The Efficiency and Robustness of Allowance Banking in the U. S. Acid Rain Program

by

A. Denny Ellerman and Juan-Pablo Montero

05-005 WP **March 2005**

The Efficiency and Robustness of Allowance Banking in the U.S. Acid Rain Program

A. Denny Ellerman and Juan-Pablo Montero[*]

Abstract

This paper provides an empirical evaluation of the efficiency of allowance banking (i.e., abating more in early periods in order to abate less in later periods) in the nationwide market for sulfur dioxide (SO_2) emission allowances that was created by the U.S. Acid Rain Program. We develop a model of efficient banking, select appropriate parameter values, and evaluate the efficiency of observed temporal pattern of abatement based on aggregate data from the first eight years of the Acid Rain Program. Contrary to the general opinion that banking in this program has been excessive, we find that it has been reasonably efficient. We also show that this optimal banking program is robust to the errors in expectation that characterized the early years of this program; however, this property is due to design features that are unique to the U.S. Acid Rain Program.

---

[*] Ellerman <ellerman@mit.edu> is with the Center for Energy and Environmental Policy Research (CEEPR) and the Sloan School of Management, Massachusetts Institute of Technology (MIT). Montero < jmontero@faceapuc.cl > is with the Department of Economics, Catholic University of Chile, and he was a visiting professor at the Sloan School of Management during the writing of this paper. We thank Paul Joskow, Matti Liski, Robert Stavins, Robert Pindyck, Dick Schmalensee, Stephen Smith, Martin Weitzman and seminar participants at Harvard, MIT and University College London for useful comments; Brice Tariel for research assistance; and CEEPR and the U.S. EPA (STAR grant award #R-82863001-0) for financial support. All correspondence should be sent to Ellerman at: MIT/CEEPR (E40-435); 77 Massachusetts Avenue; Cambridge, MA 02139 USA.

1. Introduction

Emissions trading usually refers to trades across space in the same period of time, but it can also refer to trades through time, typically by banking, which implies being able to carry over unused allowances from one period for use in later periods and which can change the temporal pattern of abatement.[1] Over the past decade, this latter dimension of emissions trading has drawn increasing attention in the literature and current proposals to decrease emission caps in the United States suggest a continuing role for banking in the future.[2] Several authors have studied the theoretical properties of intertemporal trading,[3] but no work has yet evaluated how firms have actually responded to this feature of tradable permits programs. The U.S. Acid Rain Program (Title IV of the 1990 Clean Air Act Amendments) provides a unique opportunity to do so since it allows banking (but not borrowing) and it is by far the most significant experiment in emissions trading to date.

By every measure, banking has been a major form of emissions trading in the U.S. Acid Rain Program. During the first five years of the program constituting Phase I, 1995-99, 11.65 million allowances, or 30% of all the allowances distributed in these years, were banked. Equivalently, the reduction in emissions during Phase I was about twice what was required to meet the Phase I cap.[4] Then, 3.7 million, or about a little more than a third, of these banked allowances were used to cover SO_2 emissions in

---

[1] Logically, borrowing could also be included, but it is usually not.

[2] For example, President Bush's Clear Skies Initiative and the recently proposed Clean Air Interstate Rule would reduce the existing U.S. SO_2 emissions cap by two-thirds in two steps starting in 2010 and it would allow unused allowances from the current program to be carried over for use under the lower caps. Also, proposals for reducing greenhouse gas emissions, such as the initially proposed McCain-Lieberman bill (S. 142), can be expected to feature emission caps that become more stringent over time and include banking.

[3] See Rubin (1996) and Cronshaw and Kruse (1996) for general formulations; Schennach (2000) for a formulation specific to the U.S. Acid Rain Program; and Rubin and Kling (1993) for a simulation of a potential banking program for hydrocarbon emission standards imposed on light-duty vehicle manufacturers.

[4] A reasonable estimate of the level of emissions without the program from all the units receiving allowances during the five years of Phase I is 48 million tons. The difference between this estimate and observed emissions from these same units over the same five years is about 21 million tons, or a little more than twice the 10 million tons of abatement required to achieve the cumulative cap of 38 million tons (Ellerman, 2004).

excess of the number of new vintage allowances issued for use in 2000-02 under the much tighter Phase II SO_2 emission cap.[5]

The occurrence of banking in the Acid Rain Program has not been a surprise. Given the two-phased structure of the program and the opportunity to bank, all analysts expected some amount of banking; however, there was little agreement on the likely size of the bank. Early estimates varied by a factor of five: from two to ten million tons. As Phase I began, one consulting firm created a small sensation by predicting a bank as large as 15 million tons.[6]

Enough years have passed that an evaluation of the temporal efficiency of this aspect of emissions trading can be made. The accumulation phase of the banking period is over, the size of the end-of-Phase-I bank is known, and the rate of draw down in the first three years of Phase II can be observed.[7] In contrast to the earlier papers on banking, which developed theoretical properties or simulated what might occur in a particular program, this paper looks at aggregate behavior in an actual program over a period of time that spans about half of the entire banking period and assesses whether observed behavior is efficient.[8]

---

[5] Phase II, beginning in the year 2000, differs from Phase I in both the stringency and the scope of the required emission reductions. In Phase I, units larger than 100 MW^e capacity and with 1985 emission rates of 2.5 #/mmBtu or higher were required to be subject to the SO_2 cap and to reduce emissions to an average emission rate equal to 2.5 lbs. SO_2 per mmBtu of heat input (#/mmBtu) times average 1985-87 (baseline) heat input. In Phase II, all fossil-fired generating units greater than 25 MW^e were subject to SO_2 caps, regardless of historical emission rates, and they were required to reduce emissions, absent any trading, to an amount that is less than half the Phase I rate: 1.2 #/mmBtu times baseline heat input. Units with a 1985 emission rate less than 1.2 #/mmBtu received allowances equal to baseline heat input times the 1985 emission rate.

[6] A report from the General Accounting Office published in December 1994 (USGAO, 1994) projected a Phase I bank of two million tons. An earlier and more thorough analysis by EPRI published in August 1993 (EPRI, 1993) predicted a bank "between 5 and 10 million tons, with our current projections at the higher end of the range." RDI, a coal and electric utility consulting firm, forecast a 15 million ton bank in mid-1995 as the first emission monitoring reports became available (RDI, 1995). A later EPRI report (EPRI, 1997) written with the benefit of the 1995 compliance data stated: "The bank size by 2000 is surprisingly uncertain—from 10 to 15 million short tons."

[7] We use the term "banking period" to denote the entire multi-year period over which banked, unused allowances are held consisting of an accumulation phase and a draw-down phase. In the U.S. Acid Rain Program, these phases correspond to the entirety of Phase I and some years into Phase II (often estimated to be ten years).

[8] Since a permits bank is similar to a non-renewable resource, this paper also adds to the extensive literature on the empirical validity of the Hotelling's rule to predict price and extraction paths (e.g., Farrow, 1985). In this regard, our paper provides a simpler test because we do not need to deal with extraction costs (and how they change as the resource is exhausted) and uncertainty regarding the size of the resource.

Contrary to the common perception of excessive banking during Phase I,[9] we find that the evolution of the SO_2 allowance bank has been reasonably efficient. We argue that this misperception has been due to 1) a failure to understand how an initial error in expectation concerning counterfactual emissions[10] affects banking behavior and 2) an assumption of little ability to adapt to the significantly lower than expected allowance prices at the beginning of Phase I. While efficient banking is found to be robust to changes of expectation concerning counterfactual emissions, we also show that this property cannot be generalized for all banking programs because of design features that are unique in U.S. Acid Rain Program.

In this paper, we limit attention to the years through 2002 and purposefully ignore the three-fold increase in SO_2 allowance prices that has occurred during 2004 in response to the proposed implementation of the Clean Air Interstate Rule, which would effectively lower the SO_2 cap by two-thirds in two steps beginning in 2010. Although this price behavior provides further evidence of the effect of banking provisions on allowance prices, it is too early to evaluate the efficiency of what is still not a completely implemented administrative action. Accordingly, we restrict our evaluation of the efficiency of banking in the U.S. Acid Rain Program to the years before further, phased-in reductions of the SO_2 cap, and therefore of new banking programs, became a real prospect.[11]

The rest of the paper is organized as follows. Section 2 presents the model of banking that is used to generate efficient temporal paths of abatement and marginal cost for comparison with observed banking behavior. Section 3 discusses the key assumptions underlying any banking program—program-specific parameters and assumptions about

---

[9] At least, that was our perception before writing this paper (Ellerman et al., 2000; Smith et al., 1998).

[10] Counterfactual emissions are what would have occurred in the absence of the program: what is commonly called business-as-usual (BAU) emissions in modeling exercises. Any discussion of the efficiency of abatement, marginal cost, or allowance price implies some level of counterfactual emissions since price and marginal cost reflect the quantity of abatement and abatement is the difference between what is observed and some estimate of what would occurr but for the particular policy measure. All subsequent references in this paper to "counterfactual" should be understood to mean counterfactual emissions.

[11] Ellerman (2002) applies the banking model developed in this paper to the Bush Administration's Clear Skies Proposal of February 2002, which would have implemented a similar lowering of the SO_2 cap by legislation, to analyze the consequences of enactment for allowance prices and banking behavior.

counterfactual emissions, the cost function, and the discount rate—and it provides estimates of the appropriate values for this program. Section 4 compares observed SO_2 allowance banking behavior with simulated efficient paths and draws inferences from that comparison. Section 5 explains the error in the earlier, too facile assumption of excessive banking and the reasons for the robustness of efficient banking in the face of significant errors in expectation concerning counterfactual emissions. Section 6 concludes.

2. A Model of Efficient Banking

The theory of permits banking follows directly from the theory of nonrenewable resources pioneered by Hotelling (1931). Because the cost of creating the permits that constitute the cap is zero, a banking model with no uncertainty and perfect competition would predict that during the banking period the price $P(t)$ of permits will rise at the risk-free rate of interest r, $\dot{P}(t)/P(t) = r$, where a dot denotes a time derivative. In practice, however, firms will not know with certainty the number of permits they will demand in the near and distant future, and consequently, the market equilibrium price of permits becomes an uncertain variable.[12]

When the price and demand for permits are random variables, investments in abatement and holding permits are no longer risk-free activities, and affected firms can be expected to choose an abatement path that minimizes their expected present value of compliance costs using an appropriate risk-adjusted discount rate ρ. Risk-adverse agents will diversify this risk by holding a portfolio of assets including permits.

Because modeling the efficient path of the SO_2 allowance bank is analogous to modeling the efficient extraction path of an exhaustible resource sold in a competitive market under conditions of uncertainty, we follow the approach put forward by Slade and Thille (1997), who combined the Hotelling model for pricing exhaustible resources with the capital asset pricing model (CAPM) for risky assets. Accordingly, the evolution of allowance prices during the banking period is governed by the arbitrage condition

---

[12] In the case of the SO_2 program, firms are never certain about future electricity demand and the future prices for fuels of differing sulfur content, especially that between coal and natural gas.

$$\frac{\frac{1}{dt}E_t dP(t)}{P(t)} = r + \beta(r^m - r) \equiv \rho \qquad (1)$$

where E_t is the expected value operator, r^m is the expected rate of return on a well diversified market portfolio and β, a common financial variable denoting the asset-specific risk premium, is the ratio of the covariance of ρ and r^m to the variance of r^m, that is $\beta = \sigma_{\rho m} / \sigma_m^2$. Note that both r and r^m can change overtime.

In a continuous setting efficient banking also requires instant cost minimization, i.e., at each point in time firms equalize their marginal abatement costs to the current market price. Assuming that there is a sufficiently large number of individual firms so that the aggregate abatement cost function is strictly convex, continuous and twice differentiable, the arbitrage condition can be rewritten as a function of the aggregate marginal abatement costs, $C'(q(t))$, as follows

$$\frac{\frac{1}{dt}E_t C'(q(t))}{C'(q(t))} = \rho \qquad (2)$$

where $q(t)$ is the total amount of abatement at time t.

A further condition that must hold for the entire banking period is that the cumulative number of allowances issued equal cumulative emissions. Therefore, at any time t during the banking period, the time τ at which the bank is expected to end must satisfy this condition, known as the exhaustion condition in exhaustible resource markets.[13]

---

[13] At the end of the banking period and thereafter, agents owning affected units can be expected to maintain some allowance carry-over to deal with the uncertainty about the demand for allowances in any given year or to take advantage of differences between current prices and those expected in the next year, but these amounts will be small by comparison with the carry-over during the initial banking period. Numerical exercises that add a small negative term on the left hand side of equation 3 indicate a minor effect on the efficient banking path. For example, a large carry over of as much as 20% (1.8 million allowances) would increase the size of the bank at the end of Phase I by less than 2%. Accordingly, we ignore this likely post-τ inventory phenomenon in the rest of the paper.

$$B(t)+\int_{t}^{\tau} a(t)dt = E_t\left\{\int_{t}^{\tau} u(t)dt-\int_{t}^{\tau} q(t)dt\right\} \tag{3}$$

where $B(t) \geq 0$ is the size of the allowance bank at t,[14] $a(t)$ is the number of allowances allocated at t that is specified by the legislation, and $u(t)$ represents counterfactual emissions. Observed emissions, that is, emissions constrained under the program, are $u(t) - q(t)$.

In addition, a terminal condition must hold at τ. At that point in time and thereafter, multi-year trading caused by the phased-in, discontinuous cap ceases and the predominant form of emissions trading will be spatial trading. Emissions must be equal to the cap, $a(t)$, for each period of time thereafter; and abatement, which will determine the marginal cost of abatement and the price of allowances, will be equal to the shortfall of allowances from counterfactual emissions. Thus, at any time t during the banking period, the time τ at which the bank is expected to end must also satisfy the terminal condition

$$a(\tau) = E_t\{u(\tau)-q(\tau)\} \tag{4}$$

Equations (2), (3) and (4) together with assumptions about both counterfactual emissions, $u(t)$, and the functional form of the aggregate abatement cost function, $C(q)$, can be used to solve numerically for τ and to derive explicit efficient abatement and banking paths during the banking period. Counterfactual emissions are modeled as emissions at $t = 0$ increasing at some rate, $g(t)$, that is, $u(t) = \varepsilon\, u_0\, e^{G(t)}$, where ε represents the share of total uncontrolled emissions subject to the U.S. Acid Rain Program, and $G(t) = \int_0^t g(s)\, ds$. The value of ε will equal 1.0 in Phase II, when all generating sources are included, and it will be less than 1 (0.57) in the transitional Phase I, when only a subset of sources are included. Since some amount of annual variation around the central expectation for counterfactual emissions can be expected, agents are assumed to adopt a single value of $g(t)$ reflecting the mean expectation at t for the remainder of the banking period.

---

[14] Initially, $B(0) = 0$ but it would be expected to assume positive values during the banking period.

Aggregate marginal abatement costs are assumed to depend on aggregate abatement in the following form

$$C'(q(t)) = \alpha_i [q(t)]^{\gamma} \tag{5}$$

The scaling parameter, α_i, takes the subscript 1 during Phase I and the subscript 2 thereafter. Two time-differentiated cost functions exist because Phase II expands the scope of the Acid Rain Program to include additional generating units and abatement opportunities.[15] The exponent, γ, reflects the curvature of the relationship and it is assumed to be the same for both the Phase I and Phase II aggregate cost functions and to remain unchanged during the banking period.

Given the functional form in (5), conditions (2) and (4) can be combined to obtain the expected amount of abatement in each period $q(t)$ as function of abatement at τ

$$q(t) = \begin{cases} q(\tau)\left(\dfrac{\alpha_2}{\alpha_1}\right)^{1/\gamma} e^{-\rho(\tau-t)/\gamma} & if \quad 0 \le t \le T \\ q(\tau)e^{-\rho(\tau-t)/\gamma} & if \quad T \le t \le \tau \end{cases} \tag{6}$$

where T denotes the end of Phase I and $q(\tau)$ is given by (4). Substituting (6) into (3) yields a single equation that, when fully specified with parameter values, can be expanded to solve for the end of the banking period, τ, and to derive efficient banking paths as is done below for the beginning of the trading program ($t = 0$) using g_0 to denote the expected growth rate of counterfactual emissions at $t = 0$.

---

[15] While firms that are not affected until 2000 do not abate, they are present in the market and able to accumulate a bank for use in Phase II by purchasing allowances during Phase I.

$$a_1 T + a_2\left(\tau - T\right) = \varepsilon u_0\left(\frac{e^{g_0 T} - 1}{g}\right) + u_0\left[\frac{e^{g_0 \tau} - e^{g_0 T}}{g}\right] -$$

$$\left(u_0 e^{g_0 \tau} - a_2\right) e^{\frac{-\rho\tau}{\gamma}}\left[\left(\frac{\alpha_2}{\alpha_1}\right)^{\frac{1}{\gamma}}\left(\frac{e^{\frac{\rho T}{\gamma}} - 1}{\rho/\gamma}\right) + \left(\frac{e^{\frac{\rho\tau}{\gamma}} - e^{\frac{\rho T}{\gamma}}}{\rho/\gamma}\right)\right] \qquad (7)$$

The two terms on the left-hand-side of (7) state the number of allowances available in Phase I and during the years of Phase II constituting the draw-down part of the banking period. The first two terms on the right-hand-side give cumulative counterfactual emissions for units affected during Phase I and for all units during Phase II through the end of the banking period. The third term on the right-hand side states the cumulative emission reductions over the entire banking period. The term outside the brackets is the amount of abatement required at $t = \tau$ discounted to t = 0 to express the amount of abatement that would occur at t = 0. The two terms within the brackets are indices of cumulative abatement, normalized to $q(0)$, for the Phase I units and for all units during the Phase II part of the banking period, respectively. If the two sets of covered units were identical, that is, if $\alpha_2/\alpha_1 = 1$, the two terms would collapse into one.

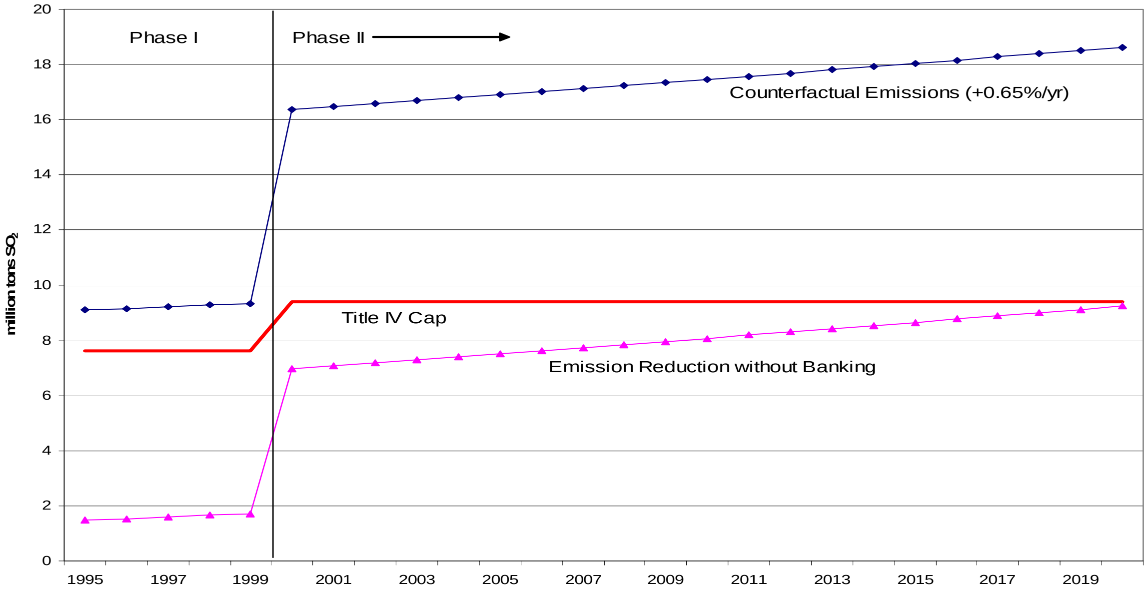
Equation (7) implicitly defines the efficient banking program at t = 0, when B (0) = 0, on the assumption that parameter values will not change during the banking period; however, it can be easily adapted to incorporate subsequent changes in agents' expectations or in other parameter values during the banking period. The observed banking path will then reflect segments of differing efficient paths each reflecting successive starting points, the accumulated bank as of each starting point, and the changed parameter values.

3. Parameter Values for the U.S. Acid Rain Program

3.1. Allowances and Counterfactual Emissions

The allowance cap and the assumptions concerning counterfactual emissions define what would be the annual required reduction of emissions absent any banking, as depicted in Figure 1. The annual allowance caps are specified in the legislation and implementing regulations. The average annual, aggregate allowance cap for Phase I, a_1, is 7.62 million allowances, the cap for Phase II, a_2, is 9.39 million allowances, and the transition from Phase I to Phase II, T, occurs at $t = 5$.[16]

Figure 1. Title IV Caps and Counterfactual Emissions



---

[16] A total of 38.09 allowances were distributed for the five Phase I years and the annual cap cited here is this cumulative sum divided by five. In fact, more allowances were allocated in 1995 and 1996 than in 1997-99; however, the distribution of the total five-year amount among years in Phase I is without importance from the standpoint of an efficient banking program at reasonable discount rates because of the short duration of Phase I. The Phase II cap is an average for the period 2000-09 including the following components: 8.9 million allowances from the basic allocation distributed annually to units and through the EPA auction, 0.10 million allowances to ¶ 410 opt-in units, and an assumed annual average of .39 million bonus and extra allowances over this period. Various bonus and extra allowances amounted to .96, .56, and .53 million allowances in 2000, 2001, and 2002, respectively, but they will be fewer in subsequent years.

Initial counterfactual emissions can be estimated with considerable accuracy. For the generating units first affected in Phase I and remaining so since then, we assume that initial counterfactual emissions, $u_1(0)$, are 9.07 million tons, as determined by a simple technique for calculating the counterfactual.[17] The initial counterfactual for the much larger universe of units affected in 2000, $u_2(0)$, is the sum of the counterfactual for the Phase I units and observed 1995 emissions of 6.72 million tons for the units first affected in 2000, or a total of 15.79 million tons.

Estimation of the expected growth rate in counterfactual emissions, g, is subject to much greater uncertainty. Both EPRI and EPA's contractor, ICF, conducted careful early studies for the purpose of analyzing the effect and cost of the Acid Rain Program and they contained estimates, as of the early 1990s, of what emissions without the program were thought likely to be (EPRI, 1993; ICF, 1989). Although growth in the demand for electricity from fossil-fuel-fired generating units was expected to be between 1.5% and 2.5% per annum, expectations for counterfactual SO_2 emissions varied greatly depending on assumptions about the retirement of coal fired units, the utilization of nuclear capacity, and the economic competitiveness of new gas-fired generating units. High emissions scenarios predicted SO_2 emissions growing at an annual rate of about 1.25% per annum through 2010, while the low emissions cases predicted either constant emissions after 2000 (EPRI, 1993) or emissions that are declining at rates between 0.5% and 1% per annum (ICF, 1989).

For the simulations in this paper, several values for expected growth in counterfactual emissions are used. For the initial expected banking program, we use a value of $g = 0.65\%$, drawn from an early-1990s EPA forecast (Pechan, 1995), to provide a single value representing expectations at the beginning of Phase I. This estimate is approximately half way between the high and low emission scenarios in the early EPRI

---

[17] A simple method for computing a counterfactual assumes that, in the absence of the SO_2 program, emission *rates* would have remained unchanged at the values observed in 1993 for Phase I units, and that total emissions would vary according to observed changes in heat input. As discussed more extensively in Ellerman et al. (2000) and especially in the appendix by Schennach, this simple counterfactual closely tracks aggregate emissions as estimated by econometric techniques that take trends in emission rates into account.

and ICF analyses of expected emissions absent the SO_2 cap. The low and high emission scenarios are represented by $g = 0\%$ and $g = 1.25\%$, respectively.[18]

3.2. The Cost Function

Two parameter values define the cost functions: the convexity parameter, γ, and the scaling parameters of the cost functions for the Phase I units, α_1, and for all units, α_2. In (7), only the ratio of the scaling parameters is needed and it can be easily shown using (6) that $(\alpha_2/\alpha_1)^{1/\gamma} = q_1/q_2$, or the ratio of abatement by Phase I units in Phase 2 to that by all units in the same year. This ratio can be estimated based on observed data for 2000, 2001, and 2002; and its value is 0.83.[19]

The convexity parameter indicates the rate at which marginal cost rises with the quantity of abatement, and values for this parameter can be inferred from several studies. The early EPRI study of abatement costs contains several charts of this relationship, and it is linear over the relevant range.[20] Ongoing analysis by the authors and colleagues at MIT concerning the cost of reducing SO_2 emissions from the 2000 levels by retrofitting scrubbers to presently unscrubbed coal-fired units indicates a similar relationship. Accordingly, we assume a linear relationship between the quantity and marginal cost of abatement, $\gamma = 1.0$.

With these assumptions, the marginal cost of the annual reduction required by the SO_2 cap without banking can be calculated as is shown by the solid line in Figure 2.[21] This is the price dual of the quantity path given in Figure 1 and the opportunity to reduce

---

[18] For comparison, the actual average annual increase in heat input (2.4% between 1993 and 2002), a reasonable proxy for electricity output, has been within but near the upper end of the expected range (1.5% to 2.5%) for growth in output, but the growth rate in estimated counterfactual emissions over the same period, 1.63%, has been greater than the 1.25% rate that was seen as the high end of the likely range for emissions growth. If our estimate of counterfactual emissions is accurate, the latter phenomenon would have led eventually to the change in expectation and in abatement quantity, price, and banking behavior that is discussed later in this paper.

[19] This ratio varies between 0.81 and 0.84 over these years. The effect of this variation on the total amount of allowances banked at the end of Phase I is less than 2%.

[20] EPRI, 1993; Figures 5-4 and 6-15.

[21] In making the estimates of optimal price in Figure 2, the scaling parameter, α_2, is assigned a value of 26 that causes the simulated 2001 prices to approximate observed 2001 prices. Average monthly prices during 2001 ranged from a low of \$159 in January to a high of \$208 in August and the average monthly price for the year was \$185. The price paths in Figure 2 will be shifted up or down to the extent that the true value of α_2 is higher or lower than 26, or that 2001 prices were below or above true equilibrium prices.

cost or increase profit by a temporal redistribution of abatement effort is immediately evident.

Figure 2. Expected Optimal Allowance Price Paths (Constant 2001$)

Since prices can be expected to be equal to marginal costs at all times during the banking period, equation (2) implies that the actual price path with banking will depend on the discount rate, which would be a real rate since the marginal abatement cost function is stated in today's dollars. The dashed lines in Figure 2 show the price paths for real discount rates of 3%, 6%, and 9%. The end of the banking period, τ, differs for each discount rate and higher discount rates are associated with shorter banking periods, lower initial prices, and greater increases in marginal abatement cost during the banking period. Once a banking program has ended, prices would cease rising at the discount rate and increase instead at a uniform lower rate characterizing the post-banking period.[22]

3.3. The Risk-Adjusted Discount Rate

SO_2 allowances are financial assets that are readily tradable and can be turned into cash. As such, holding allowances implies foregoing the return that could be earned

---

[22] This rate, which is about 1.3% per annum in Figure 2, would depend upon the rate of increase in counterfactual emissions (0.65%) and the elasticity of the marginal abatement cost curve (1.0). Some inventory might be carried from year to year and actual prices might fluctuate from this post-τ path,

during the holding period by investing the cash in other financial assets having similar risk characteristics. The relevant criterion for determining this return is the degree of undiversifiable risk associated with holding SO_2 allowances, indicated by the beta (β) coefficient in equation (1). With over nine years (mid-1994 through 2003) of SO_2 allowance price data now available, the monthly returns from holding SO_2 allowances can be readily calculated and correlated with the returns from holding a broadly diversified portfolio of equities.[23]

Table 1 provides estimates of β when the monthly returns from holding SO_2 allowances are regressed by ordinary least squares on the monthly returns from holding the S&P 500 or NASDAQ indices over the same period.

Table 1: Estimation of Beta for SO_2 Allowances
(August 1994-December 2003)

Market Index	Beta coefficient	Standard error
S&P 500	0.1725	0.1398
NASDAQ	0.0890	0.0756

In both cases β is not significantly different from zero and the same result occurs when the same regression is made over shorter periods, for instance, leaving out the earlier observations when it could be argued that markets were not as well formed, or even for periods as short as two years (24 observations). The use of robust variance estimators and corrections for serial correlation do not change the result. Since the results shown in Table 1 indicate no evident *un*diversifiable risk associated with holding SO_2 allowances, we use the risk-free rate as a reasonable approximation of the appropriate discount rate for SO_2 allowance banking.

This result, which may strike readers as surprising, as it did us, is critical to the analysis that follows, and indeed to any analysis of the extent to which banking, or any

---

reflecting year-to-year variations in demand, but the average annual increase in price would be less than the discount rate.

[23] Although data for 2004 are now available, they are not included because of the tripling of SO_2 allowance prices during 2004 in response to the proposed implementation of the Clean Air Interstate Rule. Inclusion of the 2004 data would not indicate any greater correlation of the returns from SO_2 allowances and market indices since the returns on a diversified portfolio of U.S. equities did not increase notably in 2004.

other form of temporal flexibility in emissions trading, is efficient. Two considerations specific to SO_2 allowances help to explain this result. First, the asset beta associated with producing electricity, the joint product of the SO_2 emissions covered by the U.S. Acid Rain Program, is very low. Equity betas, which take debt leverage into account, are typically around 0.5 for regulated electric utilities and 1.0 for the more highly leveraged, unregulated producers of electricity. When the observed equity betas for these two types of electricity producers are adjusted to account for the risk associated with varying debt levels, the resulting asset betas are similar, 0.2, which is low, although not zero. Since a value of +0.20 is also not rejected by the regression results reported in Table 1, a low positive valued beta is also possible, if there is reason to believe that the undiversifiable risk associated with holding SO_2 allowances is the same as that of producing electricity.

Second, and perhaps more importantly, the factors determining allowance prices are considerably different from those determining profits from generating electricity, not to mention the profits of the corporate sector as a whole. The profits of electricity producers will be influenced mostly by the price of electricity, the cost of fuel, regulatory treatment, and the growth in demand for electricity. The first three factors will have little direct influence on allowance prices, and the growth in the demand for electricity is, as explained in last section, only one of several factors determining counterfactual emissions. Factors that are at least as important in determining SO_2 emissions, and therefore the demand for allowances, are 1) the *relative* prices of fuels of differing sulfur content, in particular coal and natural gas prices, and 2) non-Title-IV regulatory requirements affecting SO_2 emissions. While these factors have some effect on the profits of the owners of electricity generating assets, the effect on equity returns for the market in equities as a whole is negligible. From this point of view, it is not so surprising that the returns from holding SO_2 allowances are uncorrelated, or at best weakly correlated, with market returns and that the *beta* for SO_2 allowances is zero or close to it. Depending on this value and that of the equity premium [$r^m - r$ in (1)], a positive risk premium for SO_2 allowances would add at most a half percentage point to a much larger risk-free rate. Keeping this in mind, we turn our attention now to this primary determinant of the discount rate for SO_2 allowance banking.

Treasury notes provide the standard for determining risk-free rates of return in the U.S. economy. Since a real rate is appropriate for this analysis, inflation-indexed Treasury notes are an obvious source; however, these notes have been offered only since the beginning of 1997 and in limited maturities, typically ten and thirty years. For the years prior to 1997, and especially for the years 1993-1995 when the owners of affected facilities were developing and deciding compliance strategies, real risk-free rates must be inferred from nominal rates less assumed inflationary expectations.

The appropriate horizon for the risk-free rate applicable to allowance banking is not obvious. Although initially the end of the banking period would have been reasonably expected to be around 2010, most of the banked allowances would be held for a shorter time. The holding period for a banked 1995 vintage allowance would be at least five years and perhaps more than fifteen, but the expected holding period would diminish with each succeeding vintage of Phase I allowances. Moreover, the holding period could be expected to be lengthened or shortened as the discount rate or expectations change. We use a ten-year maturity, partly because of the availability of a ten-year, real, risk-free rate, but also because it appears to be a reasonable average holding period for allowances banked during Phase I.

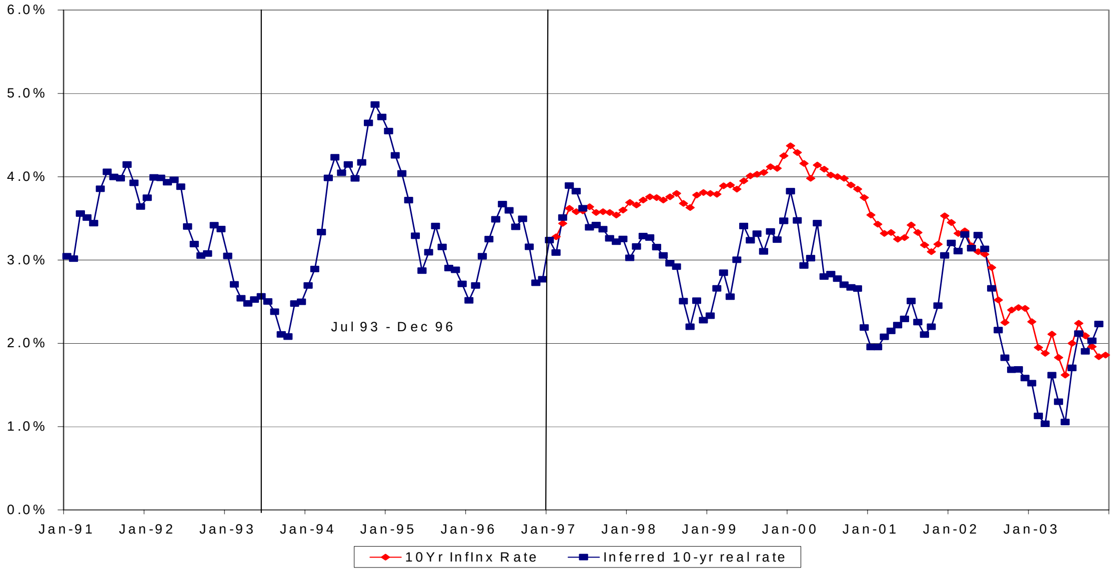
Figure 3 compares an observed 10-year real risk-free rate calculated from 10-year Treasury Inflation Protected Securities (TIPS) from 1997 through 2003 and an inferred real rate from 1992 through 2003 based on a nominal constant 10-year maturity rate less an adjustment for inflationary expectations.[24] It is evident from the period when the two overlap that the inferred real rate fluctuates more than the actual real rate. Also, the generally higher level of the actual real rate suggests that the inflationary adjustment used in this paper has generally been greater than warranted. For the period prior to 1997, when initial banking decisions would have made, the inferred rate fluctuates between 2.0% and 5.0%. Based on this data, we assume a rate of about 3.5% as an appropriate real

---

[24] The TIPS rate is constructed by taking the monthly average return for each successive ten year note so that the rate in 1997 reflects the 2007 note, that for 1998 reflects the 2008 note, etc. The Federal Reserve provides average monthly constant maturity 10-year nominal rates. These rates are adjusted for inflation by subtracting the average of the trailing 1-year and 10-year inflation rates in the all-item consumer price index for urban consumers. The use of this average assumes that inflationary expectations reflect equally the recent past and the longer term experience.

risk-free rate for the years 1993-96 when initial banking decisions were decided and implemented.

Figure 3: Actual and Inferred Real, Risk-Free Interest Rates, 1991-2003

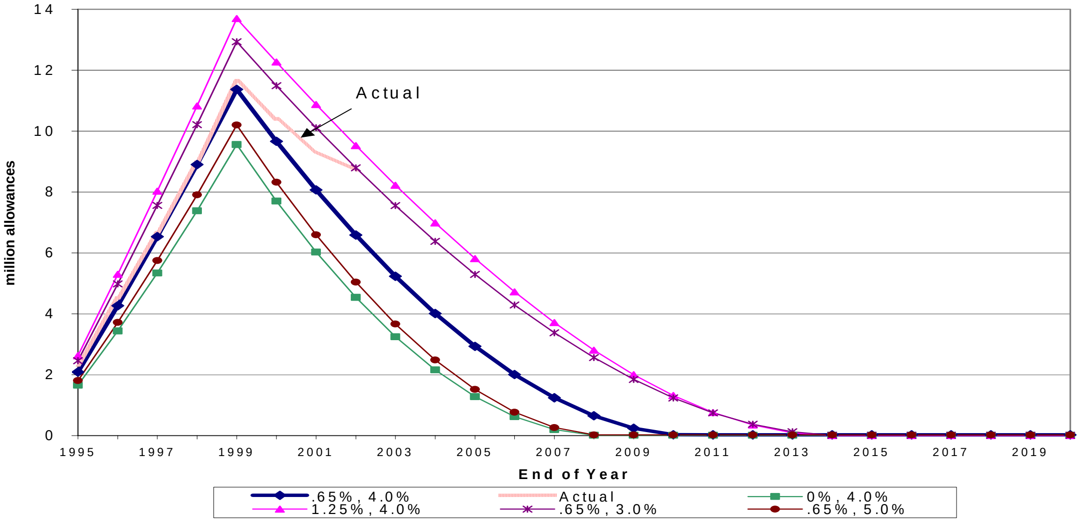


Source : Federal Reserve Bank of St. Louis, Historical series for Inflation-indexed Treasury Securities, http://www.research.stlouisfed.org/fred2/categories/82.

4. Is the Evolution of the SO_2 Bank Efficient?

One way to evaluate the efficiency of temporal trading under the U.S. Acid Rain Program is to simulate alternative banking programs as agents might have envisioned them in the early to mid-1990s when strategies to comply with the U.S. Acid Rain Program were being formulated. Figure 4 depicts alternative banking programs by the size of the accumulated bank at the end of each calendar year, as it is built up in Phase I and drawn down in Phase II.

Figure 4: Optimal Initial Banking Programs



The shaded, fuzzy line tracks the actual evolution of the SO_2 allowance bank through 2002. All the other lines indicate efficient initial banking programs with plausible assumptions about discount rates and growth in counterfactual emissions in the early to mid-1990s. The bold line in the center, which closely tracks the observed banking path through 1999, represents a program with parameter values corresponding to a 4.0% real discount rate and growth in counterfactual emissions at 0.65% per annum. The lines above and below this banking path reflect combinations of higher or lower growth in counterfactual emissions (1.25% and 0% per annum, respectively) and higher or lower discount rates (5% and 3%), as indicated in the legend to Figure 4. These variations yield larger or smaller banking programs that reflect differences in expected costs over time because of differences in expected rates of growth in counterfactual emissions or in discount rates.

It is readily apparent in Figure 4 that the actual banking path during Phase I implies a discount rate of about 4.0% and expected growth of counterfactual emissions of 0.65%, or some approximately equivalent combination of discount rate and expected growth in counterfactual emissions. In any case, the accumulation of the bank during Phase I follows a path that would be optimal within a range of values for *g* and *r* that are reasonable for this period.

Beginning in 2000, the actual banking path departs from the constant-parameter-value path indicated through 1999 in a direction that would indicate a lower discount rate, higher counterfactual emissions, or some combination of the two. Thus, an important consideration in evaluating the efficiency of banking under the U.S. Acid Rain Program is whether such a change in parameter values can be reasonably assumed to have occurred.

There was a significant decline in the real discount rate over this period: from approximately 4.0% in 2000 to 3.30% in 2001 and 2.75% in 2002 (cf. Figure 3). However, the decline in the draw-down rate was greater than what can be justified by lower discount rates as shown in Table 2, where the three data columns show the optimal draw-down rates for three cases: 1) if no change in expectation concerning the growth in counterfactual emissions or the discount rate had occurred, 2) if the observed changes in discount rate are allowed but growth in the counterfactual remained unchanged, and 3) the observed draw-down rate.

TABLE 2. Comparison of Drawdown Rates, 2000-02 (million allowances or tons)

Year	No change in g or r	Actual r, no change in g	Observed Drawdown
2000	1.71	1.71	1.25
2001	1.60	1.43	1.06
2002	1.48	1.21	0.90

The most likely explanation by our model is unobserved, a change in expectation concerning the growth in counterfactual emissions or other conditions that would make allowances more valuable. As noted earlier, the rate of increase in counterfactual emissions since 1995 had been above the high end of early 1990s expectations and by 2000 a shift upward in expectations concerning counterfactual emissions for the remainder of the banking period would be entirely plausible. Two other conditions may also have contributed to an upward adjustment in expectations the first of which was the seasonally high level of natural gas prices in the summer of 2000 that became much higher in the course of the California electricity problems of late 2000 and early 2001 and which would have led to expectations of higher utilization of coal-fired generating units and higher demand for allowances. The second contributing factor may have been the announcement of legislative proposals in Congress to lower the SO_2 cap by about two-thirds and the expression of sympathy for such an approach voiced by then presidential candidate George W. Bush during the 2000 campaign. Although some of these proposals would have limited the ability to carry-over allowances from one cap to the other, legislation having a chance of enactment could reasonably have been expected to allow carry-over if for no other reason to avoid the dumping of allowances and the concomitant increase of emissions in the final years of the old cap as agents sought to salvage what value they could from banked allowances. Allowing carry-over would have the opposite effect of not using allowances and reducing emissions during the final years of the old cap as agents initiated a new banking period reflecting the realignment of the pattern of temporal abatement to incorporate the expected effect of the lower cap. So long as this prospect was given some probability of occurring, the effect would be to diminish the drawdown rate.

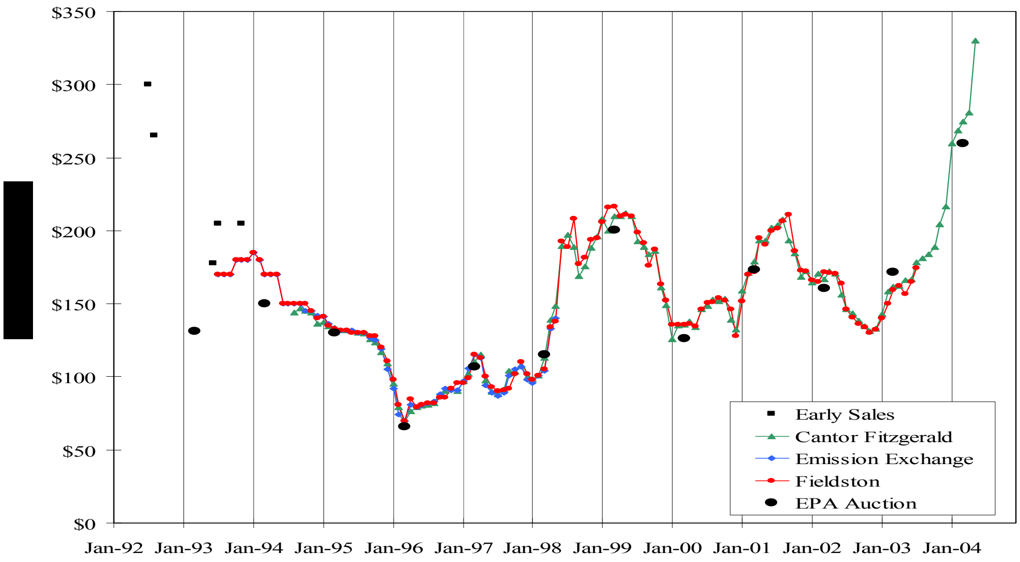
5. Discussion of Results

The finding of reasonably efficient SO_2 allowance banking is surprising in view of the earlier and widespread perception (to which we contributed) that, if anything, there had been too much banking during Phase I. An explanation of the error in our earlier thinking reveals the importance of two features that would affect banking behavior: 1) the differences in the emissions and abatement characteristics of the units included in the different phases of a phased-in emission reduction program, such as the U.S. Acid Rain Program, and 2) the extent of irreversibility in abatement decisions. In this section of the paper, we first describe the mistaken, earlier view of inefficient banking to make clear the mistaken assumptions behind the too facile conclusion of inefficient banking. Then, we proceed to an analysis of how the differing composition of units in Phases I and II created a unique condition of robustness whereby changes in expectations concerning the level of counterfactual emissions, such as occurred at the beginning of Phase I, would have had little effect on banking behavior assuming sufficient reversibility in abatement decisions. In the third subsection, we present the evidence that abatement decisions have been sufficiently reversible to permit efficient banking.

5.1. The Earlier View of Inefficient Banking

The behavior of SO_2 allowance prices in the early years of Phase I strongly suggests the existence of an error in expectation concerning the level of counterfactual emissions and therefore the amount of abatement required to comply with the new SO_2 caps. The allowance price revealed by the first EPA auction in March 1993, \$131, was about half the level of the few trades then reported at prices ranging between \$250 and \$300 and of an informed estimate of the expected Phase I price at \$250 (EPRI, 1993). Although initially these unexpectedly low prices were attributed to the design of EPA's annual auction (Cason, 1993; Cason, 1995; and Cason and Plott, 1996), the clearing price in the first EPA auction in 1993 proved to be a remarkably good predictor of Phase I prices. Subsequent research indicated that the scope for the perverse seller incentives identified by Cason and Cason and Plott were limited (Joskow, Schmalensee, and Bailey, 1998) and that SO_2 emissions in the high-sulfur-consuming Midwest had fallen for reasons unrelated to the U.S. Acid Rain Program, namely, the reduction in the delivered cost of western low-sulfur coals to the Midwest due to the deregulation of railroads during the 1980s (Ellerman and Montero, 1998). When Phase I entered into effect in 1995 and as evidence accumulated with the quarterly emission reports in April and July of 1995 that SO_2 emissions at affected sources were much lower than expected, allowance prices fell even more to an all-time low, slightly under \$70, in early 1996. Since making this initial adjustment, the SO_2 allowance price has risen steadily but erratically along what can be described as a stochastic Hotelling path as shown in Figure 5 below.

Figure 5: Current Vintage SO_2 Allowance Prices, 1993-2004



It was also evident in 1995 that the abatement decisions made by the owners of the units subject to the initial Phase I cap were to a considerable extent irreversible. Almost half of the abatement in the first years of Phase I resulted from investments in scrubbers, a capital-intensive abatement technique which requires significant lead-time in construction and is irreversible for long periods of time. Moreover, an undetermined part of the alternative technique for abatement, switching to low-sulfur coal, involved multi-year contracts some of which were signed prior to 1995 as a hedge against an expected increase in low-sulfur coal prices once the U.S. Acid Rain Program became effective in 1995. These contracts would also have contributed to irreversibility by limiting the extent to which affected units could be switched back to high-sulfur coal when the expected higher prices for low sulfur coal did not materialize, thereby increasing the demand for allowances.

The combination of an unexpectedly lower level of counterfactual emissions and of what seemed to be significant irreversibility in abatement decisions led many observers (the authors included) to believe that more allowances were being banked than would have occurred in the absence of the error in expectation. In sum, an efficient level of banking would not be achieved: a conclusion that is directly contradicted by the finding of this paper.

5.2. The Robustness of Banking to Errors in the Counterfactual

The first error in the earlier conclusion lies in the implicit assumption that changes in expectation concerning the initial level of counterfactual emissions, u_0, would have a significant effect on the amount banked during the accumulation phase. During any accumulation year, such as in Phase I, observed abatement can be divided into two components, that required to reduce emissions to the level of the cap in the current year and a further amount to produce allowances that can be banked for later use. How a change in the counterfactual, or more generally, errors in expectation concerning counterfactual emissions, might affect each of these components was not analyzed earlier and has not received any attention in the literature on permit banking.

This question can be answered using the model developed in section 2 and comparing outcomes for different levels of counterfactual emissions, u_0, assuming that there is no irreversibility and holding constant the values for the discount rate, the expected growth in counterfactual emissions, and the cost function. Table 3 reports the results of such an experiment for changes in u_0 that range from +20% to -20% of the value of initial counterfactual emissions used in the simulations reported earlier.

Table 3: Effect of Variations in the Level of Initial Counterfactual Emissions on 1995 Price and Quantities

	1995 Price (1995$/ton)	Abatement in 1995 (million tons)	Allowances Banked in 1995 (millions)	1995 Emissions (million tons)
+20%	$162	5.41	2.11	5.51
+10%	$134	4.48	2.09	5.53
Base Case	$107	3.57	2.09	5.53
- 10%	$80	2.67	2.10	5.52
-20%	$54	1.81	2.15	5.47

There is a remarkable difference between the effects expressed in the first two data columns and the last two. The initial price and quantity of abatement fall significantly with lower initial counterfactual emissions; yet, the amount of banking and the level of predicted initial emissions with the cap hardly change. It would appear that, when abatement decisions are completely reversible, all of the adjustment to this type of error in expectation will be made in current abatement, not abatement motivated by banking.

A more complete picture of the effect of a change in the counterfactual on banking behavior is given by Figure 6, which shows the optimal emission paths for affected sources for the same variations in the initial counterfactual.

Figure 6: Optimal Emission Paths with Variation of the Initial Counterfactual

As found in Table 3, the level of emissions and the amount of banking in 1995 is virtually unchanged. In the later years of Phase I, slightly lower emissions and more banking are associated with a higher counterfactual; however, the accumulated bank at the end of Phase I, which ranges from 11.06 to 12.14 allowances or tons of SO_2, is affected little by what are large changes in the level of the counterfactual. The difference between initial counterfactuals that are 20% more or less than the central estimate is 6.3 million tons annually. When this difference is accumulated over the five years of Phase I, required abatement changes by about 32 million tons, but the effect on the end-of-Phase I bank is only one million tons. The main effect of a change in the level of counterfactual emissions is seen not in Phase I, but in the draw-down phase of the banking program. While a higher counterfactual causes slightly lower emissions in Phase I, Phase II is characterized by higher emissions in the early years and a faster draw-down of the accumulated bank.

The disproportionate effect of a change in the level of counterfactual emissions on the two components of observed abatement during Phase I reflects the very different motivation of each. The current abatement component, what would be required if there were no banking, is determined by the distance from the level of counterfactual emissions to the current year cap. The banking component is motivated by the increase in marginal cost that would occur, without banking, in the transition from Phase I to Phase II. Using the notation from the earlier part of the paper, this difference is expressed in (8) below, from which the effect of a change in the initial counterfactual can be easily derived, as given at (9).

$$\Delta mc = \alpha_2 \left(u_0 e^{gT} - a_2 \right) - \alpha_1 \left(\varepsilon u_0 e^{gT} - a_1 \right) \tag{8}$$

$$\frac{\partial \Delta mc}{\partial u_0} = \alpha_1 e^{gT} \left(\frac{\alpha_2}{\alpha_1} - \varepsilon \right) \tag{9}$$

In (8), the two terms in parentheses on the right-hand-side express the quantity of required abatement at that moment in continuous time when the transition occurs and the alpha coefficients provide the scalars translating abatement into marginal cost for the two marginal abatement cost curves. The two terms outside the parentheses in (9) are scalars that translate quantity of abatement into cost and that adjust u_0 to time T. The critical variable, determining the sign of the effect, is the term in parentheses. The ratio, α_2 / α_1, is that of the slope of the Phase II cost function to the slope of the Phase I cost function and it expresses the ratio of Phase II marginal cost to Phase I marginal cost for the same level of abatement. As noted in section 3.2, this ratio is equal to observed abatement from Phase I units relative to total abatement, which has been about 0.83. The term, ε, is the ratio of counterfactual emissions from Phase I units relative to the counterfactual for all units, and it is equal to 0.57. The derivative at (9) is then positive for this program, but only because the sources included in Phase I include a greater proportion of the abatement potential than of emissions for all sources that would be covered in Phase II. It is easy to see that this difference, which reflects the characteristics of the generating units included in Phase I, could be very small and even negative for programs with different design features. And, if a program were designed such that the sources included in the two phases were the same and only the cap changed, this difference and the derivative at (9) would be zero.

The importance of this feature of program design when the composition of sources included in the two phases of banking are different can be illustrated by a simple example.

Table 4. Illustration of the Effect of a Change in Counterfactual on dMC_T

Variable (all at T and without banking)	α_2/α_1 = .50		α_2/α_1 = .75	
	Before	**After**	**Before**	**After**
Ph 1 abatement	5	7.5	5	7.5
Ph 1 marginal cost	\$5.00	\$7.50	\$5.00	\$7.50
Ph 2 abatement	15	20	15	20
Ph 2 marginal cost	\$7.50	\$10.00	\$11.25	\$15.00
Difference in marginal cost	\$2.50	\$2.50	\$6.25	\$7.50

As before, a linear marginal abatement cost function is assumed. For this illustration, we assume that 50% of the emissions are included in Phase I and we compare the effect on banking when the Phase I share of abatement opportunities, α_2 / α_1, is also 50% and when this share is greater. Assume also that the Phase I and II caps are such that, before any change in the counterfactual, the required abatement from the Phase I sources at T is 5 units and abatement required of all sources at the same time is 15 units.

A first point to note is that the difference in marginal cost between the Phase I units and all units at T (shown in the final row) is considerably greater when the Phase I units constituting half of emissions contain three-quarters of the abatement opportunities instead of only half. In this latter case, the additional units brought into the program in the second phase double the amount of emissions to be covered, but the abatement possibilities are increased by only one third when α_2 / α_1 = 0.75. Under these conditions, an efficient banking program would require more of the cumulative required abatement be undertaken by the Phase I units in the accumulation phase. Equivalently, the lower cost of Phase I abatement, relative to that in Phase II, justifies more abatement by Phase I units and more banking.

The columns labeled “after” in Table 4 show the effect of the 5 unit increase in the counterfactual at T. When the shares of the Phase I units in counterfactual emissions and in abatement possibilities are the same, the marginal costs for the Phase I units and for all units at T are both higher by $2.50 when the counterfactual is increased, but the difference in marginal costs at T, which motivates banking, remains unchanged. However, when the share of abatement possibilities included in the Phase I units is greater than their share of emissions, the difference in marginal cost increases. Again, the additional counterfactual emissions being brought under the Phase II cap are not accompanied by a corresponding increase in abatement possibilities. Consequently, an efficient banking program will distribute a greater share of the cumulative required abatement to the Phase I units, which requires an even higher price in Phase II to motivate more banking than would otherwise occur.

How much abatement can be distributed to the Phase I units depends however on the conditions expressed in equation (3), namely, that the number of allowances accumulated during Phase I must be equal to the number drawn-down in Phase II. In the case of the U.S. Acid Rain Program, the duration of the accumulation phase is fixed and efficient banking requires that marginal cost rise at the discount rate throughout the banking period. As it turns out in the U.S. Acid Rain Program, the optimal amount of abatement to be redistributed is very little.

The complex interaction that occurs can be illustrated by the following differential, which expresses the change in cumulative Phase I emissions at $t\ (\leq T)$ that is associated with a one-unit change in the counterfactual at t = 0.[25]

$$dE_t = \left[\frac{\partial U_t}{\partial u_0} - \left(\frac{\partial Q_t}{\partial u_0} + \frac{\partial Q_t}{\partial \tau} \frac{\partial \tau}{\partial u_0} \right) \right] du_0 \tag{10}$$

---

[25] Since optimal banking requires that the number allowances accumulated be equal to the number drawn down, the effect of any change on the quantity of allowances banked will be indicated by the change during the accumulation phase.

In this formula, the capital letters represent the time integrals for actual emissions, counterfactual emissions, and abatement. Appendix 1 provides the complete derivation of this cumulative differential for any *t* within Phase I, but most of the terms can be understood without reference to the more complete derivations. The first term within brackets is obviously positive. With no change in the level of the emission cap, the first term within parentheses must also be positive; but it will not necessarily be equal to the increase in cumulative counterfactual emissions for any $t \leq T$. The complete effect on abatement depends also on the second term within brackets, which is the product of the partial derivatives expressing the chained effect of a change in the counterfactual on τ and of the change in τ on abatement. As shown in the appendix, the sign of this differential is negative for the parameters applying to the U.S. Acid Rain Program. A one million ton increase in the initial counterfactual leads to a reduction of emissions during Phase I that goes from about 50,000 tons at $t = 0$ to 100,000 tons at $t = T$. Over the five years of Phase I, the cumulative reduction in actual emissions, and therefore the increase in the end-of-Phase I bank, is about 230,000 tons. When compared with the slightly more than 5 million ton increase in the counterfactual, the effect is small.

In summary, the requirements of an optimal banking program imply that most of any change in counterfactual emissions be absorbed by the change in current abatement required to reduce emissions to the Phase I cap. The effect on abatement undertaken for purposes of banking is more complicated. As explained in the appendix and in the discussion following Table 4, this effect depends strongly on the shares of counterfactual emissions and abatement possibilities comprised by the Phase I units. These values are such in the U.S. Acid Rain Program that banking behavior is robust to changes in the counterfactual, but this finding is not general and would not necessarily apply to other programs with banking that might define the sources to be included in the two phases differently.

5.3. The Condition of Sufficient Reversibility

Banking behavior would remain robust when a mistaken view of the counterfactual is recognized only if abatement decisions are sufficiently reversible to allow the required adjustment in current abatement to occur. For instance, if all abatement decisions were irreversible, it would not be possible to change the current abatement component of total abatement and all of the excess abatement would be banked. The relevant question then is whether the undeniable elements of irreversibility in the U.S. Acid Rain Program were sufficiently important to prevent the required adjustment in current abatement during Phase I. If agents had sufficient ability to adapt to the lower than expected prices, that is, if a sufficient proportion of the initial intentions and commitments to abatement were reversible in the course of Phase I, and especially in 1995 as prices were falling, no excess banking would occur.

The clear implication of the finding of reasonably efficient banking in this paper is that sufficient reversibility existed. Evidence in support of this conclusion is provided by the choices of abatement technique since 1995 by the units continuously subject to the Acid Rain Program since 1995, as shown in Table 5.

Table 5: Emission Reduction due to Scrubbing and Switching (Phase I units only)

	Scrubbing (million tons)	Switching (million tons)	Scrubbing Share
1995	1.71	2.19	44%
1996	1.83	2.20	45%
1997	1.92	2.29	46%
1998	1.89	2.47	43%
1999	1.80	2.74	40%
2000	1.94	3.38	36%
2001	1.99	3.27	38%
2002	1.97	3.56	36%

Switching, the form of abatement requiring the least lead time and having the least irreversibility, increased by about 63%, or by almost a million and a half tons, while the amount of abatement from scrubbing increased relatively little from 1997, when all the Phase I scrubbers were first operating for the full year. The magnitude of the initial error in expectations can only be estimated, but the magnitudes are such that earlier intentions to abate one or even two million tons more by switching in 1995 could be presumed either to have been cancelled or to have been quickly reversed as prices fell from around $150 in late 1994 to the all time low of $70 in early 1996. Then, as allowance prices increased in the ensuing years to highs of as much as $200, much of what may have been initially cancelled abatement by switching was restored.

This pattern of incremental abatement since 1995 also identifies the cost of inefficiency created by the elements of irreversibility for the particular error in expectation observed in the U.S. Acid Rain Program. Efficient banking requires sufficient reversibility, but sufficient reversibility does not imply that the shares of abatement by competing techniques are optimal. Instead of the approximately 45:55 split between abatement by scrubbing and by switching observed in 1995, the efficient split—given perfect foresight or a complete absence of irreversibility—might have been closer to the 35:65 split observed in 2002. The error in expectation led to an earlier commitment to irreversible investments in abatement than would otherwise have occurred; however, this inefficiency dissipated as allowance prices rose over time to the level of the erroneous expectation that justified the investments in the first place.[26] With sufficient reversibility, the cost of the inefficiency associated with the irreversible commitments showed up not so much in the price of allowances but in the supply of the reversible technique, in this case, switching to low sulfur coal, the suppliers of which were faced with greater reduction in demand than would have been the case had the initial expectations been correct or had it been possible to undo the irreversible abatement decisions.

6. Conclusion

The conclusion of this evaluation of the temporal efficiency of SO_2 allowance banking—that the temporal pattern of abatement under the U.S. Acid Rain Program has been reasonably efficient—is both reassuring and surprising. It is reassuring in affirming once again that properly constructed markets produce good results. The surprise results from the discrepancy with the earlier view that banking in this program was excessive. Reconciling the results of this paper with that view reveals the importance of

---

[26] Nearly all of the units retrofitted with scrubbers at the beginning of Phase I received bonus allowances that were contingent on this form of investment. Although this incentive may have made a difference in some cases, an analysis of the value of the bonus allowances relative to those of the allowances freed up for sale by scrubbing shows that the latter are far more important and that the decision to invest rests more on the expected value of allowances however obtained than on the additional bonus allowances.

understanding both how changes in expectations concerning counterfactual emissions affect banking and the extent of irreversibility in abatement decisions. In the case of the U.S. Acid Rain Program, efficient banking is relatively robust with respect to changes in expectation concerning counterfactual emissions; however, this robustness reflects elements of program design that are specific to this program. Also, the elements of irreversibility in abatement decisions were not so large as to prevent a reduction in abatement approximately equal to the error in expectation concerning the level of initial counterfactual emissions. The undeniable elements of irreversibility were, however, not without a cost in efficiency, which showed up not in the amount of banking but in the choice of abatement technique during the early years of Phase I when less switching to lower sulfur coal occurred than would have been optimal given perfect foresight or complete reversibility.

Several of the findings in this paper suggest more general implications for banking behavior. First, the correlation of the returns from holding SO_2 allowances with those from holding a diversified portfolio of equities is very low if not zero. This means that the discount rate will be relatively low and the role of banking relatively more important than it would be with more correlation and a higher discount rate. Caution should be used in generalizing this result to other tradable permit systems, such as for NO_x or CO_2 emissions; however, neither should it be assumed that the allowance beta, and the discount rate, is high. Second, changes in the level of counterfactual emissions will always have a much larger effect on the quantity of abatement and on allowance prices than on banking behavior which depends on a number of other factors. Banking may be affected by changes in the level of counterfactual emissions, but the effect will be small and its magnitude and sign will depend on design features specific to each program. Third, when allowed in phased-in cap-and-trade programs, banking can be expected to produce more abatement and higher allowance prices in the early phases of the program in what can be seen as voluntary "early action." This implication will be particularly important in designing CO_2 cap-and-trade programs. Since the level of the counterfactual is inherently uncertain and the initial CO_2 caps are likely to be relatively undemanding, the expectation (or the reality) of later, more stringent caps will tend to produce a positive price and early abatement even if the initial cap is non-binding.

The results of this paper should not be read as asserting that SO_2 allowance banking has been efficient in any exact sense; few real-world examples of economic behavior meet this test. The uncertainties about discount rates, growth in counterfactual emissions, and abatement cost functions are too great to allow such a statement. Nevertheless, the uncertainties can be bounded within relatively narrow ranges and when these likely values are used, reasonably efficient banking is indicated. Some agents may have hoarded or even dumped banked allowances in a manner that could not be judged to be economically efficient, but these exceptions have not been important enough to affect aggregate behavior noticeably. More importantly, the U.S. Acid Rain Program has shown that firms will use banking provisions in a rational and predictable way. There is no support in the experience with this program for the often expressed fear that firms will create emission spikes by using banked permits all at once. To the contrary, abatement is moved forward in time and the temporal pattern of emissions is remarkably smooth and stable. Allowance prices are considerably more variable than emissions, reflecting the many vagaries that affect any market, but this property should be of less importance to environmental regulators.

Appendix

Optimal banking program under the U.S. Acid Rain Program consists of two phases, an accumulation phase occurring from t = 0 to t = T and a drawdown phase from T to t = τ, for which the quantities of allowances accumulated and drawn-down are equal by definition. With a fixed cap during both phases, any change in cumulative banking behavior will be reflected in a change in cumulative emissions. Moreover, since the size of the bank is determined during the accumulation phase, it suffices to focus only on the change in emissions during this phase to determine the effects of a change in the counterfactual on the size of the bank.

To simplify notation, capital letters are used to signify cumulative amounts so that e_t denotes emissions at *t* and E_t denotes cumulative emissions from t_0 through *t*. Moreover, consideration is restricted to values of *t* within the range $0 \le t \le T$, or within Phase I. Furthermore, the abatement cost function is assumed to be linear, that is, $\gamma = 1$.

$$\begin{aligned}
&a)\quad e_t = u_t - q_t = \varepsilon u_0 e^{gt} - \frac{\alpha_2}{\alpha_1} q_\tau e^{-\rho(\tau-t)} \\
&b)\quad U_t = \varepsilon u_0 \int_0^{t'} e^{gt}\,dt = \varepsilon u_0 \left(\frac{e^{gt}-1}{g}\right) \\
&c)\quad Q_t = q_\tau \frac{\alpha_2}{\alpha_1} \int_0^{t'} e^{\rho(\tau-t)}\,dt = \left(u_0 e^{g\tau} - a_2\right) e^{-\rho\tau} \frac{\alpha_2}{\alpha_1} \left(\frac{e^{\rho t}-1}{\rho}\right) \\
&d)\quad E_t = U_t - Q_t = A_t - B_t
\end{aligned} \qquad \text{A.1)}$$

From (A.1d), the change in the bank for any t > 0 is:

$$dE_t = dU_t - dQ_t = -dB_t \quad \text{since } dA_t = 0. \qquad \text{A.2)}$$

From (A.1b) and (A.1c), U_t and Q_t are directly dependent on u_0. In addition, Q_t is a function of τ, which is a also a function of u_0 by the implicit function for τ stated at equation (7) of the main paper. Thus,

A.3) $\tau = \tau(u_0, \cdots)$

Accordingly, differentiation of (A.2) yields

$$\frac{dE_t}{du_0} = \frac{\partial U_t}{\partial u_0} - \left(\frac{\partial Q_t}{\partial u_0} + \frac{\partial Q_t}{\partial \tau}\frac{\partial \tau}{\partial u_0} \right) = -\frac{dB_t}{du_0} \tag*{A.4)}$$

as at (10) of the main text.

The direct partial derivatives with respect to u_0 can be quickly found and they are:

$$\begin{aligned} &a) \quad \frac{\partial U_t}{\partial u_0} = \varepsilon \left(\frac{e^{gt} - 1}{g} \right) > 0 \\ &b) \quad \frac{\partial Q_t}{\partial u_0} = e^{(g-\rho)\tau} \frac{\alpha_2}{\alpha_1} \left(\frac{e^{\rho t} - 1}{\rho} \right) > 0 \end{aligned} \tag*{A.5)}$$

As would be expected with an unchanging cap, these two terms are off-setting and the exhaustion condition for optimal banking would require them to be equal if there were no change in τ as a result of the change in the counterfactual. In the event, τ will be recalculated and it is therefore necessary to calculate both the effect of a change in τ on cumulative abatement and the change in τ occasioned by the change in the counterfactual.

The partial derivative of cumulative abatement with respect to τ is:

$$\frac{\partial Q_t}{\partial \tau} = \frac{\alpha_2}{\alpha_1} \left(\frac{e^{\rho t} - 1}{\rho} \right) e^{-\rho\tau} \left[g u_0 e^{g\tau} - \rho \left(u_0 e^{g\tau} - a_2 \right) \right] = e^{-g\tau} \frac{\partial Q_t}{\partial u_0} \left(g\, u_\tau - \rho\, q_\tau \right) \tag*{A.6)}$$

The partial derivative for τ with respect to u_0 can be found by differentiating the implicit function for τ [equation (7) in the main paper] by u_0 recognizing that τ is a function of u_0 and then rearranging terms. The result is:

$$\begin{aligned} &\frac{\partial \tau}{\partial u_0} = \frac{\dfrac{\rho}{g}\left[e^{g\tau} - e^{gt}(1-\varepsilon) - \varepsilon \right] - e^{g\tau} X}{\left(g u_\tau - \rho\, q_\tau \right) X} \\ &where \quad X = 1 - e^{-\rho\tau} \left[e^{\rho t} - \frac{\alpha_2}{\alpha_1}\left(e^{\rho t} - 1 \right) \right] > 0 \end{aligned} \tag*{A.7)}$$

The term $g u_\tau - \rho q_\tau$ cancels out in the product of (A.6) and (A.7). Furthermore, substitution of (A.5), (A.6), and (A.7) into (A.4), cancellation, and rearrangement leads to the following result for the effect of a change in the counterfactual on cumulative emissions for any *t* in Phase I, $0 < t \leq T$.

$$\frac{dE_t}{du_0} = \varepsilon\left(\frac{e^{gt}-1}{g}\right) - \frac{\alpha_2}{\alpha_1}\left(\frac{e^{\rho t}-1}{g}\right)\left[\frac{\left(e^{g\tau}-e^{gt}\right)+\varepsilon\left(e^{gt}-1\right)}{\left(e^{\rho\tau}-e^{\rho t}\right)+\left(\alpha_2/\alpha_1\right)\left(e^{\rho t}-1\right)}\right] \qquad \text{A.8)}$$

A similar derivation can be used to find the effect on the level of emissions at any point in time during Phase I.

$$\frac{de_t}{du_0} = \varepsilon e^{gt} - \frac{\alpha_2}{\alpha_1}e^{\rho t}\frac{\rho}{g}\left[\frac{\left(e^{g\tau}-e^{gt}\right)+\varepsilon\left(e^{gt}-1\right)}{\left(e^{\rho\tau}-e^{\rho t}\right)+\left(\alpha_2/\alpha_1\right)\left(e^{\rho t}-1\right)}\right] \qquad \text{A.9)}$$

Which simplifies at t = 0 to become

$$\frac{de_0}{du_0} = \varepsilon - \frac{\alpha_2}{\alpha_1}\frac{\rho}{g}\frac{e^{g\tau}-1}{e^{\rho\tau}-1} = \varepsilon - \frac{\alpha_2}{\alpha_1}\frac{\left(e^{g\tau}-1\right)/g}{\left(e^{\rho\tau}-1\right)/\rho} \qquad \text{A.9a)}$$

In all three of these equations, (A.8), (A.9) and (A.9a), the parameters governing the effect of a change in the initial counterfactual on emissions during Phase I, and therefore on banking, are four exogenous variables, two reflecting program design, ε and α_2/α_1, and two, *g* and *ρ*, expressing the growth of emissions and the discount rate, as well as the endogenous variable, τ, that defines an optimal banking program.

Analyzing the simpler equation (A.9a) helps to understand the other, more complicated equations. The two right-hand-side terms in all equations express the offsetting changes in counterfactual emissions and in abatement that will determine the change in constrained or observed emissions. At *t* = 0, the change in counterfactual emissions is simply the fraction of emissions that are covered in Phase I, or ε. The change in the abatement is considerably more complicated but it reduces to the product of the fraction of abatement opportunities among the Phase I sources and a ratio expressing the expected growth in counterfactual emissions and in abatement for the entire banking period both normalized to the value of these variables at t = 0. For g < ρ, as would be true of any banking program, this ratio will be always less than unity. The net effect on initial emissions will depend then on the relation between emissions and abatement opportunities for the Phase I sources and what might be considered the burden of future emissions growth in relation to planned abatement over the banking program. The greater the share of abatement opportunities and the greater the burden of expected emissions growth, the greater will be the effect of an increase in counterfactual emissions in lowering e_0 and the greater will be abatement during Phase I and the end-of-Phase I bank. As expressed in (A.8) and (A.9) for $t > 0$, the change in emissions reflects the same balance of forces modified only by the effects of the passage of time.

The signs and the values of the derivatives at (A.8) and (A.9) can be calculated for the U.S. Acid Rain Program as indicated in the following table for a one million ton change in the initial counterfactual.

Differential	Value
de_0	-0.049
de_T	-0.100
dE_T	-0.228

A one million ton increase in the initial counterfactual would result in a 49,000 ton decrease in emissions at $t = 0$, a 100,000 ton decrease at $t = T$, and a cumulative reduction in emissions and commensurate increase in the end of Phase I bank of 228,000 tons.

Calculating these values when the sources included in Phase I are the same as those in Phase II reveals some of the unique features associated with the U.S. Acid Rain Program. If the sources covered in the two phases are the same, (A.9a) would be always positive, since $g < \rho$ for any banking program to exist, and it would become:

$$\frac{de_0}{du_0} = 1 - \frac{\left(e^{g\tau} - 1\right)/g}{\left(e^{\rho\tau} - 1\right)/\rho} > 0 \qquad \text{A.10)}$$

With the same values for g, ρ, and τ, the value of the derivative would be +0.253. The value observed for the U.S. Acid Rain Program is negative because α_2/α_1 is considerably greater (0.83) than ε (0.57). When the Phase I affected sources contain a sufficiently larger fraction (depending on the values of g and ρ) of the abatement opportunities than of emissions, the requirements of optimal banking place a greater call on the Phase I sources for abatement than would be the case if their share of abatement opportunities were less.

The clear implication of this relation between the shares of emissions and abatement potential is that when these shares are equal an increase in the counterfactual leads to an increase in actual emissions and a decrease in banking, and vice versa. The intuitive explanation behind this behavior is that banking becomes less important because the absolute value of the no-banking difference in marginal cost at T (which will be unchanged when the shares are equal) constitutes a smaller percentage of the higher marginal cost in Phase II. Since the marginal cost of abatement in Phase I is no longer as cheap relative to Phase II marginal cost as before (even though both have risen), less banking is justified. Thus, for any given difference in no-banking marginal cost at T, banking becomes more or less important as a component of total abatement as the level of the counterfactual decreases or increases. In the U.S. Acid Rain Program, this relationship does not obtain because the Phase I units include a sufficiently greater share of abatement opportunities than emissions and the magnitude of the effect so small because the values of the variables in (A.8) and (A.9) are so nearly compensating.

Setting aside the complicated effects of a change in the level of counterfactual emissions on banking behavior, the predominant effect of a change in the counterfactual is that abatement for banking is far less affected than the abatement required to achieve the cap before any consideration of banking. The equations in this appendix show that the change in abatement may be more or less than the change in counterfactual emissions depending on the shares of emissions and of abatement potential included within each of the phases of the banking program.

References

Cason, Timothy N., 1993. “Seller Incentive Properties of EPA’s Emissions Trading Auction.” *Journal of Environmental Economics and Management*, 24, no. 2 (September): 177-95.

Cason, Timothy N., 1995. “An Experimental Investigation of the Seller Incentives in EPA’s Emission Trading Auction.” *American Economic Review*, 85, no. 4 (September): 905-22.

Cason, Timothy N. and Plott, Charles R., 1996. “EPA’s New Emissions Trading Mechanism: A Laboratory Evaluation.” *Journal of Environmental Economics and Management*, 30, no. 2 (March): 133-60.

Cronshaw, Mark B. and Jamie Brown Kruse, “Regulated Firms in Pollution Permit Markets with Banking,” *Journal of Regulatory Economics*, 9:179-189 (1996).

Electric Power Research Institute (EPRI), Integrated Analysis of Fuel, Technology and Emission Allowance Markets: Electric Utility Responses to the Clean Air Act Amendments of 1990, EPRI TR-102510 (August 1993).

Electric Power Research Institute (EPRI), *SO_2 Compliance and Allowance Trading: Developments and Outlook*, EPRI TR-107897 (April 1997).

Ellerman, A. Denny, *Analysis of the Bush Proposal to Reduce the SO_2 Cap*, MIT/CEEPR Working Paper 2002-002 (February 2002).

Ellerman, A. Denny, “The US SO_2 Cap-and-Trade Programme” in *Tradeable Permits: Policy Evaluation, Design, and Reform,* Organization for Economic Development and Cooperation (97-2004-07-1-P), Paris, France (2004).

Ellerman, A. Denny, Paul L. Joskow, Richard Schmalensee, Juan-Pablo Montero, and Elizabeth M. Bailey, *Markets for Clean Air: The U.S. Acid Rain Program*, Cambridge University Press, 2000.

Ellerman, A. Denny and Juan-Pablo Montero. “The Declining Trend in Sulfur Dioxide Emissions: Implications for Allowance Prices,” *Journal of Environmental Economics and Management* 36: 26-45 (1998).

Hotelling, H., “The Economics of Exhaustible Resources”, *Journal of Political Economy* 39, 137-175 (1931).

ICF Resources Inc. (ICF), “Economic Analysis of Title V (Acid Rain Provisions) of the Administration’s Proposed Clean Air Act Amendments (H.R. 3030/S. 1490),” prepared for the U.S. Environmental Protection Agency, September 1989.

Joskow, Paul L.; Schmalensee, Richard; and Bailey, Elizabeth M. 1998. “The Market for Sulfur Dioxide Emissions.” *American Economic Review*, ii, no. 4 (September): 669-85.

Pechan, E.H., and Associates, Inc. 1995. The Acid Rain Data Base Version 1 (ARDBV1). Contract #68-D3-0005, prepared for the U.S. Environmental Protection Agency’s Office of Atmospheric Programs, Acid Rain Division, Washington.

Resources Data International, Inc. (RDI, 1995), RDI’s Phase I Databook—Performance under the Clean Air Act Amendments of 1990, Boulder, CO.

Rubin, Jonathan D., “A Model of Intertemporal Emission Trading, Banking, and Borrowing,” *Journal of Environmental Economics and Management* 31, 269-286 (1996).

Rubin, Jonathan and Catherine Kling, “An Emission Saved is an Emission Earned: An Empirical Study of Emission Banking for Light-Duty Vehicle Manufacturers,” *Journal of Environmental Economics and Management*, 25, 257-274 (1993)

Schennach, Susanne, “The Economics of Pollution Permit Banking in the Context of Title IV of the Clean Air Act Amendments”, *Journal of Environmental Economics and Management* 40, 189-210 (2000).

Farrow, Scott, “Testing the Efficiency of Extraction from a Stock Resource,” *Journal of Political Economy* 93, 452-487.

Slade, Margaret and Henry Thille, “Hotelling Confronts CAPM: A Test of the Theory of Exhaustible Resources,” *Canadian Journal of Economics* 30, 687-708 (1997).

Smith, Anne E., Jeremy Platt and A. Denny Ellerman, “The Costs of Reducing Electric Utility SO_2 Emissions may not be as Low as You Might Think,” Working Paper 98-010, MIT Center for Energy and Environmental Policy Research, (August 1998).

United States General Accounting Office (USGAO), *Allowance Trading Offers an Opportunity to Reduce Emissions at Less Cost*, GAO/RCED-95-30 (December 1994).