



**CEEPR**

**Center for Energy and Environmental Policy Research**

The Temporal Efficiency of SO<sub>2</sub> Emissions Trading

by

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

**02-003 WP**

**September 2002**

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

# The Temporal Efficiency of SO<sub>2</sub> Emissions Trading

A. Denny Ellerman and Juan-Pablo Montero<sup>\*</sup>

## Abstract

This paper provides an empirical evaluation of the temporal efficiency of the U.S. Acid Rain Program, which implemented a nationwide market for trading and banking sulfur dioxide (SO<sub>2</sub>) emission allowances. We first develop a model of efficient banking and select appropriate parameter values. Then, we use aggregate data from the first seven years of the Acid Rain Program, to assess the temporal efficiency of the observed banking behavior. We find that banking has been surprisingly efficient and we discuss why this finding disagrees with the common perception of excessive banking in this program.

---

<sup>\*</sup> 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.

## 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.<sup>1</sup> Over the past decade, this latter dimension of emissions trading has drawn increasing attention in the literature and current proposals to decrease emission caps suggest a larger role for banking in the future.<sup>2</sup> Several authors have studied the theoretical properties of intertemporal trading,<sup>3</sup> but no work has yet evaluated how firms have actually responded to the possibility of trading emissions through time. 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, only 26.4 million of the 38.1 million allowances distributed were used to cover emissions. The remaining 11.65 million allowances (30% of all the allowances distributed) were banked. Equivalently, the reduction in emissions during Phase I was about twice what was required to meet the Phase I cap.<sup>4</sup> Then, in 2000-01, 2.8 million, or about a quarter, of these banked allowances were used to cover SO<sub>2</sub> emissions that were

---

<sup>1</sup> Logically, borrowing could also be included, but it is usually not.

<sup>2</sup> President Bush's Clear Skies Initiative would reduce the existing U.S. SO<sub>2</sub> emissions cap by another 70% in two steps starting in 2010 and current legislative proposals before the U.S. Congress would effect similar reductions. Moreover, an effective policy for reducing atmospheric greenhouse gas concentrations would likely include emission caps that would become more stringent over time.

<sup>3</sup> 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.

<sup>4</sup> A reasonable estimate of counterfactual emissions from all the units receiving allowances during the five years of Phase I is 48 million tons. The difference between this estimate of what emissions would have been without the Acid Rain Program and observed emissions from these same units is about 21 million tons, or a little more than twice the amount of abatement required to achieve the five-year cap of 38 million tons.

greater than the issuance of allowances for use in these years under the much tighter Phase II SO<sub>2</sub> emission cap.<sup>5</sup>

Banking was expected to occur in the Acid Rain Program, but there was little agreement on the size of the bank at the end of Phase I. The two-phased structure of the program created what would have been, without banking, an obvious and significant difference in required abatement between Phases I and II, and thus in the expected marginal cost of abatement. The provision permitting an unlimited carry-over of unused allowances made the equalization of marginal costs on a discounted basis possible and thereby provided the incentive for agents to abate more in Phase I and to use the banked allowances to abate less in the early years of Phase II. While all analyses agreed that banking would occur, early estimates of the size of the bank at the end of Phase I 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.<sup>6</sup>

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 two years of Phase II can be observed. 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

---

<sup>5</sup> 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<sup>e</sup> capacity and with 1985 emission rates of 2.5 #/mmBtu or higher were required to be subject to the SO<sub>2</sub> cap and to reduce emissions to an average emission rate equal to 2.5 lbs. SO<sub>2</sub> 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<sup>e</sup> were subject to SO<sub>2</sub> 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.

<sup>6</sup> 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.”

time that spans about half of the entire banking period and assesses whether observed behavior is efficient.<sup>7</sup>

Contrary to the common perception of excessive banking during Phase I,<sup>8</sup> we find that the evolution of the SO<sub>2</sub> allowance bank has been reasonably efficient. We argue that this misperception has been largely based on two mistakes: the use of a higher discount rate for SO<sub>2</sub> allowances than is warranted by price behavior in allowance markets and an assumption of less ability to adapt to the significantly lower than expected allowance prices at the beginning of Phase I coupled with a misunderstanding of the effect of that adjustment on the banking behavior of electric utilities.

The rest of the paper is organized as follows. Section 2 presents the model of banking that is used to generate efficient banking paths for comparison with observed banking behavior. Section 3 discusses the key assumptions underlying any banking program—program-specific parameters and assumptions about counterfactual emissions, the cost function, and the discount rate—and it provides estimates of the appropriate values for simulating efficient SO<sub>2</sub> allowance banking. Section 4 compares observed banking with simulated efficient paths and draws inferences from that comparison. Section 5 discusses why the findings from this research differ from the common perception of excessive SO<sub>2</sub> allowance banking. We conclude in section 6.

## 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,

---

<sup>7</sup> 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. We do have, however, some uncertainty on the demand side due to counterfactual emissions.

<sup>8</sup>At least, that was our perception before writing this paper (Ellerman et al., 2000; Smith et al., 1998).

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.<sup>9</sup>

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 will choose an abatement path that minimizes their expected present value of compliance costs using an appropriate risk-adjusted discount rate  $\rho$ . Risk-adverse agents will diversify this risk by holding a portfolio of assets including permits.

Because modeling the efficient path of the SO<sub>2</sub> 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

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

where  $E_t$  is the expected value operator,  $r^m$  is the expected rate of return on a well diversified market portfolio, and  $\beta$ , a common financial variable that determines the asset-specific risk premium, is the ratio of the covariance of  $\rho$  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

---

<sup>9</sup> In the case of the SO<sub>2</sub> program, firms are never certain about future electricity demand and the future prices for fuels of differing sulfur content.

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

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

A further condition that must hold during the banking period is that the cumulative number of allowances issued equal cumulative emissions.<sup>10</sup> Therefore, at any time  $t$  during the banking period, the time  $\tau$  at which the bank is expected to end must satisfy the exhaustion condition<sup>11</sup>

$$B(t) + \int_t^\tau a(t)dt = E_t \left\{ \int_t^\tau u(t)dt - \int_t^\tau q(t)dt \right\} \quad (2)$$

where  $B(t) \geq 0$  is the size of the allowance bank at  $t$ ,<sup>12</sup>  $a(t)$  is the number of allowances allocated at  $t$  that is specified by the legislation, and  $u(t)$  are counterfactual emissions, i.e., emissions that would have been observed in the absence of the SO<sub>2</sub> trading program, so actual emissions are  $u(t) - q(t)$ .

In addition, a terminal condition must hold at  $\tau$ . At that point in time and thereafter, temporal trading ceases and the only form of emissions trading observed will be spatial trading.<sup>13</sup> 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

---

<sup>10</sup> In exhaustible resource markets, this condition is commonly known as the exhaustion condition.

<sup>11</sup> Note that to allow for an allowance carry over at the end of the banking period a small negative term on the left hand side should be added. Numerical exercises, however, indicate a minor effect on the efficient banking path. For example, a large carry over of as much as 20% (2 million allowances) would increase the size of the bank at the end of Phase I by less than 2%.

<sup>12</sup> Initially we have that  $B(0) = 0$ .

<sup>13</sup> Differences between the current price and price expected in the next period may cause an allowance inventory to be built up and drawn down after  $\tau$ . This form of intertemporal trading will occur mostly between adjacent years and never lead to the multi-year accumulation and draw-down that characterizes the transition between Phase I and Phase II.

emissions. Thus, at any time  $t$  during the banking period, the time  $\tau$  at which the bank is expected to end must also satisfy the terminal condition

$$a(\tau) = E_t \{u(\tau) - q(\tau)\} \quad (3)$$

Conditions (1), (2) and (3) together with assumptions about both counterfactual emissions  $u(t)$  and the functional form of the aggregate abatement cost function  $C(q)$  allows us to solve for  $\tau$  and derive explicit efficient abatement and banking paths during the banking period. Counterfactual emissions are modeled as emissions at  $t = 0$  increasing at some (varying) rate,  $g(t)$ , that is,  $u(t) = u_i(0)e^{G(t)}$ , where  $i = 1$  during the Phase I years and 2 thereafter, and  $G(t) = \int_0^t g(s)ds$ . To estimate actual counterfactual emissions at time  $t$ , we let  $g$  to change from 0 to  $t$  according to past information on actual emissions and fuel use.<sup>14</sup> To estimate the evolution of expected counterfactual emissions beginning at  $t$ , we simply assume that agents expect  $g$  to remain constant throughout.

For abatement costs we assume that aggregate marginal abatement costs depend on aggregate abatement in the following form

$$C'(q(t)) = \alpha_i [q(t)]^\gamma \quad (4)$$

The scaling parameter,  $\alpha_i$ , 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.<sup>15</sup> The exponent,  $\gamma$ , reflects the curvature of the relationship and we assume it is the same for both the Phase I and Phase II aggregate cost functions and that it remains unchanged during the banking period.<sup>16</sup>

---

<sup>14</sup> See Ellerman et al. (2000) for more details.

<sup>15</sup> Firms that are not affected until 2000 are nonetheless present in the market and able to accumulate a bank during Phase I by purchasing allowances.

<sup>16</sup> The possibility of technological change will be discussed later.



Given the functional form in (4), conditions (1) and (3) can be combined to obtain the expected amount of abatement in each period  $q(t)$  as function of abatement at  $\tau$

$$q(t) = \begin{cases} q(\tau) \left( \frac{\alpha_2}{\alpha_1} \right)^{1/\gamma} e^{-\rho(\tau-t)/\gamma} & \text{if } 0 \leq t \leq T \\ q(\tau) e^{-\rho(\tau-t)/\gamma} & \text{if } T \leq t \leq \tau \end{cases} \quad (5)$$

where  $T$  denotes the end of Phase I and  $q(\tau)$  is given by (3). Substituting (5) into (2) yields a single equation that solves for the expected end of the efficient banking path  $\tau$  as function of  $B(t)$  and expectations formed at  $t$ .

We can now use the model to estimate  $\tau$  and the evolution of the efficient bank at the beginning of the trading program ( $t = 0$ ) by solving

$$a_1 T + a_2 (\tau - T) = u_1(0) \left( \frac{e^{g_0 T} - 1}{g} \right) + u_2(0) \left[ \frac{e^{g_0 \tau} - e^{g_0 T}}{g} \right] - \left( u_2(0) e^{g_0 \tau} - a_2 \right) \left[ \left( \frac{\alpha_2}{\alpha_1} \right)^{1/\gamma} e^{-\rho \tau / \gamma} \left( \frac{e^{\rho T / \gamma} - 1}{\rho / \gamma} \right) + \left( \frac{1 - e^{-\rho(\tau-T) / \gamma}}{\rho / \gamma} \right) \right] \quad (6)$$

where  $g_0$  denotes the expected growth rate of counterfactual emissions at  $t = 0$ .

The two terms on the left-hand-side of (6) state the number of allowances available in Phase I and during the years of Phase II constituting the draw-down phase of 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 up to 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 in parentheses outside the brackets is  $q(\tau)$ , the amount of abatement required at  $t = \tau$ , and the two terms in the brackets are indices of cumulative abatement, normalized to  $q(\tau)$ , for Phase I and for the Phase II part of the banking period, respectively.

Thus, once given a set of parameter values expressing the allowance cap, counterfactual emissions, and the abatement cost function, equation (6) can be solved for  $\tau$ , and an expected efficient banking path  $B(t)$  can be easily computed as

$$B(t) = \int_0^t [a(s) - u(s) + q(s)] ds \quad (7)$$

Although the model has been solved for parameter values estimated at the beginning of the banking period and assumed to remain constant throughout, it can be easily adapted to incorporate changes in agents' expectations or in market conditions during the banking period. With such changes, the observed banking path will reflect segments of differing efficient paths each reflecting successive starting points and the associated parameter values and accumulated banks.

### 3. Parameter Values

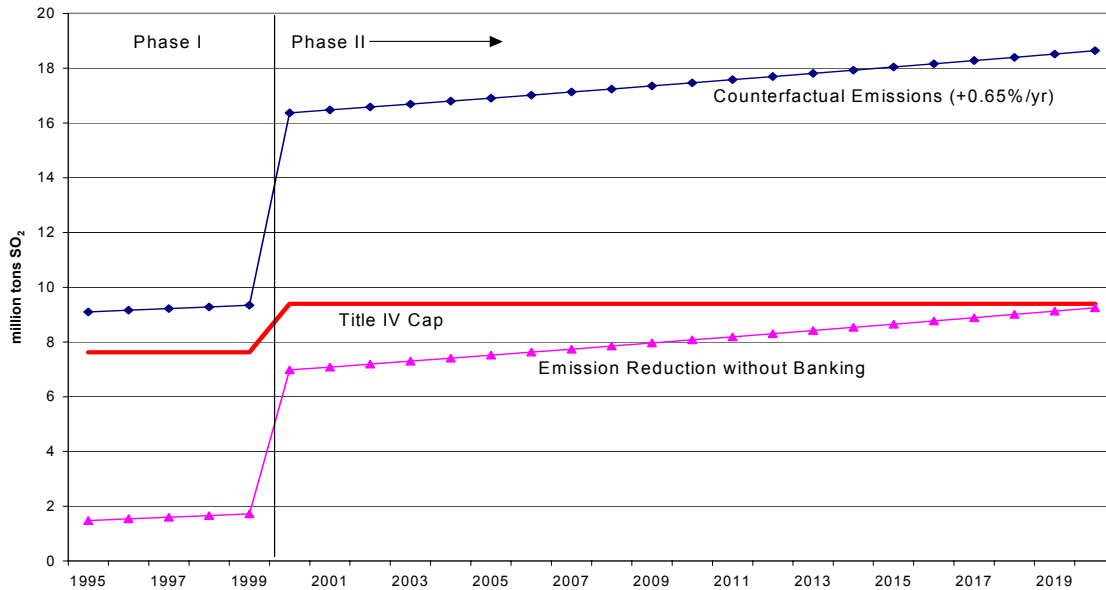
#### 3.1. Allowances and Counterfactual Emissions

The allowance cap and the assumptions concerning counterfactual emissions define 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 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  $t = 5$ .<sup>17</sup>

---

<sup>17</sup> 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 the last three years; 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. The assorted bonus and extra allowances amounted to .96 and .56 million allowances in 2000 and 2001, respectively, but they will be fewer in subsequent years.

Fig 1: Title IV Caps and Counterfactual Emissions



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 defined by a simple technique for calculating the counterfactual, 9.07 million tons of  $SO_2$ .<sup>18</sup> 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 for the units first affected in 2000. This value is 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 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

<sup>18</sup> The simple 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.

retirement of coal fired units, the utilization of nuclear capacity, and the economic competitiveness of new gas-fired generating units. High emissions scenarios predicted emissions growing at annual rate of about 1.25% per annum through 2010, while the low emissions cases predicted either constant SO<sub>2</sub> emissions after 2000 (EPRI, 1993) or emissions that are declining at rates between 0.5% and 1% per annum (ICF, 1989). The predicted increase in heat input was remarkably accurate (2.1% between 1993 and 2001), but the growth rate in counterfactual emissions has been much higher than expected. Over this period, the average annual rate of increase was 2.0%, considerably more than the 1.25% rate that was seen as the high end of the likely range.

While the early-1990s predictions of constant or declining counterfactual emissions are no longer plausible, it is also unlikely that the relatively high growth rate in counterfactual emissions since 1993 will continue. Two considerations are relevant. The first concerns the ability of existing coal-fired generating units to meet incremental demand by increasing utilization over the rest of the banking period, as they have over the past ten years. If these units are near capacity, so that incremental demand will have to be met by new units with low emissions, whether gas-fired units or coal units meeting stringent New Source Performance Standards, the rate of increase in emissions will be relatively low.<sup>19</sup> Announcements and construction of new base-load capacity, both coal and natural gas fired, over the past few years suggest that the ability of existing units to meet incremental demand is reaching a limit, even if further increases in the utilization of coal-fired units are possible, especially if the recent increase in the relative price of natural gas to coal persists.

The second consideration affecting assumptions about counterfactual emissions is what requirements may be placed on existing coal-fired units in response to other provisions of the Clean Air Act, such as new source review, visibility, fine particulates, or mercury regulations, to name only a few of the possibilities. Any requirement like these would reduce counterfactual emissions.<sup>20</sup>

---

<sup>19</sup> In fact, about ninety percent of planned generating capacity additions are gas-fired.

<sup>20</sup> There is always the possibility that allowances freed up by such actions would be “sterilized.” If so, the allowance cap is reduced by the same amount of as the counterfactual and the amount of reduction required would be the same as before.

The effect of both considerations is that a confident prediction of the growth in counterfactual emissions cannot be made, but the possibilities can be bounded. A forecast of no growth in emissions absent the cap imposed by the Acid Rain Program is no longer plausible, much less a decline in counterfactual SO<sub>2</sub> emissions. At the other extreme, it seems unlikely that counterfactual emissions could increase at the rate of growth in electricity production, as has been the case for the past seven years, because of technical limits on increased utilization of existing coal-fired generating capacity, the effects of other provisions on the Clean Air Act on these units, and the entry of a large amount of new combined-cycle gas-fired capacity after 2000.

For the simulations in this paper, we use four different values for expected growth in counterfactual emissions,  $g$ . Based on EPA's forecast (Pechan, 1995), we use a value of  $g = 0.65\%$  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 and ICF analyses of expected emissions absent the SO<sub>2</sub> cap. Based on the observed growth in demand for electricity during the 1990s and the recent change in the price of natural gas relative to coal, we use three values for  $g$ , 1.0%, 1.25%, and 1.50%, to reflect revised expectations that would be more appropriate for evaluating the remainder of the post-2000 banking period.

### 3.2. The Cost Function

Two parameter values define the cost functions: the convexity parameter,  $\gamma$ , and the scaling parameters of the cost functions for the Phase I units,  $\alpha_1$ , and for all units,  $\alpha_2$ . In equation (6), only the ratio of the scaling parameters is needed and it can be easily shown using equation (3) that  $(\alpha_1/\alpha_2)^{1/\gamma} = q_2/q_1$ , or the ratio of abatement by all units in Phase 2 to that by Phase I units in the same year. This ratio can be observed in 2000 and 2001 it is 1.21.<sup>21</sup>

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 various studies. The early EPRI study of abatement costs contains several charts of this relationship,

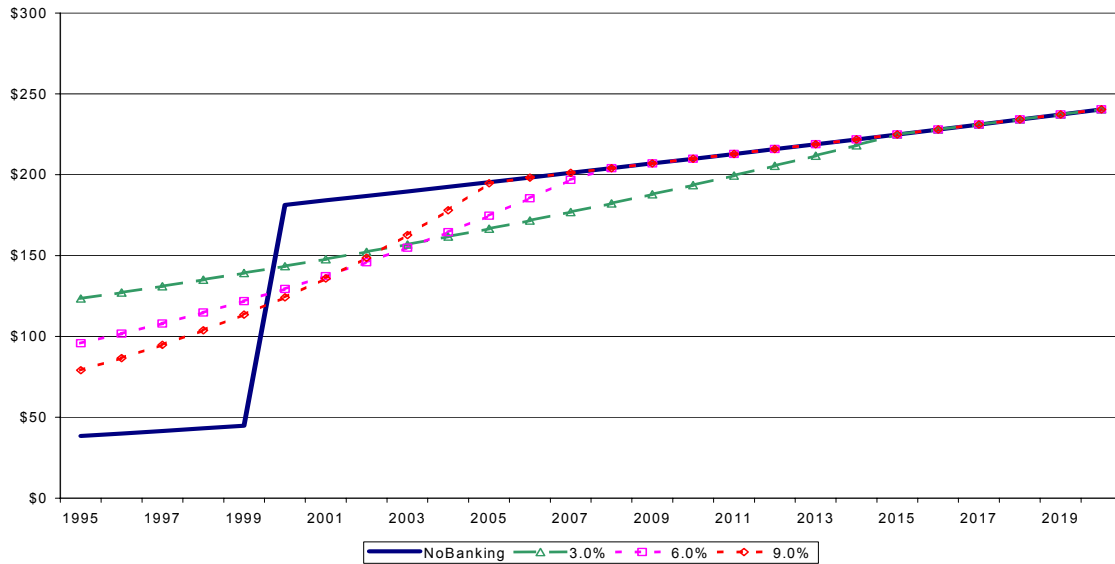
---

<sup>21</sup> Depending on assumption about the counterfactual this ratio varies between 1.19 and 1.23. The effect of this variation on the total amount of allowances banked at the end of Phase I is less than 2%.

which is linear over the relevant range.<sup>22</sup> Ongoing analysis by the authors and colleagues at MIT concerning the cost of reducing SO<sub>2</sub> emissions from the 2000 levels by retrofitting scrubbers to unscrubbed coal-fired units indicates a similar relationship (Ellerman and Joskow, 2002). 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<sub>2</sub> cap without banking can be calculated as is shown by the solid line in Figure 2.<sup>23</sup> This is the price dual of the quantity path given in Figure 1 and the opportunity to reduce cost or increase profit by banking is immediately evident.

Figure 2: 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 (1) 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,  $\tau$ , differs for each

<sup>22</sup> EPRI, 1993; Figures 5-4 and 6-15.

<sup>23</sup> In making this estimate, the scaling parameter,  $\alpha_2$ , is assigned a value of 26 that causes the simulated 2001 prices to approximate observed 2001 prices. Average monthly prices during 2001 ranged from \$159 in January to \$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  $\alpha_2$  is higher or lower than 26, or that 2001 prices were below or above true equilibrium prices.

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.<sup>24</sup>

### 3.3. The Risk-Adjusted Discount Rate

SO<sub>2</sub> 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 by investing the cash during the holding period 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<sub>2</sub> allowances, indicated by the beta ( $\beta$ ) coefficient in equation (1). With over seven years of SO<sub>2</sub> allowance price data now available, the monthly returns from holding SO<sub>2</sub> allowances can be readily calculated and correlated with the returns from holding a broadly diversified portfolio of equities.

Table 1 provides estimates of  $\beta$  when the monthly returns from holding SO<sub>2</sub> allowances are regressed by ordinary least squares on the monthly returns from holding various market indices over the same period.

<b>Table 1: Estimation of Beta for SO<sub>2</sub> Allowances (August 1994-December 2001)</b>		
Market Index	Beta coefficient	Standard error
S&P 500	-0.098	0.258
NYSE	-0.219	0.228
NASDAQ	-0.039	0.104

In all cases  $\beta$  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

<sup>24</sup> 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- $\tau$  path, reflecting year-to-year variations in demand, but the average annual increase in price would be less than the discount rate and as determined by the interaction between the rate of growth of counterfactual emissions and the marginal abatement cost curve.

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. With a beta of zero, there is no *undiversifiable* risk associated with holding SO<sub>2</sub> allowances and the appropriate discount rate is the risk-free rate.

This result, which will strike most 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 other form of temporal flexibility in emissions trading, is efficient.

Two factors explain this result. First, the beta associated with producing electricity is very low. Equity betas are typically around 0.5 for regulated electric utilities and 1.0 for unregulated producers of electricity, who are more highly leveraged. When the observed equity betas for these two types of electricity producers are un-leveraged to account for the equity risk associated with varying debt levels, the resulting asset betas are similar, 0.2, which is low, although not zero.

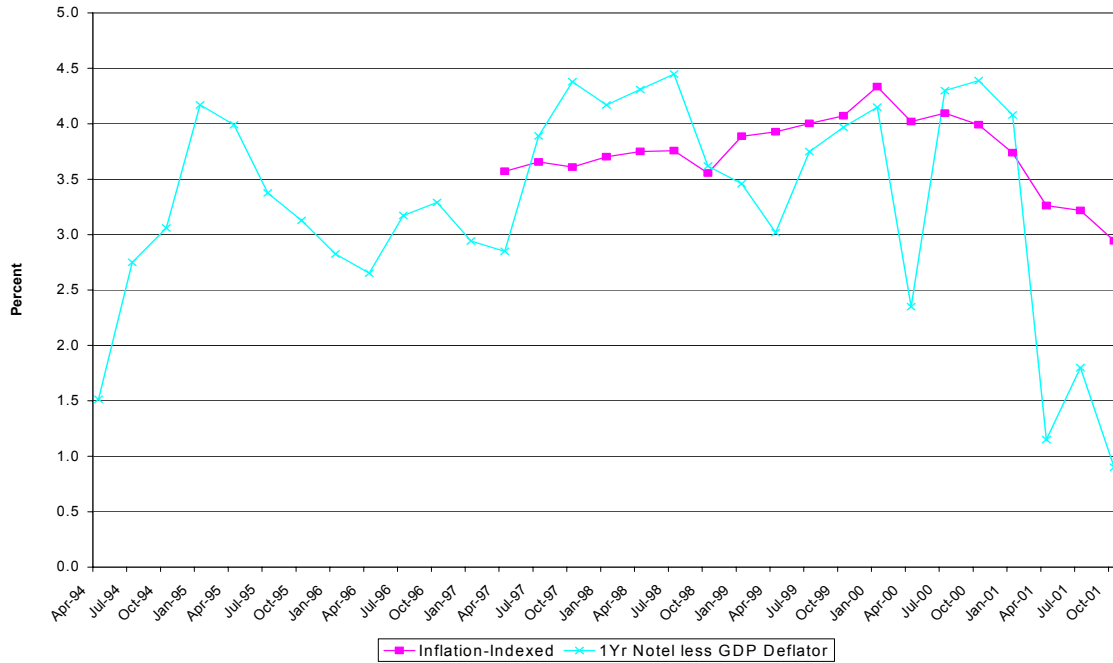
Second, and perhaps more importantly, the factors determining allowance prices are considerably different from those determining profits from generating electricity, not to mention the profits of the corporate sector as a whole. The profits of electricity producers will be influenced mostly by the price 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 only one of several factors determining counterfactual emissions. More important factors for the latter are the *relative* prices of fuels of differing sulfur content and non-Title-IV regulatory requirements affecting SO<sub>2</sub> emissions. Whatever the effect of these factors on the profits of the owners of electricity generating assets, the effect on equity returns for the market as a whole is negligible. Thus, it is not surprising that the returns from holding SO<sub>2</sub> allowances are uncorrelated with market returns and that the *beta* for SO<sub>2</sub> allowances is zero.

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.<sup>25</sup> For estimating the risk-free, real rate for years prior to 1997, we use the one-year Treasury note less the inflation rate over the past year as indicated by the GDP deflator. Figure 3 shows the real risk-free rate determined by both of these methods.

**Figure 3: Risk-free Rates**



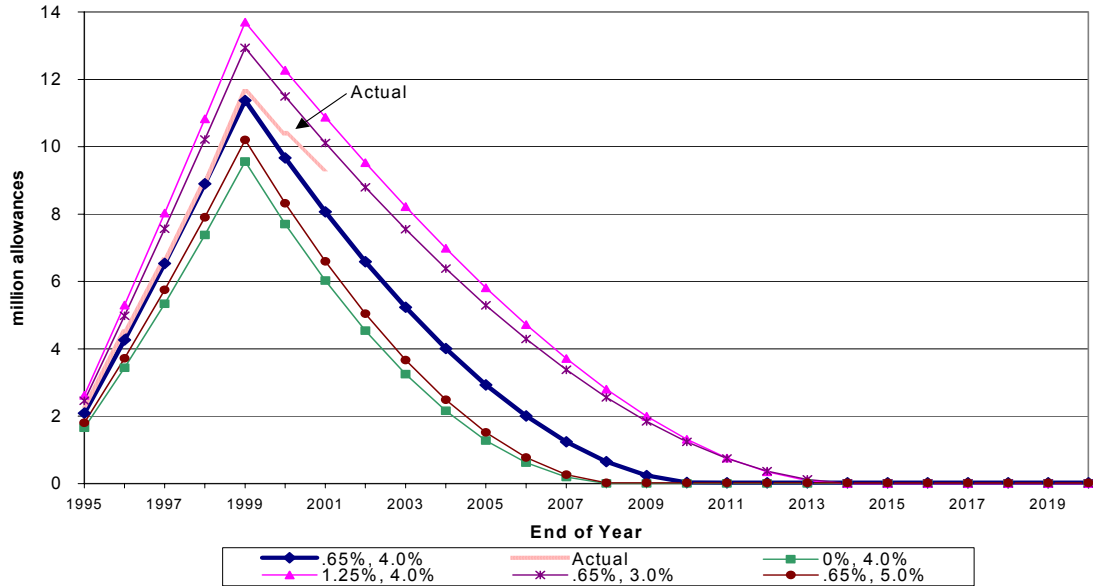
For the period prior to 1997, the real risk-free rate fluctuates between 1.5% in early 1994 to about 4.0% in early 1995 and averages 3.1% for the period 1994-1996. From 1997 through 2000, the period when the two methods overlap, similar rates are indicated: an average of 3.9% by the inflation-indexed notes and 3.7% by the one-year note less the past year’s rate of inflation. The two methods diverge significantly at the end of 2001, but lower rate reflects the steep forward curve prevailing at this time because of actions taken by the Federal Reserve to inject liquidity into the U.S. economy. For the simulations in this paper, we assume a range from 3.0% to 5.0%. The real risk-free rate for the appropriate maturity seems have been between 3% and 4% for most of the period and we use a 5% rate to illustrate the effect of using a higher than the risk-free rate.

<sup>25</sup> In particular, we use inflation-indexed notes with maturities in 2007-2008 to span most of the likely banking period.

#### 4. Is the Evolution of the SO<sub>2</sub> Bank Efficient?

A convenient means of depicting alternative banking programs is to plot the accumulation and draw down of banked allowances at the end of each calendar year, as is done in Figure 4.

Figure 4: Optimal Banking Programs

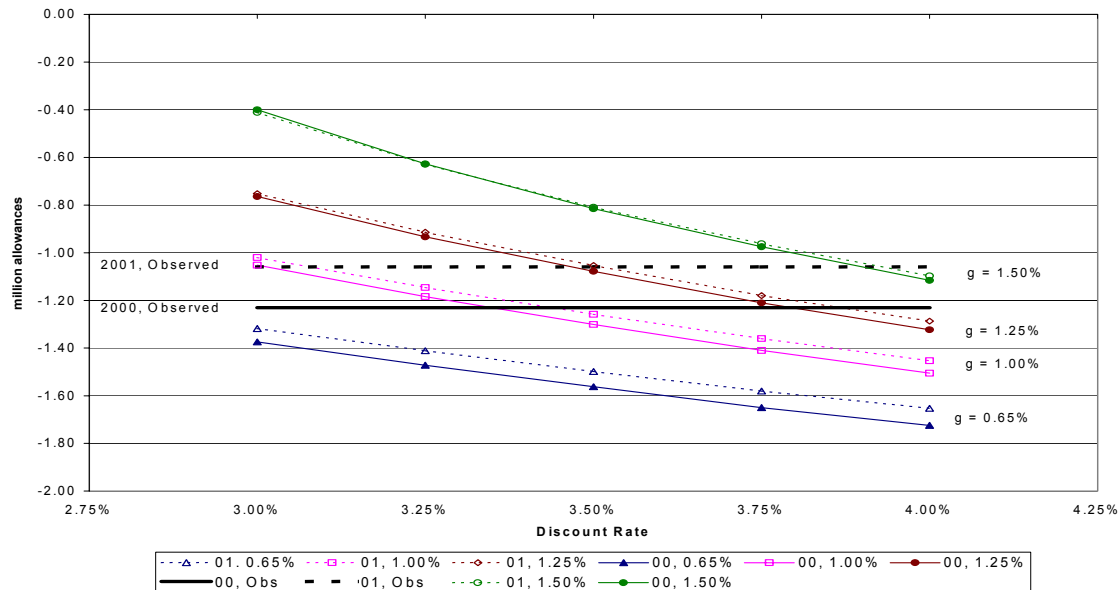


The shaded, fuzzy line tracks the actual evolution of the SO<sub>2</sub> allowance bank through 2001. All the other lines indicate efficient banking programs with plausible assumptions about discount rates and growth in counterfactual emissions as of 1995. The bold line in the center, which closely tracks 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 (smaller) banking programs reflecting higher (lower) expected costs without banking because of higher (lower) rates of growth in counterfactual emissions or a higher (lower) valuation of future costs relative to the present reflecting lower (higher) discount rates.

The efficiency of the observed banking path can be assessed by comparison with these simulated banking programs either directly, by reference to a best estimate of the parameter values that would have prevailed at the beginning of the banking period, or inversely by evaluating whether the parameter values that replicate the observed banking path are reasonable. It is readily apparent in Figure 4 that the actual banking path during Phase I implies a discount rate of about 4.0% and growth of counterfactual emissions of 0.65%. These are approximately correct values; and, if anything, the real discount rate in 1995 might have been closer to 3% than to 4%, which would imply that there has been too little banking. These paths also assume that the single values of  $g$  and  $r$  representing expectations at the start of Phase I remained unchanged in the succeeding years. During Phase I, assumptions about both  $g$  and  $r$  would likely have increased somewhat, but the effects of each would have been off-setting, thereby imparting more consistency to the observed banking path than was true of the underlying parameters.

The inverse approach is helpful in evaluating the apparent departure in 2000-01 from the constant 4.0%-0.65% efficient path that fits the 1995-99 data. Figure 5 plots draw-down rates on the vertical axis against real discount rates on the horizontal axis for different assumptions about the growth in counterfactual emissions.

Figure 5: Title IV Banking Drawdown Rates



The bold, horizontal lines in Figure 5 indicate the observed draw-down rates in 2000 (solid line) and 2001 (dashed line) and the downward sloping lines map the points corresponding to efficient programs for the indicated values for 2000 and 2001 given the initial banks in those years. Intersections of the horizontal lines with the downward sloping lines indicate combinations of parameter values characterizing efficient banking programs. For example, the observed 2000 draw-down is consistent with the following pairs ( $g = 1.00\%$ ,  $r = 3.35\%$ ) and ( $g = 1.25\%$ ,  $r = 3.80\%$ ). If the real discount rate in late 1999 and early 2000 was between 3.75% and 4.00%, as seems likely, the observed 2000 draw-down is consistent with assumptions concerning  $g$  ranging from about 1.20% and 1.35%. As discussed earlier, the mean expectation concerning growth in the counterfactual emissions is more likely to have been close to these numbers than to the value of 0.65% that represents a reasonable reflection of expectations in 1995. Similarly, the observed 2001 draw-down is consistent with the pairs ( $g = 1.00\%$ ,  $r = 3.10\%$ ) and ( $g = 1.25\%$ ,  $r = 3.50\%$ ). Given the decline in the real discount rate during 2001, and perhaps also in expectations concerning the growth in counterfactual emissions, these combinations of  $g$  and  $r$  cannot be dismissed as implausible or clearly irrational.

## 5. Discussion of Results

The finding of reasonably efficient SO<sub>2</sub> allowance banking is surprising in view of the common perception that, if anything, there has been too much banking during Phase I. Since we shared this point of view prior to conducting this research, an explanation of the misperception helps to clarify these results and the source of the misunderstanding.

### 5.1. Risk-free Discount Rate

The most obvious explanation for the discrepancy is the discount rate. The *beta* for SO<sub>2</sub> allowances could not have been estimated in the years preceding the start of Phase I because no market existed and there were no analogous financial assets with which comparison might have been made. Most electric utilities probably used a higher

discount rate appropriate for evaluating investments in electricity generation.<sup>26</sup> Whether utilities used a project-specific discount rate reflecting the asset beta for electricity generation or a weighted average cost of capital, the rate would have been higher than the risk-free rate. We are not aware of any published analysis of banking using a higher than risk-free rate, but anyone making back-of-the-envelope calculations using such a rate would have concluded that SO<sub>2</sub> allowance banking was excessive.

## 5.2. Initial Error and Irreversibility

Another reason for the perception of excessive banking is the unexpected and persistent fall in SO<sub>2</sub> allowance prices as Phase I got underway and a concomitant belief that much of the abatement undertaken then was irreversible. The allowance price revealed by the first EPA auction in March 1993, \$131, was lower than the few trades then reported at prices ranging between \$250 and \$300 and an informed estimate of \$250 (EPRI, 1993). More importantly, as the Phase I requirements took effect beginning in 1995, allowance prices fell even more to an all-time low, slightly under \$70, in early 1996. Almost half of the abatement in the first years of Phase I resulted from investments in scrubbers, which require significant lead-times in construction and are irreversible for long periods of time. Moreover, some of the switching to low-sulfur coal involved long term contracts, which would have contributed additional elements of irreversibility.

While an initial error in expectations and significant irreversibility in abatement cannot be doubted, the critical issue is the extent to which the irreversibility would have led to a larger bank. Excessive banking would occur only if most of the abatement undertaken at the beginning of Phase I had been irreversible for the duration of Phase I. If agents had sufficient ability to adapt to lower prices, that is, if a sufficient proportion of the initial commitments to abatement were reversible in the course of Phase I, and especially in 1995 as prices were falling, the amount of excess banking would not have been very large.

The effect of sufficient ability to adapt to lower prices can be illustrated by the simulation reported in Table 2 that assumes irreversibility is completely absent. These results show the effect of an initial counterfactual,  $u(0)$ , that is 10% and 20% above and

---

<sup>26</sup> For instance, a real discount rate of 6% was used in EPRI (1993) for evaluating investments in scrubbers.

below the base case on allowance price, abatement, allowances banked, and emissions in the first year of the banking program, while assumptions about the discount rate, the expected growth in counterfactual emissions, and the cost function are held constant.

<b>Table 2: Effect of Variations in the Level of Initial Counterfactual Emissions on 1995 Price and Quantities</b>				
	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. Since abatement in the accumulation phase of a banking program can be divided into that required for achieving the cap without banking and the additional abatement for banking, it is obvious that, when irreversibility is absent, all of the adjustment will be made in the former. The amount banked remains relatively unchanged since the different counterfactual assumptions affect the required reductions in Phase I and Phase II more or less equally and leave the difference in marginal cost—which determines the amount of banking—relatively unchanged.

Whatever the level of the counterfactual, and therefore of the abatement to meet the cap, efficient banking would always call for more abatement in Phase I to equalize expected marginal costs between Phases I and II, even when the cap is initially non-binding (zero marginal cost), as it would be in the minus 20% case in Table 2. In this instance, all the abatement is undertaken for the purpose of banking, and the amount

banked and the initial emissions level would be only slightly changed from what they would be with an initially binding cap.

The clear implication of this simulation, when considered together with the evidence of efficient banking, is that the responses to the Acid Rain Program have been characterized by less irreversibility than has been commonly assumed. Further support for 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 3.

<b>Table 3: Emission Reduction due to Scrubbing and Switching (Phase I units only)</b>		
	Scrubbing (million tons)	Switching (million tons)
1995	1.77	2.12
1996	1.88	2.15
1997	1.95	2.23
1998	1.94	2.40
1999	1.85	2.67
2000	2.03	3.28
2001	2.05	3.20

Switching, the form of abatement requiring the least lead time and having the least irreversibility, increased by about 50%, or slightly more than a million 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. Table 2 indicates that a 10% variation in the level of the initial counterfactual (about 1.5 million tons of SO<sub>2</sub>) translates into about 0.9 million tons of initial abatement. The magnitude of the initial error in expectations is not knowable, 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 cancelled abatement by switching was restored.

### **5.3. Technological Change**

A final explanation for the belief that there has been too much banking in SO<sub>2</sub> allowance trading program concerns technological change. It can be easily shown using

the equations in Section 2 that the effect of including some positive rate of continuous cost diminution is the same as an equivalent increase in the discount rate. If the impact of technological change on costs is to reduce  $\alpha_i$  at a rate of  $\delta$  per period, the new efficient banking paths can be obtained by simply replacing the original discount rate  $\rho$  by  $\rho + \delta$ .

Although it is evident now that the cost of scrubbing has diminished over the past decade, it is not clear that this improvement in abatement costs was expected, much less that it was included in banking calculations. There is, for instance, no record of studies discussing the effects of technological change on the costs of the Acid Rain Program, much less on banking behavior. Even if expected, the effect on banking would have depended on whether the reduction in scrubber cost was a one-time event or a continuing process. If it is a one-time event, the aggregate cost function is shifted downward in both Phases I and II and the effect is the same as that of the error in initial counterfactual emissions just discussed: current abatement and price change, but not the amount of banking. The banking program would change only if future, as yet unrealized, cost reductions are expected to reduce the difference between marginal costs in Phase I and Phase II.

Whether some rate of cost diminution should now be included, and if so, what rates, are good questions. It seems unlikely that all the cost diminution in abatement has been exhausted, but it also seems evident from the inverse calculations in Figure 5 that technological change is not being included in banking calculations, perhaps inappropriately.

## **6. Conclusion**

The results of this evaluation of the temporal efficiency of SO<sub>2</sub> allowance banking are both reassuring and surprising. The results are reassuring in affirming once again that properly constructed markets produce good results. The surprise arises from the widespread perception of excess banking among participants and analysts in this market.

Since we shared this misperception, a good part of this paper (and an even greater part of its preparation) has been devoted to explaining it. Two assumptions seem to explain the error. The first is the belief that the appropriate discount rate for holding SO<sub>2</sub>



allowances is that for investments in electricity generating capacity, which would be higher than the rate appropriate for the zero-beta asset that we find SO<sub>2</sub> allowances to be. The second assumption concerns the effect of an initial over-commitment to abatement on banking. Observers appear to have assumed that initial abatement was more irreversible than it was in fact, and they failed to understand the extent to which the adjustment to this error affects current abatement instead of the quantity banked.

The results of this paper should not be read as asserting that SO<sub>2</sub> 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. The aggregate behavior of the SO<sub>2</sub> bank indicates that most agents have made reasonably efficient abatement decisions during Phase I and the first two years of Phase II.<sup>27</sup> Perhaps, this conclusion should not be surprising. SO<sub>2</sub> allowances are financial assets and agents should be expected to treat them accordingly, despite their novelty and peculiar attributes.

---

<sup>27</sup> Montero (2002) shows that when some agents do not fully participate in the market, or are able to exercise market power, the evolution of the actual bank would noticeably differ from the evolution of an efficient bank.

## References

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<sub>2</sub> Compliance and Allowance Trading: Developments and Outlook*, EPRI TR-107897 (April 1997).

Ellerman, A. Denny and Paul L. Joskow, “To retrofit or to replace?”, MIT/CEEPR Working Paper (forthcoming).

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.

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.

Montero, Juan-Pablo, “Testing the Efficiency of a Tradable Permits Market,” MIT/CEEPR Working Paper WP 0200x (August 2002).

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<sub>2</sub> 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).