the Title IV requirements. That is, the units will operate in the upward sloping section of the new marginal cost function.



We now examine the effect of the greater penetration of PRB coal on the path of the allowance price. In the long run, that is, after the bank has been depleted, the decrease in A simply results in an identical decrease in the allowance price, since $P_t = m(a_t) = A + Ba_t$.

We then see, from equation (8), that the length of the banking period τ increases when $\tilde{\varepsilon}$ decreases. Since the allowance price has to rise at rate r, regardless of the presence of PRB coal, the longer banking period, combined with the lower long-run allowance price, inevitably leads to a drop in the initial price, P_0.[16] Thus, the whole price path drops.

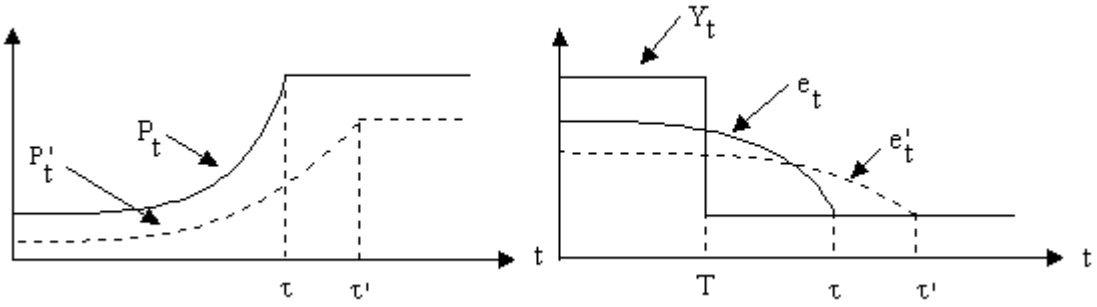
To examine the effect on the emissions path, we express $\tilde{\varepsilon}$ as a function of τ (using equation (33) from Appendix B) and substitute the result into equation (11):

$$e_t = \begin{cases} \dfrac{rT\left(Y_H - Y_L\right)}{1 - r\tau - e^{-r\tau}}\left(e^{r(t-\tau)} - 1\right) + Y_L & \text{if } t \le \tau \\ Y_L & \text{if } t > \tau \end{cases}$$

An increase in τ cannot yield an increase of e_t at all times, since the total number allowances, and thus cumulative emissions, do not change. Similarly, it is incompatible with an overall decrease of e_t. Since it can be shown (see Appendix C) that $\frac{\partial e_t}{\partial \tau}$ is increasing in t, it follows that $\frac{\partial e_t}{\partial \tau}$ must be negative for small t and gradually increase until

[16] This result can easily be shown formally by using the equality between $m(a_0)$ and $e^{-r\tau}m(a_\tau)$ to calculate the total derivative of $m(a_0)$ with respect to $\tilde{\varepsilon}$.

it becomes positive. In short, an increase in τ that results from a decrease in $\tilde{\varepsilon}$ must affect the path of emissions as illustrated below: The initial decrease in emissions becomes progressively smaller, until the emissions eventually increase over what they would have been without PRB coal.



2.4 Growth in Demand for Electricity

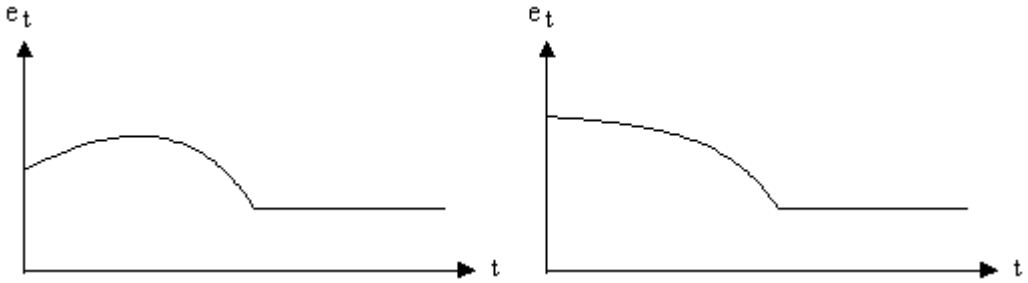
The analysis in the previous section assumes constant demand for electricity produced by the generating units affected by Title IV, to provide some initial insights about the economics of allowance banking. The general framework of Section 2.2, however, allows for the possibility of growth in demand through the time-dependence of ε_t. This section will thus relax the constant demand assumption, thus providing a more realistic picture of the allowance market.

A time-dependence of ε_t translates into a time-dependence of $\tilde{\varepsilon}_t$, since $\tilde{\varepsilon}_t = \frac{A_t}{B} + \varepsilon_t$.[17] Starting from the path of price and the path of emissions given by equations (2) and (3), we can derive equations analogous to equations (10) and (11) derived in the case of constant demand:

$$P_t = \begin{cases} B\left(\tilde{\varepsilon}_\tau - Y_L\right)e^{r(t-\tau)} & \text{for } t \le \tau \\ B\left(\tilde{\varepsilon}_\tau - Y_L\right) & \text{for } t > \tau \end{cases} \qquad (12)$$

$$e_t = \begin{cases} \tilde{\varepsilon}_t - \left(\tilde{\varepsilon}_\tau - Y_L\right)e^{r(t-\tau)} & \text{for } t \le \tau \\ Y_L & \text{for } t > \tau \end{cases}. \qquad (13)$$

These equations, which make no assumption about the type of growth in demand,[18] are very revealing. The price increases at the rate of interest during the banking period, as before. Yet, while the price stops increasing after the banking period in the model without growth, it now keeps increasing, yet at a slower rate - otherwise, banking would not have stopped - governed by the growth in demand. These expressions also underline that in the model with growth in demand, the emissions path is determined by two counteracting forces, namely the interest rate and the growth in demand. If the future were not discounted, changes in emissions would follow changes in demand during the banking period[19], as this would equalize marginal costs across time periods. Hence, an increasing demand would yield an increasing emissions path. However, in the presence of discounting, marginal abatement costs must increase over the banking period. It follows that the abatement must increase over time, which implies that the emissions must lie further and further below the demand $\tilde{\varepsilon}_t$ as time progresses. The discounting of the future eventually bring the emissions down toward Y_L in Phase II. Unlike the effect of growth, the effect of r on e_t increases over time due to the term e^{rt}. It is thus possible that, for small t, the effect of growth in $\tilde{\varepsilon}_t$ dominates the effect of r: In contrast to the constant demand case, where only a decreasing path of emissions can be observed, the path of emissions can actually increase initially. Eventually, the effect of r has to dominate, so that e_t decreases towards Y_L at the end of the banking period. Two cases thus emerge, depending on the relative value of the parameters:



---

[17] Note that the mathematical results presented in this section remain valid whether the changes in $\tilde{\varepsilon}_t$ are due to changes in ε_t or in A_t. To keep the exposition clear, we will focus on the effect of changes in ε_t, rather than A_t. For simplicity, B is for now kept constant.

[18] Except that, as noted before, growth must not cause $m_t(\varepsilon_t - Y_t)$ to rise at a rate faster than the interest rate.

[19] Assuming a time-independent cost function.

The next logical step would now be to solve for τ and e_0 to entirely determine the path of price and emissions, as done in the constant demand case. However, this approach is not analytically tractable, unless a functional form for the growth in demand is postulated. Yet even without assuming anything about the type of growth, we can determine in which direction the beginning of the emissions path (e_0) is shifted due to the introduction of growth into the model, by examining the change in e_0 due to any small change in $\tilde{\varepsilon}_t$ (for all t). Once the change in e_0 is known, the changes in the paths of price and emissions is derived straightforwardly by applying equation (2) and (3) of Section 2.2:

$$P_t = \begin{cases} B\left(\tilde{\varepsilon}_0 - e_0\right)e^{rt} & \text{for } t \le \tau \\ B\left(\tilde{\varepsilon}_t - Y_L\right) & \text{for } t > \tau \end{cases} \qquad (14)$$

$$e_t = \begin{cases} \tilde{\varepsilon}_t - \left(\tilde{\varepsilon}_0 - e_0\right)e^{rt} & \text{for } t \le \tau \\ Y_L & \text{for } t > \tau \end{cases}. \qquad (15)$$

It is worth noting that the price continues to rise after τ, at a rate less than r, following the increase in $\tilde{\varepsilon}_t$.

The change in e_0 due to an arbitrary change in $\tilde{\varepsilon}_t$ is obtained by noting that, despite changes in $\tilde{\varepsilon}_t$, the total number of allowance issued during the banking period must equal the number of allowances used in the banking period, as required by equation (5). If $\tilde{\varepsilon}_t$ changes, e_0 and τ must adapt to preserve this equality. As described in more detail in Appendix D, this amounts to taking the total differential of equation (5) with respect to e_0, τ and $\tilde{\varepsilon}_t$ (for all t). This yields a surprisingly simple result:

$$\Delta e_0 = \Delta\tilde{\varepsilon}_0 - \frac{r}{e^{r\tau} - 1}\int_0^{\tau} \Delta\tilde{\varepsilon}_t \, dt \,. \qquad (16)$$

The corresponding change in price is then given by:

$$\Delta P_0 = B\frac{r}{e^{r\tau} - 1}\int_0^{\tau} \Delta\tilde{\varepsilon}_t \, dt \,. \qquad (17).$$

Introducing growth in demand into the model is equivalent to setting $\Delta\tilde{\varepsilon}_0 = 0$, and $\Delta\tilde{\varepsilon}_t > 0$ for all t > 0. Hence, $\Delta e_0 < 0$ and $\Delta P_0 > 0$. Intuitively, the decrease in emissions at the beginning of Phase I makes sense, since the growth in demand provides an incentive to

save allowances for the future above and beyond the banking due to the decline in the number of allowances issued. Of course, this drop in emissions below the path found in the previous section is not permanent, since cumulative emissions have to remain unchanged. Also, the entire path of the allowance price shifts up, as seen in equation (14). This follows directly from the increase in demand for allowances due to the increasing demand for electricity.

Up to now, the discussion has relied on essentially no assumption about the time-dependence of electricity demand and has lead us to general results linking changes in demand and marginal cost to changes in prices and emissions. An analysis of the effect of growth on the length of the banking period τ necessitates assumptions regarding the type of growth. Focusing on a special type of growth also enables us to obtain more precise results, namely a determination of the length of the banking period and the absolute level of the paths of price and emissions.

The particular case of linear growth in electricity demand provides results which are both simple to derive and to easy to interpret. Linear growth can be viewed as exponential growth in the limit of small growth. We thus expect our results to translate qualitatively to an exponential growth model in the limit of slow growth.[20]

Following the usual procedure, we need to solve equations (4) and (6) for e_0 and τ. With $m_t(\varepsilon_t - e_t) = B(\tilde{\varepsilon}_t - e_t)$ and $\tilde{\varepsilon}_t = \tilde{\varepsilon}_0 + gt$, these equations become:

$$\tilde{\varepsilon}_0 + g\tau - \left(\tilde{\varepsilon}_0 - e_0\right)e^{r\tau} = Y_L$$

$$\int_0^\tau \left(\tilde{\varepsilon}_0 + gt - \left(\tilde{\varepsilon}_0 - e_0\right)e^{rt}\right)dt = TY_H + \left(\tau - T\right)Y_L \ .$$

The solution of these equations (derived in Appendix E) can be written as:

$$\tau = \frac{1}{r}h\left(rT\frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon}_0 - Y_L\right)} \quad , \quad \frac{g}{r\left(\tilde{\varepsilon}_0 - Y_L\right)}\right) \qquad (18)$$

$$e_0 = \tilde{\varepsilon}_\tau - \left(\tilde{\varepsilon}_\tau - Y_L\right)e^{-r\tau}, \text{ with } \tau \text{ given by (18)}, \qquad (19)$$

where h(s,γ) is the solution to the following transcendental equation:

$$\left(1-e^{-h}\right)=\frac{h\left(\frac{1}{2}\gamma h+1\right)-s}{1+\gamma h} \qquad (20)$$

The parameter s is identical to the one used in the constant demand case, while the parameter γ is a measure of growth in demand. Note that when γ = 0, equation (18) reduces to equation (8), that is, h(s,0) = f(s). The function h(s,γ) is increasing in s, but increasing in γ if and only if $s < 1-e^{-2s}$. [21] Consequently, the introduction of growth in demand can lead to either an increase or a decrease in the length of the banking period. The number of extra allowances given out in Phase I greatly influences the result through parameter s.

2.5 Change in the number of units

The reader interested in an analysis of the SO_2 Program might be especially critical of our focus on the behavior of a fixed number of units that experience a decrease in the number of allowances issued per unit of time from Phase I to Phase II. This focus conforms to the reality of a phase-in program such as Title IV in the sense that Phase II of Title IV is indeed the more stringent phase in terms of the number of allowances issued to each unit. More strictly speaking, however, the number of units subject to SO_2 regulation increases sharply between Phase I and Phase II from 445 units in 1995 to over 2000 units in Phase II. The total number of allowances issued actually increases slightly (from, for instance, 8694 in 1995 to around 9400 in Phase II), yet the dramatic increase in the number of units still makes Phase II the more restrictive period.[22] We will now show that

[20] In the case of electricity demand, the growth is indeed quite slow: about 2% per year (from Ellerman ad Montero (1996), Figure 1).
[21] Under a mild condition, described in Appendix E, which is usually satisfied in practice.
[22] As explained in Joskow and Schmalensee (1997), the actual number of allowances varies from year to year. At least as long as there is banking, this does not affect our results in any way, as mentioned earlier.

accounting for the change in the number of units leaves our previous results qualitatively unchanged.[23]

In terms of our model, there are essentially two changes that have to be taken into account. First, the demand of electricity from the units affected by Title IV jumps up at the beginning of Phase II. Let us call ε_t^1 the demand for electricity from units affected in Phase I, and ε_t^2 the demand for electricity from *all* the units included in Phase II.

Second, the marginal cost function will change. Put into the context of Title IV, the marginal cost function considered so far has included all the units eligible for compliance in Phase I.[24] In Phase II, all Title IV units have to be taken into account. In the following analysis, we will consider the most general case, in which both A and B change as we enter Phase II.

Phase I: $$m_t^1(a_t) = A_t^1 + B^1(\varepsilon_t^1 - e_t) = B^1(\tilde{\varepsilon}_t^1 - e_t)$$

Phase II: $$m_t^2(a_t) = A_t^2 + B^2(\varepsilon_t^2 - e_t) = B^2(\tilde{\varepsilon}_t^2 - e_t).$$

The result that the discounted marginal cost is equalized during the banking period remains unchanged:

$$B^1(\tilde{\varepsilon}_t^1 - e_t) = e^{rt} B^1(\tilde{\varepsilon}_0^1 - e_0) \qquad \text{if } t \le T$$

$$B^2(\tilde{\varepsilon}_t^2 - e_t) = e^{rt} B^1(\tilde{\varepsilon}_0^1 - e_0) \qquad \text{if } t > T$$

which implies that the emissions path is given by:

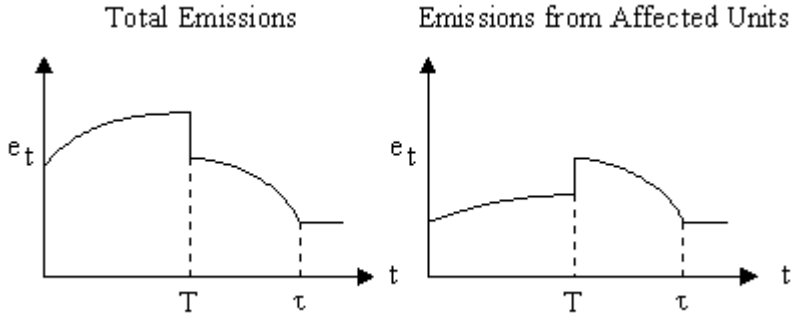
$$e_t = \begin{cases} \tilde{\varepsilon}_t^1 - (\tilde{\varepsilon}_0^1 - e_0)e^{rt} & \text{if } t \le T \\ \tilde{\varepsilon}_t^2 - \dfrac{B^1}{B^2}(\tilde{\varepsilon}_0^1 - e_0)e^{rt} & \text{if } T < t \le \tau \\ Y_L & \text{if } t > \tau \end{cases}. \qquad (21)$$

---

[23] Also, the derivations of the results are essentially identical to the ones in the previous section and are therefore not included in the Appendix.

[24] Eligibility, instead of designation as a Table A unit, is important due to the Substitution and Reduced Ulitization provisions of Title IV. Montero (1997) finds that 623 Phase II units were eligible to opt into the program early in addition to the original 263 Table A units. Units with low marginal costs will of course be likely to opt in.

Note that equation (21) measures the SO_2 emissions by units affected under Title IV in each phase. Total emissions in Phase I are given by e_t *plus* the emissions of non-affected units. The following two graphs illustrate this difference.



In each phase, the emissions have a similar functional form as in equation (15). Again, growth in demand and discounting of the future work against each other. Depending on the relative importance of growth in the first Phase (for Phase I units), growth in the second Phase (for all units) and the interest rate, various paths of emissions can be observed. Thus, a hump in the emissions path can occur in one period, in both, or not at all.

While we observe a jump in e_t at t = T, the path of the allowance price remains, of course, continuous:

$$P_t = \begin{cases} B^1\left(\tilde{\varepsilon}_0^1 - e_0\right)e^{rt} & \text{if } t \le \tau \\ B^2\left(\tilde{\varepsilon}_t^2 - Y_L\right) & \text{if } t > \tau \end{cases}.$$

As before, we can also determine the effect of changes in $\tilde{\varepsilon}_t$ (the demand from affected units at time t) on e_0 and thus on the path of emissions and price. We can use equation (36) of Appendix D to derive an equation similar to equation (16):

$$\Delta e_0 = \Delta\tilde{\varepsilon}_0 - r\left(\left(e^{rT} - 1\right) + \frac{B^1}{B^2}\left(e^{r\tau} - e^{rT}\right)\right)^{-1} \int_0^\tau \Delta\tilde{\varepsilon}_t\, dt \qquad (22)$$

Interestingly, changes in Phase II parameters, such as an increase in growth of demand affecting only non-Phase I units in Phase II, affect the path of emissions in Phase I through e_0.

As before, we can assume linear growth ($\tilde{\varepsilon}_t^1 = \tilde{\varepsilon}_0^1 + g^1 t$ and $\tilde{\varepsilon}_t^2 = \tilde{\varepsilon}_0^2 + g^2 t$) to find τ and its dependence on parameters such as growth. We obtain the following equation, which is the analogue to equation (18):

$$\tau = \frac{1}{r} k\left(s + \Delta s, \gamma, M\right)$$

where:

$$s = rT \frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon}_0 - Y_L\right)}$$

$$\Delta s = \frac{rT\left(\tilde{\varepsilon}_0^2 - \tilde{\varepsilon}_0^1\right) + \frac{rT^2\left(g^2 - g^1\right)}{2}}{\tilde{\varepsilon}_0^2 - Y_L}$$

$$\gamma = \frac{g^2}{r\left(\tilde{\varepsilon}_0 - Y_L\right)}$$

$$M = e^{rT} - \frac{B^2}{B^1}\left(e^{rT} - 1\right)$$

and k(s',γ,M) is the solution to the following transcendental equation

$$\left(1 - Me^{-k}\right) = \frac{k\left(\frac{1}{2}\gamma k + 1\right) - s'}{1 + \gamma k}. \qquad (23)$$

As in our previous analysis, s is mainly a measure of how many extra allowances are given out over Phase I relative to Phase II. Similarly, Δs is roughly (up to a constant) a measure of how many allowances would have to be distributed to non-Phase I units in Phase I to cover their emissions without requiring abatement.[25] Parameter M is a measure of the variation in the slope B of the marginal cost. If $B^1=B^2$, then M=1 and equation (23) reduces to equation (20) obtained in the case of a constant number of units, that is, k(s',γ,1) = h(s',γ). We would thus fall back to the usual equation under linear growth with the exception that s is replaced by s+Δs.

---

[25] This can be seen from:

$$\int_0^T \left(\tilde{\varepsilon}_t^2 - \tilde{\varepsilon}_t^1\right) dt = \int_0^T \left(\left(\tilde{\varepsilon}_0^2 + g^2 t\right) - \left(\tilde{\varepsilon}_0^1 + g^1 t\right)\right) dt = \left(\tilde{\varepsilon}_0^2 - \tilde{\varepsilon}_0^1\right) T + \left(g^2 - g^1\right) \frac{T^2}{2}.$$

The functional form of equation (23) being similar to the one of equation (20), we obtain a similar result concerning the effect of the introduction of growth. Condition (41) simply turns into: k(s+Δs, γ,M) is increasing in γ if and only if[26] $(s+\Delta s) \leq 1 - Me^{-2(s+\Delta s)}$.

In short, although the inclusion of a change in the number of units into the model makes the derivations more complex and certainly changes the results numerically, the analysis essentially remains unchanged. More importantly, the change in the number of units itself only necessitates minor changes in the model. It is mainly the change in the slope of the marginal cost function (B) due to the change in the number units which complicates the matter. When $B^1 = B^2$, the only change required is a redefinition of the electricity demand $\tilde{\varepsilon}_t$. In light of these results, the change in the number of units will be neglected in the remainder of our analysis.

2.6 Technological Change

The flexibility provided by an allowance trading program such as the SO_2 program encourages the use of the most cost-effective abatement methods, which, in turn, may stimulate innovations in environmental control technologies. Since the enactment of Title IV, for instance, reports on improvements in the blending of coal as well as in the removal efficiency and implementation costs of scrubbers have flooded the literature on coal and electricity markets.[27] As these technological innovations are expected to continue and to influence the units' compliance strategies and thus banking behavior, we will now briefly address the issue of technological progress in our model.

In the framework of our model, the effect of technological change can be modeled by allowing the parameters of the marginal cost function to be time-dependent. An arbitrary time-dependence of the marginal cost parameter A_t, keeping B_t constant over time, is straightforward to include; all results we derived concerning the time-dependence

[26] Provided, again, that $1 + \gamma (k - 1) \geq 0$.

of $\tilde{\varepsilon}_t$ can be directly applied, as A_t influences the results only through $\tilde{\varepsilon}_t = \frac{A_t}{B} + \varepsilon_t$. In particular, equations (16) and (17) can be applied to determine the initial shifts in emissions and prices under very few assumptions. Moreover, these equations allow for both continuous and one-time technological progress.

An expected future decrease in marginal cost (that is, $\Delta A_0=0$, and $\Delta A_t<0$ for $t>t_1$), for instance, can be handled by merely reversing the results we obtained when investigating the effect of the introduction of growth. Hence, an expected future decrease in marginal cost causes an overall price decrease and an initial increase in the emissions. The increase in emissions only lasts for part of the banking period, the length of which also adjusts.[28]

Another interesting case arises when the decrease in A_t is uniform over time and starts immediately: $\Delta A_t = \Delta A < 0$ for all t. This is equivalent to the example of PRB coal analyzed in the case of constant demand. We now show that similar conclusions can be obtained when relaxing the constant demand assumption. A uniform decrease in A_t causes an initial decrease in the path of emissions, as a simple application of equation (16) shows:

$$\begin{aligned}\Delta e_0 &= \Delta\tilde{\varepsilon} - \frac{r}{e^{r\tau}-1}\int_0^{\tau}\Delta\tilde{\varepsilon}\,dt \\ &= \Delta\tilde{\varepsilon} - \frac{r}{e^{r\tau}-1}\tau\Delta\tilde{\varepsilon} \\ &= \frac{\overbrace{\Delta\tilde{\varepsilon}}^{<0}}{\underbrace{e^{r\tau}-1}_{\geq 0}}\underbrace{\left(e^{r\tau}-(1+r\tau)\right)}_{\geq 0} \leq 0\end{aligned}$$

($e^{r\tau} \geq 1 + r\tau$ because $e^{r\tau}$ is convex and $1 + r\tau$ is tangent to $e^{r\tau}$ at $\tau=0$.)

This initial decrease cannot persist forever because the cumulative number of allowances used must remain the same, despite the change in the costs of abatement. Also, equation (14) indicates an overall decrease of the path of price.

The two previous cases illustrate the importance of the timing of technological breakthroughs for banking behavior. First, we emphasize that the drop in price is always

---

[27] Examples include Arnold and Smith (1994), Kindig and Godfrey (1991), Walz et al. (1995), and Greenberger (1991). Note that our exposition does not imply that none of the innovations observed in reality would have surfaced in the absence of Title IV.

permanent, while the initial change in emissions does not even persist through the banking period, since cumulative emissions must remain unchanged. Also, while a decrease in the marginal cost always leads to a decrease in price at t = 0, such a decrease will only lead to a decrease in emissions at t=0 if the decrease in the marginal cost at t=0 is large enough relative to the future decrease. In particular, if there is no decrease in marginal cost at t=0, there will actually be an increase in emissions at t=0. The essential difference in the direction of the initial shift in emissions originates from the fact that a future decrease in A_t discourages saving allowances for future use. On the contrary, if the decrease in A_t starts immediately, the lower costs make more allowances available now, thus encouraging savings.

Technological change affecting the parameter B_t is more difficult to handle. In the previous section on the change in the number of units, we indirectly considered the effect of a one-time change in the parameter B. Since the results derived do not depend on whether the change in B is due to the change in the number of units or due to a technological innovation, the same approach can be applied to determine the effects of changes in B, simply replacing T by the time t_1 when the one-time drop occurs.[29]

[28] The adjustment could go in either direction.
[29] We will not consider continuous changes in B here.

III. THE ALLOWANCE MARKET UNDER UNCERTAINTY

3.1 The Path of Price under Uncertainty

Unfortunately, no environmental policy is implemented in the ideal world of certainty we have modeled so far. On the contrary, the electricity units subject to Title IV have to face uncertainty for instance in their marginal cost of abatement, in the regulatory environment, and in the demand for electricity. That is, expectations about the future have to replace perfect foresight. In this section, we will examine how some of the results derived so far are affected by the presence of uncertainty. As we will see, uncertainty changes affects our results in two ways. First, when accurate information about the future is replaced by expected information about the future, the expected path of price and emission are slightly modified: in a sense, the units become slightly more pessimistic. Secondly, the future expected path changes as time progresses, when new information about the future becomes available.

In the context of uncertainty, and assuming risk-neutrality, the units seek to minimize the sum of their *expected* discounted cost:[30]

$$\min_{\{e_0,e_1,\ldots\}}\left\{E_0\left[\sum_{t=0}^{\infty}(1+\mu)^{-t}c_t(\varepsilon_t-e_t)\right]\right\} \text{ subject to } \begin{cases} S_{t+1}=S_t+Y_t-e_t \\ S_{t+1}\geq 0 \end{cases} \quad (24)$$

where

$E_t[.]$ is the expectation value given all the information known at time t,

μ $= r + \rho$, where r is the riskless interest and ρ is the asset-specific risk premium.[31]

As shown in Appendix F, the solution to equation (24) can be obtained by using dynamic programming:

---

[30] We momentarily switch back to discrete time to ease the derivation. The continuous time version of the results follows straightforwardly.

[31] The risk premium is taken as given here; it can be derived from the CAPM, where it represents the nondiversifiable risk of the asset, i.e. the covariance of its rate of return with the market return. See, for instance, Blanchard and Fisher (1996), Ch. 10.

$$E_t\left[\left(m_{t+1}\left(\varepsilon_{t+1}-e_{t+1}\right)\right)\right]=\left(1+\mu\right)\left(m_t\left(\varepsilon_t-e_t\right)-\lambda_t\right). \qquad (25)$$

Thus, as long as we have banking, so that $\lambda_t = 0$, we equalize the discounted *expected* marginal cost of abatement, regardless of the functional form of the marginal cost.

Iterating equation (25) gives:

$$\begin{aligned} m_t\left(a_t\right) &= \lambda_t + \left(1+\mu\right)^{-1} E_t\left[\left(m_{t+1}\left(a_{t+1}\right)\right)\right] \\ &= \lambda_t + \left(1+\mu\right)^{-1} E_t\left[\lambda_{t+1} + \left(1+\mu\right)^{-1} E_{t+1}\left[\left(m_{t+2}\left(a_{t+2}\right)\right)\right]\right] \\ &\vdots \\ &= \sum_{s=0}^{N-1}\left(1+\mu\right)^{-s} E_t\left[\lambda_{t+s}\right] + \left(1+\mu\right)^{-N} E_t\left[\left(m_{t+N}\left(a_{t+N}\right)\right)\right] \end{aligned}$$

so that

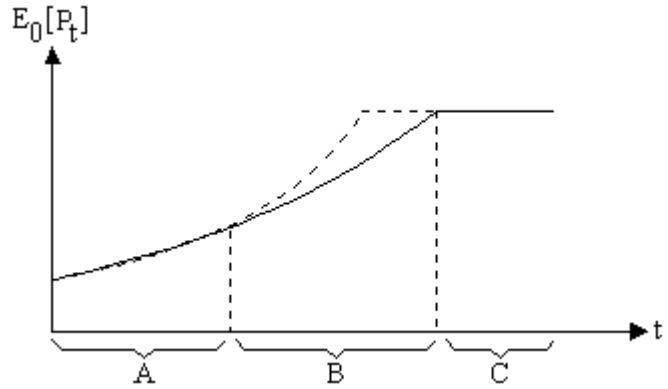
$$E_t\left[\left(m_{t+N}\left(a_{t+N}\right)\right)\right]=\left(1+\mu\right)^N m_t\left(a_t\right)-\left(1+\mu\right)^N\sum_{s=0}^{N-1}\left(1+\mu\right)^{-s} E_t\left[\lambda_{t+s}\right]. \qquad (26)$$

Equation (26) describes the expectation at time t of the path of the allowance price. It tells us that when we expect the bank to be non-empty, so that $E_t\left[\lambda_{t+s}\right]=0 \quad \forall s < N$, the expected price rises at rate μ between time t and t+N. When $E_t\left[\lambda_{t+s}\right]>0 \quad \forall s \in \left[\underline{s},\bar{s}\right]$, so that we assign a positive probability to an empty bank, the expected price will rise at a rate less than μ when $N \in \left[\underline{s},\bar{s}\right]$. The downward offset will increase over the interval $\left[\underline{s},\bar{s}\right]$. Note that realistically, once we have assigned a positive probability to an empty bank in some period $\underline{s}$, we will expect that the bank might possibly be empty in all subsequent time periods as well. When we expect the bank to be depleted without any doubt, then $E_t\left[e_{t+N}\right]=E_t\left[Y_{t+N}\right]$, and the expected price is uniquely determined by

$$E_t\left[P_{t+N}\right]=E_t\left[m_{t+N}\left(\varepsilon_{t+N}-Y_{t+N}\right)\right].$$

To summarize the interpretation of equation (26),

$$E_t\left[\left(P_{t+N}\right)\right]=\underbrace{\left(1+\mu\right)^N P_t}_{\text{Rise at rate } \mu}-\underbrace{\left(1+\mu\right)^N\sum_{s=0}^{N-1}\left(1+\mu\right)^{-s} E_t\left[\lambda_{t+s}\right]}_{\substack{\text{Downward correction increasing with} \\ \text{the likelyhood of an empty bank}}}.$$



In the above figure, A is the period when the bank is certainly nonempty, B, the period when the bank may be empty and C, the period when the bank is certainly empty.[32] The dashed line represents the price that would be observed if there were no uncertainty about whether the bank is empty or not.[33]

In the reality of the SO_2 allowance market, the ranges A and C do exist: Due to the particular phase-in setup of Title IV, with a more restrictive Phase II, market observers are certain that the bank won't be empty a year from now, while they do expect the bank of allowances accumulated in Phase I to be depleted late in Phase II.

It is interesting to note that the form of equation (25) reminds us of the difference equation

$$E_t\left[P_{t+1}\right] = \left(1+\mu\right)P_t - \psi_t$$

on which the present value model of rational commodity pricing[34] is based, where ψ_t is the convenience yield to holding an asset. Bailey (1996a) put this convenience yield into the context of the SO_2 allowance market, attributing it mainly to the transaction costs associated with transferring an allowance.

---

[32] As in the model under certainty, we assume that the banking period is unique by requiring that $m_t(\varepsilon_t - Y_L)$ has a probability zero of increasing at a rate greater than the interest rate.

[33] The graph compares the path obtained under certainty with the expected path under uncertainty given a known initial price P_0 at t=0. In reality, as we will see later, the initial expected price is also affected by the presence of uncertainty, but here, we have artificially set the initial price under certainty and under uncertainty equal to simplify the discussion.

[34] Pindyck (1993).

Yet in our model, we have ignored the effect of transaction costs[35] and the resulting convenience yield. Instead, our λ_t results from the possibility of a stockout, that is, a depletion of the allowance bank.[36] The importance of the non-negativity constraint and its impact on the rate of price increase has been extensively recognized in the commodity market literature.[37]

We now provide two interpretations of λ_t. The term λ_t is nonzero when the constraint $S_t + Y_t - e_t \geq 0$ is binding. As is well known, λ_t is the shadow price of relaxing this constraint when it is binding. More precisely, from equation (25) we can write

$$\begin{aligned}\lambda_t &= m_t(\varepsilon_t - e_t) - (1+\mu)^{-1} E_t\left[\left(m_{t+1}(\varepsilon_{t+1} - e_{t+1})\right)\right] \\ &= -\frac{\partial}{\partial(\Delta e)}\left(c_t(\varepsilon_t - (e_t + \Delta e_t)) - (1+\mu)^{-1} E_t\left[\left(c_{t+1}(\varepsilon_{t+1} - (e_{t+1} - \Delta e_t))\right)\right]\right)\end{aligned}$$

In other words, λ_t is the cost savings that would be incurred if an allowance allocated in period t+1 were made available in period t.

Examining equation (26) gives another, maybe more intuitive, explanation for the importance of λ_t. The electricity generating units as a whole are willing to save an allowance for future use, even when the rate of price increase is less than μ, provided that there is the possibility of an empty bank. The fact that the market as a whole cannot borrow allowances from future allocations introduces a nonlinearity into the model. The market can, for instance, absorb an unexpected downward fluctuation around the expected trend in electricity demand by saving more allowances than expected, thereby smoothing the effect of this shock on the price level. An unexpected upward fluctuation of the same size, however, can only be absorbed by a use of allowances from the existing bank. The buffering capability is limited, since an optimal smoothing may require more allowances than available at the present time. That is, in the case of a stockout, an upward shock in demand increases the price by more than the corresponding downward shock decreases it.

---

[35] Transaction costs constitute only about 1% of the allowance price. For more on this, see Bailey (1996a)
[36] Note that a convenience yield arising from transaction costs and our findings are not mutually exclusive; both effects can be expected to have an influence on the allowance prices actually observed in the real world.
[37] See, for example, Williams and Wright (1991).

Hence, when there is a possibility of a stockout, the units as a whole are willing to save allowances even though $E_t\left[\left(P_{t+N}\right)\right] < \left(1+\mu\right)^N P_t$.

3.2 Certainty Equivalence Property

Although equation (25) along with the bank non-negativity constraint fully define the expected path of price and emissions, an analytic solution is impractical. Numerical simulations have been implemented in similar problems.[38]

To get further analytical insight, we propose instead an approximate solution to equation (25), identifying the direction of the bias caused by the approximation. Taking the expectation at time 0 in equation (25):

$$E_0\left[\left(m_{t+1}\left(\varepsilon_{t+1}-e_{t+1}\right)\right)\right] = \left(1+\mu\right)\left(E_0\left[m_t\left(\varepsilon_t-e_t\right)\right]-E_0\left[\lambda_t\right]\right).$$

Assuming linear marginal cost and a parameter B constant over time, we have:

$$\left(B\left(E_0\left[\tilde{\varepsilon}_{t+1}\right]-E_0\left[e_{t+1}\right]\right)\right) = \left(1+\mu\right)\left(B\left(E_0\left[\tilde{\varepsilon}_t\right]-E_0\left[e_t\right]\right)-E_0\left[\lambda_t\right]\right) \qquad (27)$$

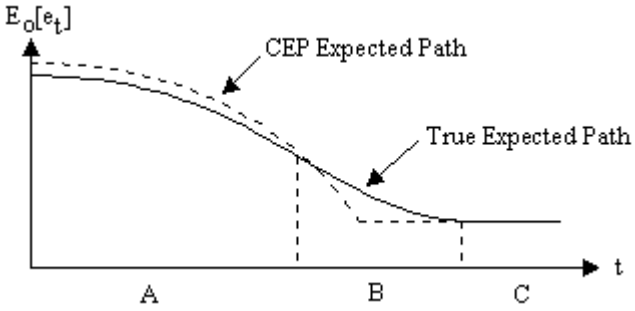
Note that equation (27) is exactly the same as equation (1) obtained under certainty, except for the fact that all quantities are replaced by their expectations and r by μ. It is tempting to find the expected path as we did under certainty. However, the term $E_0[\lambda_t]$ does not behave in the same way as λ_t in the model under certainty: It is non-zero as soon as there is a *possibility* of stockout, even though the *expected* amount of allowances available in the bank is non-zero. It is only when the bank has a zero probability of being empty (region A in the previous graph) that $E_0[\lambda_t]=0$. In this case, the so-called certainty-equivalence property (CEP)[39] applies and iterating equation (27) does give the correct expected path, once $E_0[e_0]$ is known. Of course, the assumption of a linear marginal cost is essential for this.

[38] Deaton and Laroque (1992), Williams and Wright (1991).
[39] See, for example, "The consumption/savings decision under uncertainty in Blanchard and Fisher (1996).

When a stockout becomes a possibility (region B in the previous graph), the term $E_0[\lambda_t]$ can no longer be neglected. Rearranging equation (27), we obtain an upward correction to the CEP path in region B:

$$E_0[e_{t+1}] = \underbrace{E_0[\tilde{\varepsilon}_{t+1}] - (1+\mu)\left(E_0[\tilde{\varepsilon}_t] - E_0[e_t]\right)}_{\text{Certainty equivalent path}} + \underbrace{\frac{(1+\mu)}{B} E_0[\lambda_t]}_{\text{Upward correction}}.$$

Due to this correction relative to the certainty equivalent path in region B, using the CEP over the whole banking period (regions A and B) to solve for $E_0[e_0]$ will give slightly biased values of $E_0[e_0]$. The value of $E_0[e_0]$ obtained through CEP has to be slightly lowered, so that the total expected cumulative emissions remain equal to the cumulative number of allowances issued, despite of the upward correction in region B. Thus, two corrections have to be made to the path obtained with the CEP, as illustrated below.



First, the true expected emissions path lies somewhat below the path predicted through the CEP in the period of certain banking. Second, the uncertainty regarding the end of the banking period smoothes out the cusp obtained at the end of the banking period in the model under certainty. Keeping these corrections in mind, we will use the certainty equivalence property in what follows.

3.3 Changes in Expectations

It is worth pointing out that equation (26) gives the path of the expected price based on information available at time t. When our expectations about the various parameters of the model change, we have to recompute the path from that point on, taking

into account the number of allowances currently in the bank. When new information becomes available, there may be a cusp or even a discontinuity in the path of emissions and of price. Obviously, the path is changed as soon as the information about an event becomes public, even when the event itself occurs in the future.

The analyze the effect of changes in expectations, we first have to translate real world uncertainties into uncertainties regarding the parameters in our model:

- Uncertainty regarding the deregulation in the electric utility market ⇒ uncertainty in ε_t.
- Uncertainty regarding environmental regulation with an effect on the use of coal in the generation of electricity ⇒ uncertainty in ε_t.
- Uncertainty regarding technological change ⇒ uncertainty in A_t or B_t.

We are now ready to consider a few examples that illustrate the impact of changes in expectations on the path of emissions and prices.

Changes in expectations regarding ε_t or A_t can both be simply regarded as changes in $E[\tilde{\varepsilon}_t]$. Under the CEP we can adapt equation (16) to determine the changes in the expected emissions when our expectations change at some time t_0 during the banking period:

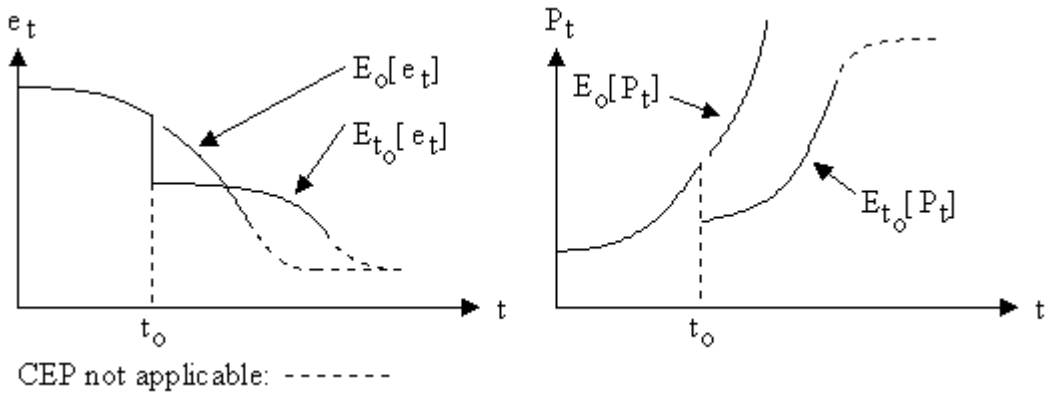
$$\Delta e_{t_0} = \Delta\tilde{\varepsilon}_{t_0} - \frac{r}{e^{r(\tau - t_0)} - 1}\int_{t_0}^{\tau} \Delta\tilde{\varepsilon}_t dt \qquad (28)$$

where $\Delta\tilde{\varepsilon}_t = E_{t_0}\left[\tilde{\varepsilon}_t\right] - E_0\left[\tilde{\varepsilon}_t\right]$.

The changes in future emissions can then be found through:

$$E_{t_0}\left[e_t\right] = E_{t_0}\left[\tilde{\varepsilon}_t\right] - \left(\tilde{\varepsilon}_{t_0} - e_{t_0}\right)e^{r(t - t_0)}. \qquad (29)$$

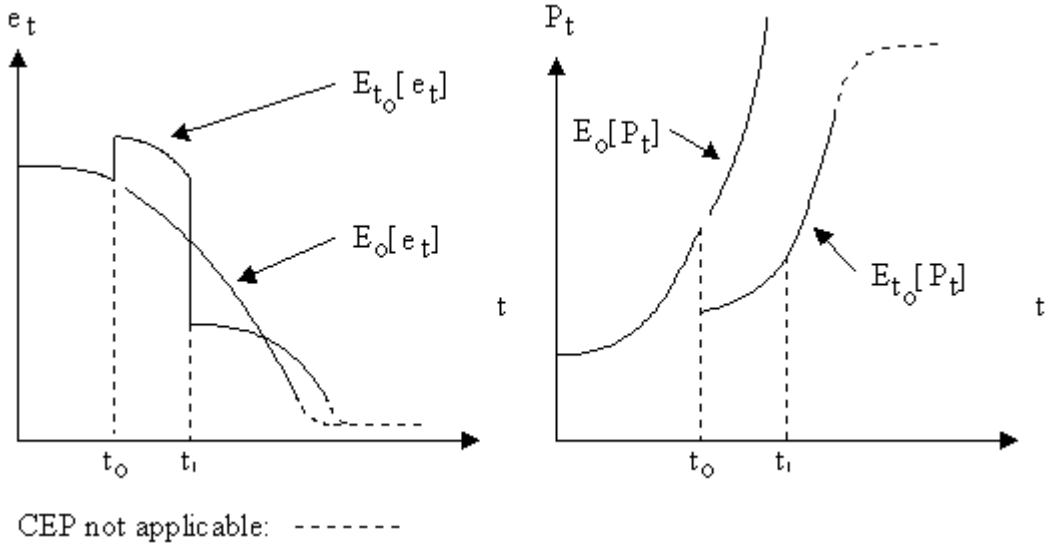
Using these results, we now consider an unexpected permanent decrease in A_t, starting right now at t_0 - as, for instance, in the case of PRB coal - or an unexpected permanent decrease in the level of demand ε_t for coal produced electricity, starting right now at t_0.



A permanent, uniform decrease in A_t or ε_t that is expected for $t \geq t_1 > t_0$ - for instance, the government announces at time t_0 that it plans to implement stricter environmental regulations disfavoring coal at time t_1 - has somewhat different implications, as we will now see:

$$\Delta\tilde{\varepsilon}_{t_0} = 0 \text{ and } \Delta\tilde{\varepsilon}_t < 0 \text{ for } t \geq t_1 > t_0$$

Application of equation (28) shows that $\Delta e_{t_0} > 0$, although demand at t_0 is unchanged, while equation (29) indicates that at $t=t_1$, when the demand is actually expected to drop, we will expect a drop in emissions. This drop in expected emissions at $t=t_1$ is simply due to the drop in expected counterfactual emissions at $t=t_1$. The expected path of price will, of course, not jump at t_1.



The opposite picture emerges if the electricity generating units find out at t_0 that deregulation of the electricity market becomes a lot more likely in the future: Since coal is

a fairly cheap fuel, we can interpret deregulation as an increased demand for electricity generated by already existing coal-fueled units.

We shall not study the effect of uncertainty and changes in expectations concerning B, since we would have to make a correction to CEP even in the period where banking is certain. The amount of work required for a rigorous treatment does not seem to be justified by the few intuitive results that might be obtained.

IV. CONCLUSION

This paper has analyzed in detail the economics of allowance banking in the context of Title IV of the 1990 Clean Air Act Amendments. Going beyond simple first order conditions, we have determined the paths of the allowance price and of SO_2 emissions as well as the length of the banking period as a function of well defined parameters describing the SO_2 allowance market environment under certainty. In contrast to previous theoretical studies, special attention has been given to the fact that programs like Title IV do not allow emissions permits to borrowed from future allocations. We have also determined the impact of changes in these parameters on the allowance price and on emissions under mild assumptions about the time-dependence of electricity demand and the marginal cost of abatement. We have first derived explicit solutions in the case of constant demand for electricity, before handling the more realistic case of growing demand. Incorporating into the model the fact that the number of units subject to SO_2 regulation changes drastically has not changed our results qualitatively. The effect of technological progress has been treated as a straightforward application of our previous analysis.

The most notable effect of the inclusion of uncertainty into the model is the units' willingness to bank allowances even when the expected price rises at a rate less than the rate of interest. We show that this effect is present even in the absence of any transaction costs, as it arises from the non-negativity constraint on the bank of allowances. Using the results obtained under certainty, we derive simple, although admittedly approximate, relationships which enable us to determine the impact of the changes in expectations on the units' banking behavior.

Even hough our analysis has focused on the SO_2 Program, the model is general enough to be adapted to any other tradable permits program that contains a phase-in provision. This is quite important in light of the fact that this type of market-based approach to environmental regulation is likely to become increasingly common in the future and has even been suggested on a world-wide basis to regulate emissions of greenhouse gases.

References

Arnold, B.J., and D.B. Smith (1994). "Coal Blending - Meeting Specifications." *Mining Engineering*. October.

Bailey, E.M. (1996a). "Prices and Pricing in the Market for Sulfur Dioxide Allowances." Unpublished Manuscript. Department of Economics, MIT. October.

Bailey, E.M. (1996b). "Allowance Trading Activity and State Regulatory Rulings: Evidence from the U.S. Acid Rain Program." Working Paper 96-002. MIT-CEEPR, March.

Blanchard, O.J., and S. Fischer (1996). *Lectures on Macroeconomics*. Cambridge: The MIT Press.

Cronshaw, M.B., and J.B. Kruse (1996). "Regulated Firms in Pollution Permit Markets with Banking." *Journal of Regulatory Economics*. Vol. 9, p179-186.

Deaton, A., and G. Laroque (1992). "On the Behavior of Commodity Prices." *Review of Economic Studies*. Vol. 59, p.1-23.

Ellerman, A.D.,and J.P. Montero (1996). "Why Are Allowance Prices so Low? An Analysis of the SO_2 Emissions Trading Program." Working Paper 96-001. MIT-CEEPR, February.

Ellerman, A.D., R. Schmalensee, P.L. Joskow, J.P. Montero and E.M. Bailey (1997). *Emissions Trading Under the U.S. Acid Rain Program. Evaluation of Compliance Costs and Allowance Market Performance.* MIT Center for Energy and Environmental Policy Research.

Energy Information Administration (EIA, 1997). *The Effects of Title IV of the Clean Air Act Amendments of 1990 on Electric Utilities: An Update.* DOE/EIA-0582 (97). Washington, DC.

Environmental Protection Agency (EPA, 1997). *Acid Rain Program Homepage*. http://www.epa.gov/acidrain.

Greenberger, L.S. (1992). "Shopping for Acid Rain Control Strategies." *Public Utilities Fortnightly*. January 15.

Joskow, P., and R. Schmalensee (1998) "The Political Economics of Market-Based Environmental Policy." *Journal of Law and Economics,* april 1998. Also Working Paper 96-003. MIT CEEPR. March 1996.

Kindig, J.K., and R.L. Godfrey (1991). "Contorlling the Cost of Clean Air - A New Clean Coal Technology." *Public Utilities Fortnightly.* June 15.

Kling, C. and J. Rubin (1997). "Bankable Permits for the Control of Environmental Pollution." *Journal of Public Economics.* Vol.64, p.101-115.

Montero, J.P. (1997). "Volunteering for Market-Based Environmental Regulation: The Substitution Provision of the SO_2 Emissions Trading Program." Working Paper 97-001. MIT CEEPR. January.

Montgomery, W.D. (1972). "Markets in Licenses and Efficient Pollution Control Programs." *Journal of Economic Theory*. Vol.5, p.395-418.

Pindyck, R.S. (1993). "The Present Value Model of Rational Commodity Pricing." *The Economic Journal*. Vol. 103, p.511-530.

Rubin, J.D. (1996). "A Model of Intertemporal Emission Trading, Banking, and Borrowing." *Journal of Environmental Economics and Management.* Vol. 31, p.269-286.

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

Walz, M.A., D. Edeal and J.S. Snow (1995). "Mixing it up. Coal Blending in the 90's." *Coal.* April.

Williams, J.C., and B. Wright (1991). *Storage and Commodity Markets*. Cambridge: Cambridge University Press.

Appendix A: Definition of the symbols

P_t Undiscounted allowance price at time t (spot price).

e_t actual number of tons of SO_2 emissions at time t, after any abatement has taken place.

ε_t SO_2 emissions that would be needed to satisfy the demand for electricity from Title IV units at time t without any SO_2 abatement requirements (counterfactual emissions).

a_t tons of SO_2 abatement at time t; $a_t = \varepsilon_t - e_t$.

$c_t(a_t)$ cost of abating a_t tons of SO_2 at time t.

$m_t(a_t)$ marginal cost of abating another ton of SO_2 at time t, when a_t tons are already being abated. $m_t(a_t) = A_t + B_t a_t = A_t + B_t (\varepsilon_t - e_t) = B_t (\tilde{\varepsilon}_t - e_t)$

$\tilde{\varepsilon}_t$ $= \frac{A_t}{B_t} + \varepsilon_t$

Y_t total number of allowances with vintage time t issued to all affected units.

Y_H Y_t in Phase I.

Y_L Y_t in Phase II.

S_t stock of allowances (bank) available at the beginning of period t.

T length of Phase I.

τ time when the allowance bank runs out.

r riskless rate of interest.

μ asset-specific rate of interest.

ρ risk premium; $\mu = r + \rho$.

g growth in electricity demand ($\tilde{\varepsilon}_t = \tilde{\varepsilon}_0 + gt$).

Appendix B: Constant Electricity Demand, Determination of e_0 and τ

This section describes how to obtain the solution of equations (4) and (6) for τ and e_0 in the case of constant electricity demand and no technological changes ($\tilde{\varepsilon}_t = \tilde{\varepsilon}$, and B_t = B). The marginal cost is then given by $m_t(a_t) = B(\tilde{\varepsilon} - e_t)$ and equalization of the discounted marginal cost during the banking period yields

$$e_t = \tilde{\varepsilon} - e^{rt}(\tilde{\varepsilon} - e_0) \quad \text{for } t \leq \tau \qquad (30).$$

In this simple case, equation (6) becomes:

$$\int_0^{\tau} \left(\tilde{\varepsilon} - e^{rt}\left(\tilde{\varepsilon} - e_0\right)\right)dt = TY_H + (\tau - T)Y_L$$

$$\Rightarrow \quad \tau\tilde{\varepsilon} + \left(\tilde{\varepsilon} - e_0\right)\frac{\left(1 - e^{r\tau}\right)}{r} = TY_H + (\tau - T)Y_L, \qquad (31)$$

while equation (4) becomes:

$$\left(\tilde{\varepsilon} - e_0\right) = e^{-r\tau}\left(\tilde{\varepsilon} - Y_L\right) \qquad (32)$$

which we substitute into (31) to obtain:

$$\tilde{\varepsilon}\tau + \left(\tilde{\varepsilon} - Y_L\right)\left(e^{-r\tau} - 1\right)\frac{1}{r} = TY_H + (\tau - T)Y_L .$$

Rearranging yields:

$$\left(1 - e^{-r\tau}\right) = r\tau - rT\frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon} - Y_L\right)}. \qquad (33)$$

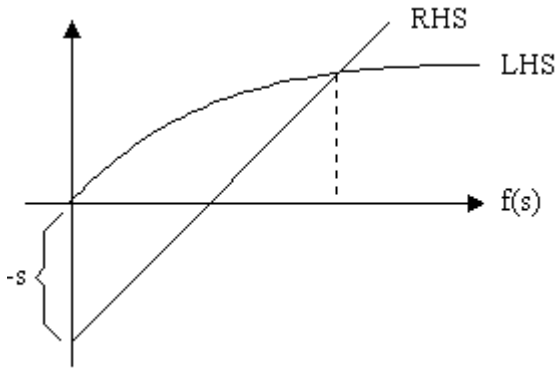
If we define

$$s = rT\frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon} - Y_L\right)}$$

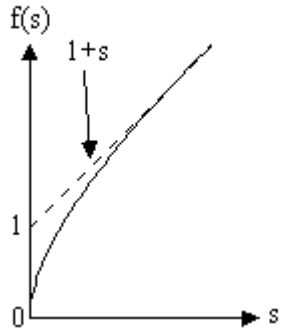
and let $r\tau = f$ we can write

$$\left(1 - e^{-f}\right) = f - s . \qquad (34)$$

Intuitively, the value of f as a function of s can be found by determining the point where the left hand side and the right hand side coincide:



Here is the graph of the function f(s) thus defined:



It is strictly increasing, intersects the origin and tends asymptotically to 1+s for large s.

The fact that f(s) is increasing in s follows from the fact that the slope of the right hand side of equation (34) is 1 while the slope of the left hand side is less or equal to 1 for $f \geq 0$. Using the function f(s) we have just defined, we can express τ as:

$$\tau = \frac{1}{r} f\left(r \frac{T\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon} - Y_L\right)} \right).$$

The value of e_0 can then be found by simply rearranging equation (32):

$$e_0 = \tilde{\varepsilon} - \left(\tilde{\varepsilon} - Y_L\right) e^{-r\tau}.$$

Appendix C: Constant Electricity Demand, Proof that $\frac{\partial e_t}{\partial \tau}$ is Increasing in t

During the banking period, we have:

$$e_t = \frac{rT\left(Y_H - Y_L\right)}{1 - r\tau - e^{-r\tau}}\left(e^{r(t-\tau)} - 1\right) + Y_L$$

so that

$$\frac{\partial e_t}{\partial \tau} = \frac{-rT\left(Y_H - Y_L\right)}{\left(1 - r\tau - e^{-r\tau}\right)^2}\left(-r + re^{-r\tau}\right)\left(e^{r(t-\tau)} - 1\right) + \frac{rT\left(Y_H - Y_L\right)}{1 - r\tau - e^{-r\tau}}(-r)e^{r(t-\tau)}$$

$$\frac{\partial^2 e_t}{\partial t \partial \tau} = \frac{-rT\left(Y_H - Y_L\right)}{\left(1 - r\tau - e^{-r\tau}\right)^2}\left(-r + re^{-r\tau}\right)re^{r(t-\tau)} + \frac{rT\left(Y_H - Y_L\right)}{1 - r\tau - e^{-r\tau}}\left(-r^2\right)e^{r(t-\tau)}$$

$$= \frac{r^3T\left(Y_H - Y_L\right)}{\left(1 - r\tau - e^{-r\tau}\right)}e^{r(t-\tau)}\left(\frac{\left(1 - e^{-r\tau}\right)}{1 - r\tau - e^{-r\tau}} - 1\right)$$

$$= \underbrace{\underbrace{\overbrace{\frac{r^3T\left(Y_H - Y_L\right)}{\left(1 - r\tau - e^{-r\tau}\right)}}^{\geq 0}}_{\leq 0}\underbrace{e^{r(t-\tau)}}_{\geq 0}\left(\underbrace{\overbrace{r\tau}^{\geq 0}}_{\leq 0} - \underbrace{1 - r\tau - e^{-r\tau}}_{\leq 0}\right)}_{\geq 0}$$

which shows that $\frac{\partial e_t}{\partial \tau}$ is increasing in t.

Appendix D: Non-Constant Electricity Demand, Effect of Changes in $\tilde{\varepsilon}_t$

We now derive the effect of an arbitrary variation in the path of $\tilde{\varepsilon}_t$. This includes any variation in the level of the marginal cost (A_t parameter) or in the electricity demand ε_t right now or at any point in the future. From equation (5), we have:

$$\int_0^\tau e_t dt - \left(T\left(Y_H - Y_L\right) + \tau Y_L\right) = 0 . \qquad (35)$$

We let $\tilde{\varepsilon}_t$ change to $\tilde{\varepsilon}_t + \Delta\tilde{\varepsilon}_t$ and let τ and e_0 adapt to this change so that (35) remains satisfied. This amounts to taking the total differential of (35) with respect to τ, e_0 and $\tilde{\varepsilon}_t$ at every time t:

$$\frac{\partial L}{\partial \tau}\Delta\tau + \frac{\partial L}{\partial e_0}\Delta e_0 + \int_0^\tau \frac{\partial L}{\partial \tilde{\varepsilon}_s}\Delta\tilde{\varepsilon}_s ds = 0$$

where L represents the left hand side of (35). Evaluations of each derivative yields:

$$\left(e_\tau - Y_L\right)\Delta\tau + \left(\int_0^\tau \frac{\partial e_t}{\partial e_0} dt\right)\Delta e_0 + \int_0^\tau \left(\int_0^\tau \frac{\partial e_t}{\partial \tilde{\varepsilon}_s} dt\right)\Delta\tilde{\varepsilon}_s ds = 0 .$$

Since $e_\tau = Y_L$ the first term vanishes and rearranging yields:

$$\Delta e_0 = \frac{-\int_0^\tau \int_0^\tau \frac{\partial e_t}{\partial \tilde{\varepsilon}_s}\Delta\tilde{\varepsilon}_s ds\, dt}{\int_0^\tau \frac{\partial e_t}{\partial e_0} dt} . \qquad (36)$$

In the case of linear marginal cost, $e_t = \tilde{\varepsilon}_t - \left(\tilde{\varepsilon}_0 - e_0\right)e^{rt}$ for $t \le \tau$ and we obtain:

$$\frac{\partial e_t}{\partial \varepsilon_s} = \delta(t-s) - \delta(s)e^{rt}$$

$$\frac{\partial e_t}{\partial e_0} = e^{rt}$$

where $\delta(.)$ is the Dirac delta (i.e., the point mass distribution). Thus,

$$\Delta e_0 = \frac{-\int_0^{\tau}\int_0^{\tau}\left(\delta(t-s)-\delta(s)e^{rt}\right)\Delta\tilde{\varepsilon}_s\,ds\,dt}{\int_0^{\tau}e^{rt}dt}$$

$$= \frac{-\int_0^{\tau}\left(\Delta\tilde{\varepsilon}_t-\Delta\tilde{\varepsilon}_0 e^{rt}\right)dt}{\int_0^{\tau}e^{rt}dt}$$

$$= \frac{\frac{\left(e^{r\tau}-1\right)}{r}\Delta\tilde{\varepsilon}_0-\int_0^{\tau}\Delta\tilde{\varepsilon}_t\,dt}{\frac{\left(e^{r\tau}-1\right)}{r}}$$

so that:

$$\Delta e_0 = \Delta\tilde{\varepsilon}_0 - \frac{r}{e^{r\tau}-1}\int_0^{\tau}\Delta\tilde{\varepsilon}_t\,dt\,,$$

which is equation (16) in the body of the paper.

Appendix E: Linear Growth in Electricity Demand, Determination of e_0 and τ

This section describes how the obtain the solution of equations (4) and (6) for τ and e_0 in the case of linear growth of parameter $\tilde{\varepsilon}_t$:

$$\tilde{\varepsilon}_t = \tilde{\varepsilon}_0 + gt\ .$$

Equation (4) becomes:

$$\tilde{\varepsilon}_0 + g\tau - \left(\tilde{\varepsilon}_0 - e_0\right)e^{r\tau} = Y_L \Rightarrow \left(\tilde{\varepsilon}_0 - e_0\right) = \left(\tilde{\varepsilon}_0 + g\tau - Y_L\right)e^{-r\tau},$$

which can be substituted into equation (6):

$$\int_0^\tau \left(\tilde{\varepsilon}_0 + gt - \left(\tilde{\varepsilon}_0 - e_0\right)e^{rt}\right)dt = TY_H + \left(\tau - T\right)Y_L$$

to yield:

$$\int_0^\tau \left(\tilde{\varepsilon}_0 + gt - \left(\tilde{\varepsilon}_0 + g\tau - Y_L\right)e^{r(t-\tau)}\right)dt = TY_H + (\tau - T)Y_L$$

$$\Rightarrow \tilde{\varepsilon}_0\tau + \frac{g\tau^2}{2} - \left(\tilde{\varepsilon}_0 + g\tau - Y_L\right)e^{-r\tau}\left(\frac{e^{r\tau} - 1}{r}\right) = TY_H + (\tau - T)Y_L\ .$$

$$\Rightarrow \left(1 - e^{-r\tau}\right) = \frac{r\tau\left(\frac{g}{2r\left(\tilde{\varepsilon}_0 - Y_L\right)}r\tau + 1\right) - rT\frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon}_0 - Y_L\right)}}{1 + \frac{g}{r\left(\tilde{\varepsilon}_0 - Y_L\right)}r\tau}$$

which can be written as:

$$\left(1 - e^{-h}\right) = \frac{h\left(\frac{1}{2}\gamma h + 1\right) - s}{1 + \gamma h} \qquad (37)$$

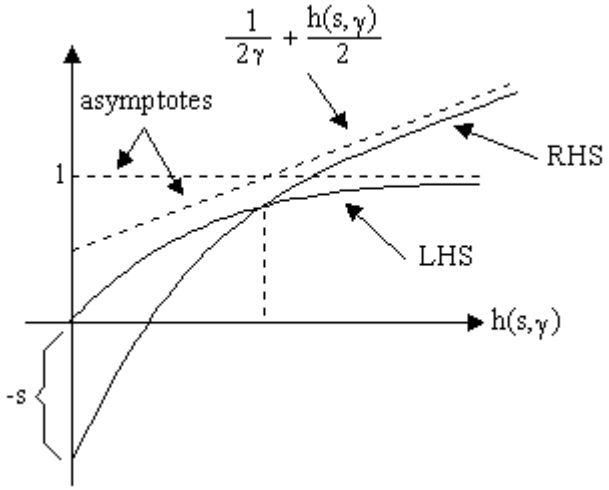
where

$$\left.\begin{aligned} h &= r\tau \\ s &= rT\frac{\left(Y_H - Y_L\right)}{\left(\tilde{\varepsilon}_0 - Y_L\right)} \\ \gamma &= \frac{g}{r\left(\tilde{\varepsilon}_0 - Y_L\right)} \end{aligned}\right\} \qquad (38)$$

Equation (37) implicitly defines a function h(s,γ) and the solution can expressed as:

$$\tau = \frac{1}{r} h(s,\gamma) \qquad (39)$$

Graphically, the solution of (37) can be obtained by finding at which value of h(s,γ) the left hand side and the right hand side of equation (37) intersect:



Note that h(s,0)=f(s), where f(s) is the function defined in the constant demand case. We now determine in which direction the value of the function h(s,γ) changes when its two arguments change. Taking the total differential of equation (37) with respect to s and g shows that h(s,γ) is increasing in s if and only if:

$$\gamma(r\tau - 1) + 1 \geq 0 \qquad (40)$$

In the case of the allowance market, one can verify that this condition is satisfied for realistic values of the parameters.[40] Hence, for all practical purposes, the result that τ

[40] Condition (40) can be written as:

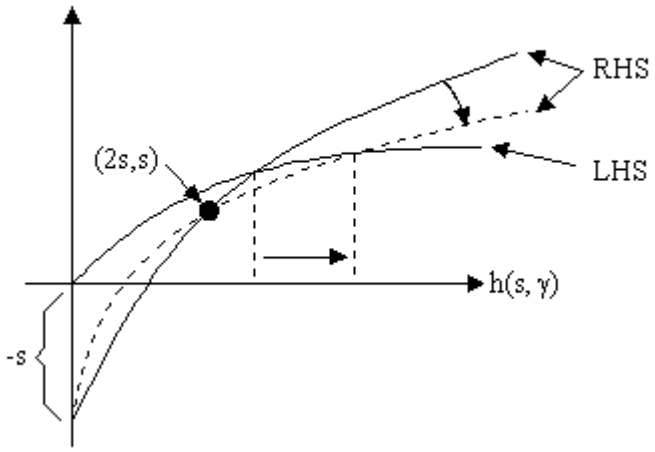
$$\gamma(r\tau - 1) + 1 \geq 0 \quad \Leftrightarrow \quad \frac{g(r\tau - 1)}{r\left(\tilde{\varepsilon}_0 - Y_L\right)} + 1 \geq 0 \quad \Leftrightarrow \quad \frac{\left(g/\tilde{\varepsilon}_0\right)(r\tau - 1)}{r\left(1 - Y_L/\tilde{\varepsilon}_0\right)} + 1 \geq 0$$

The worst case is obtained when rτ < 1 and when r and τ are as small as possible, while $g/\tilde{\varepsilon}_0$ and $Y_L/\tilde{\varepsilon}_0$ are as large as possible. Taking the following extreme values: r = 4%, τ = 5 years, $g/\tilde{\varepsilon}_0 \leq g/\varepsilon_0 \approx 2\%$ (from Ellerman and Montero (1996), Figure 1) and $Y_L/\tilde{\varepsilon}_0 \leq Y_L/\varepsilon_0 \approx 30\%$ (from Ellerman et al. (1997), p. 14), we obtain

$$\frac{(2\%)\left((4\%) \times (5 \text{ years}) - 1\right)}{(4\%)(1 - 30\%)} + 1 = 0.429 \geq 0.$$

increases with s, which we had obtained in the case of constant demand applies for the linear growth in demand case as well.

We now seek to determine when the function h(s,γ) is increasing in γ. It can be shown that condition (40) also guarantees that the right hand side of equation (37) is steeper than its left hand side and under that condition it is straightforward to determine the impact of changes in γ. Differentiation of the right hand side of (37) with respect to γ show that an increase in γ leaves the point (2s,s) unchanged while lowering the curve for h(s, γ) > 2s and increasing the curve for h(s, γ) < 2s, as illustrated below.



Hence, if the RHS is steeper than the LHS, the direction of the change in h(s, γ) only depends on whether the point (2s,s) lies above or below the curve representing the left hand side. We thus conclude that h(s, γ) is increasing in γ if and only if:

$s < 1 - e^{-2s}$ or, equivalently s<0.7968 (41)

(assuming that condition (40) always holds).

In summary, since we can reasonably assume that condition (40) holds, h(s,γ) is always increasing in s and is only increasing in γ if condition (41) holds.

Appendix F: Path of Price under Uncertainty

We now derive the solution to equation (24):

$$\min_{\{e_0,e_1,\ldots\}}\left\{E_0\left[\sum_{t=0}^{\infty}(1+\mu)^{-t}c_t(\varepsilon_t-e_t)\right]\right\} \qquad \text{subject to} \qquad \begin{cases}S_{t+1}=S_t+Y_t-e_t\\ S_{t+1}\geq 0\end{cases}$$

through dynamic programming. We define a valuation function

$$V_t(S_t)=\min_{e_t}\left\{F_t(S_t,e_t)\right\}$$

where $F_t(S_t,e_t)=c_t(\varepsilon_t-e_t)+(1+\mu)^{-1}E_t\left[V_{t+1}(S_{t+1})\right]+\lambda_t(e_t-S_t-Y_t)$.

The first-order condition is:

$$\frac{\partial F_t}{\partial e_t}(S_t,e_t)=0 \;\Rightarrow\; -m_t(\varepsilon_t-e_t)+(1+\mu)^{-1}E_t\left[\frac{\partial V_{t+1}(S_{t+1})}{\partial S_{t+1}}\underbrace{\frac{\partial S_{t+1}}{\partial e_t}}_{-1}\right]+\lambda_t=0$$

$$\Rightarrow\; E_t\left[\frac{\partial V_{t+1}(S_{t+1})}{\partial S_{t+1}}\right]=(1+\mu)\left(-m_t(\varepsilon_t-e_t)+\lambda_t\right) \qquad (42)$$

while the envelope condition is:

$$\frac{\partial V_{t+1}(S_{t+1})}{\partial S_{t+1}}=\frac{\partial F_{t+1}(S_{t+1},e_{t+1})}{\partial S_{t+1}}$$

$$=(1+\mu)^{-1}E_{t+1}\left[\frac{\partial V_{t+2}(S_{t+2})}{\partial S_{t+2}}\underbrace{\frac{\partial S_{t+2}}{\partial S_{t+1}}}_{1}\right]-\lambda_{t+1}.$$

Taking the expectation at time t of the envelope condition gives

$$E_t\left[\frac{\partial V_{t+1}(S_{t+1})}{\partial S_{t+1}}\right]=(1+\mu)^{-1}E_t\left[E_{t+1}\left[\frac{\partial V_{t+2}(S_{t+2})}{\partial S_{t+2}}\underbrace{\frac{\partial S_{t+2}}{\partial S_{t+1}}}_{1}\right]\right]-E_t\left[\lambda_{t+1}\right].$$

Equation (42) can then substituted in, which yields:

$$(1+\mu)\left(-m_t(\varepsilon_t-e_t)+\lambda_t\right)=(1+\mu)^{-1}E_t\left[(1+\mu)\left(-m_{t+1}(\varepsilon_{t+1}-e_{t+1})+\lambda_{t+1}\right)\right]-E_t\left[\lambda_{t+1}\right]$$

.

Rearranging gives

$$E_t\left[\left(m_{t+1}\left(\varepsilon_{t+1}-e_{t+1}\right)\right)\right]=\left(1+\mu\right)\left(m_t\left(\varepsilon_t-e_t\right)-\lambda_t\right). \qquad (43)$$