

# THE US SO<sub>2</sub> ALLOWANCE TRADING PROGRAM

BY A. DENNY ELLERMAN

Massachusetts Institute of Technology (\*)

*The US SO<sub>2</sub> Allowance Trading Program is the world's first large-scale application of a cap-and-trade mechanism for limiting emissions, and it is often cited as an example for the control of other pollutants and of greenhouse gases. Drawing upon experience with this novel approach to emissions control since 1995, this article makes five observations that address common misunderstandings about emissions trading and that are applicable to the control of greenhouse gases. First, emissions trading did not compromise environmental effectiveness, and even enhanced it. Second, the program works because of the simplicity of the compliance requirement, the unavoidably strict accountability of the system, and the complete flexibility given to emitting sources. All three go together to form what may be regarded as a virtuous circle. Third, despite fears to the contrary, allowance markets developed in response to trading opportunities. Fourth, the politics of allowance allocation can be helpful in overcoming objections to emission control measures. Finally, provisions for voluntary accession present problems of moral hazard that must be carefully considered.*

Few discussions of emissions trading get very far without reference to the U.S. SO<sub>2</sub> allowance trading program. This program is the world's first large-scale application of a cap-and-trade system for addressing an environmental problem, and it has worked exceedingly well, surpassing even proponents' expectations. As such, it presents an attractive alternative to the usual command-and-control approach to environmental problems, and it has become a standard for such alternatives. More importantly, the U.S. experience has revealed other attributes, besides the conventional argument about cost savings, that make this instrument attractive.

Of course, the U.S. SO<sub>2</sub> program cannot be applied blindly to other environmental problems. Every problem is different and the

institutional circumstances in which the program is implemented can differ significantly between countries and between the national and international level. A full discussion of the applicability of this program to international GHG emissions trading is beyond the scope of this paper, but some observations of broad applicability in the design of national GHG emissions trading systems can be made. The five observations offered below are directed both at common misunderstandings about emissions trading and aspects of the program that are particularly relevant for GHG emissions trading. Readers interested in more detail about U.S. SO<sub>2</sub> emissions trading are referred to the recently published *Markets for Clean Air: The U.S. Acid Rain Program* (Ellerman et. al., 2000).

## A – Emissions trading does not compromise environmental effectiveness

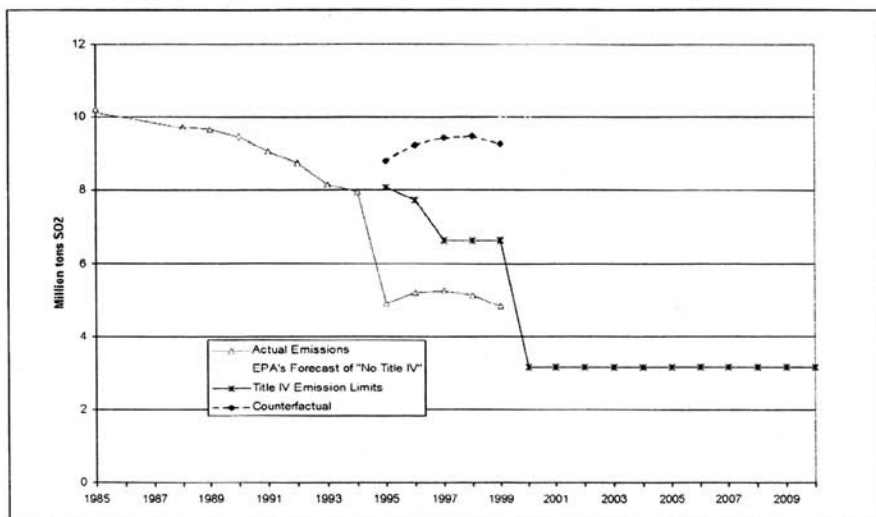
Perhaps the greatest misunderstanding about emissions trading, especially in the international arena, is the perception that it allows polluters to evade their emission reduction commitments. In part, this criticism rests on differing conceptions of the emission reduction requirement. If the goal is that each affected party reduce emissions

(\*) The author is senior lecturer, Sloan School of Management, and executive director, Joint Program on the Science and Policy of Global Change, both at the Massachusetts Institute of Technology.

in a specified manner, then the criticism is valid. Under a cap-and-trade system, complete flexibility is afforded affected firms and any single source may choose not to reduce at all. The catch is that for every source not reducing emissions other sources must reduce more. Alternatively, if a broader view of the emission reduction commitment is adopted – namely, that aggregate emissions matter, not emissions from individual sources – the criticism that polluters are able to evade the requirement is unfounded. The U.S. SO<sub>2</sub> emissions trading program has demonstrated decisively that a tight, binding cap will reduce the relevant emissions, even though who, where, and how are left completely open. In fact, SO<sub>2</sub> emissions have been reduced far more than required and more than expected in the first transitional phase of the program, as illustrated in Figure 1.

The solid line beginning in 1985 and continuing through 1999 indicates actual total SO<sub>2</sub> emissions from the 374 generating units that were subject to the cap during the first five years of the program. SO<sub>2</sub> emissions had been falling regularly before the acid rain program became effective in 1995, but the reduction of emissions in that first year was sharp and unprecedented. No one doubts that the cause was the SO<sub>2</sub> emissions trading program. The solid line beginning in 1995 and extending to 2010 indicates the cap, the number of allowances distributed to these units for the first fifteen years of the program (1). Because of some special distributions of allowances, the cap was not particularly binding in the first two years of the program, yet emissions were reduced well beyond what was required in this and later years. The dotted line at the top indicates the baseline estimate of emissions from these units when the legislation was passed in 1990 and the dashed line provides a counterfactual estimate of what emissions would have been in the first five years given actual electricity demand in Phase I. The cumulative reduction of emissions in these

**FIGURE 1 - U.S. SO<sub>2</sub> Emissions Trading : Emissions, Caps, and Counterfactuals**



first five years was approximately 20 million tons of which slightly more than half exceeded what would have been required to meet the cap in these years. This « excess » reduction has created « banked » allowances that will be used to smooth the transition to the lower Phase II cap by covering emissions greater than the annual allocations in the first years of Phase II. Thus, « over-compliance » in Phase I will be offset by « under-compliance » in Phase II ; however, if earlier reductions are preferred to later reductions, the over-compliance in Phase I has brought an extra environmental benefit.

The more important aspect of environmental effectiveness is not the temporal acceleration of the required emission reductions, but the absence of exemptions and exceptions. Inevitably, the implementation of standards and regulations involve exemptions and exceptions that recognize unique circumstances at a unit that would result in undue hardship if the standard were to be applied uniformly. In many cases, such exception is equitable, but the special dispensations invariably detract from the environmental effectiveness of the prescribed standard. The problem is not so much the exceptions, but the lack of incentive to do more than is required where meeting the

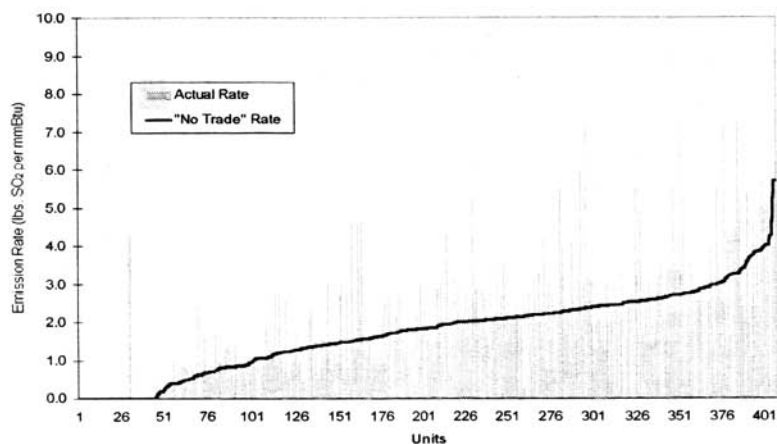
standard is relatively cheap. As a result, all the deviations from the standard are in the direction of loosening the requirement. In contrast, tight caps ensure that deviations in one direction are offset by deviations in the other direction. The flexibility afforded every source is one reason, but it is also true that no firm can claim undue hardship due to unique circumstance. With a market for allowances, the cost of an allowance represents the greatest hardship to be endured, and in a market with many buyers, none will be unique.

Figure 2 illustrates the extent to which the deviations in the SO<sub>2</sub> emissions trading program were offsetting.

The columns represent observed emission rates and the bold line rising from the left to the right shows the emission rates that would

(1) The first five years, known as Phase I, form a transitional phase during which an intermediate cap, allowing 2.5 lbs. SO<sub>2</sub> per million Btu of average 1985-87 heat input, was placed on the largest and most highly emitting generating units. A lower permanent cap begins in 2000. Known as Phase II, it covers all generating units and allows 1.2 lbs. SO<sub>2</sub> per mmmBtu of average 1985-87 heat input.

FIGURE 2 - SO<sub>2</sub> Emissions « Allowed » and Actual, 1998



have been observed at each affected unit in 1998 if every unit required to reduce emissions to meet its allowance allocation had done so without any trading (2). Columns above the bold line indicate units that acquired permits from other units to cover emissions greater than the quantity of allowances allocated to that unit. Without trading, these units would have been candidates for exception or exemption. Columns below the bold line indicate units that reduced more than necessary to be within the unit's allowance allocation. Without trading, some of these might have reduced emissions more than required, but certainly not in the large number or large amounts observed here.

The arguments for emissions trading are always based on cost-savings, and in an ideal, textbook world in which parties complied with environmental requirements however onerous they might be in individual cases, cost savings would be the only attraction of a cap-and-trade system. But, in an imperfect real world, in which equitable exception occurs, equal environmental effectiveness cannot be taken for granted. As noted in *Markets for Clean Air*, no other provision in three decades of experience regulating air emissions in the United States has been as effective as the SO<sub>2</sub> cap-and-trade program in reducing and limiting emissions. As a

result, economists entranced with cost-saving possibilities are no longer the sole champions of emissions trading; an increasing number of environmentalists have come to advocate cap-and-trade systems as more environmentally effective than the traditional command-and-control alternatives.

### B – Simplicity, accountability, and flexibility go together

The U.S. SO<sub>2</sub> emissions program is remarkable for the complete flexibility given affected sources. What is not so well appreciated is that this flexibility is made possible by a level of accountability that is far stricter than what is usually applied to environmental regulations. The requirement imposed upon emitters could not be simpler: to hold a valid permit to be surrendered for each ton of SO<sub>2</sub> emitted. Such singular simplicity both requires strict accountability and makes strict accountability possible. When nothing matters other than surrendering an allowance for each ton of emissions, there is no other basis for judging compliance than this simple one-to-one correspondence (3). Not only does strict accountability result by default, but it also becomes eminently practicable when no complicating conditions introduce admini-

nistrative discretion into the compliance decision. And, with such strict accountability in a well-designed program, the regulator can be very relaxed about whether and how individual emitters reduce emissions. Like the banker who cares not for what purpose a check is written but only that sufficient money is in the account, the environmental regulator need not ask how an emission reduction was accomplished, or even whether, but look only to see that there is an allowance to cover the debit.

This relationship between simplicity, accountability and flexibility also helps to explain the environmental effectiveness of the SO<sub>2</sub> emissions trading system. The simplicity removes wiggle-room, and the resulting strict accountability means that flexibility will have no effect on the cap. More complicated, tailored rules diminish accountability and make flexibility less possible if loopholes are to be avoided. The lesson that emerges from this experiment – that simplicity encourages accountability, and strict accountability enables flexibility – should have wide applicability.

### C – Allowance markets will develop

One of the surprises about the SO<sub>2</sub> emissions trading program was the rapidity with

(2) The extent to which this line departs from the 2.5 lbs./mmBtu standard used to allocate allowances indicates the extent to which heat input at these units has changed since 1985-87. Heat input has been reduced for units on the right hand side that could have emitted at above 2.5 lbs./mmBtu, while many on the left hand side have increased heat input. Other units to the left of the graph had either switched to a low sulfur coal prior to 1995 or shut down (those on the extreme left) and were therefore not constrained by the allowance allocation.

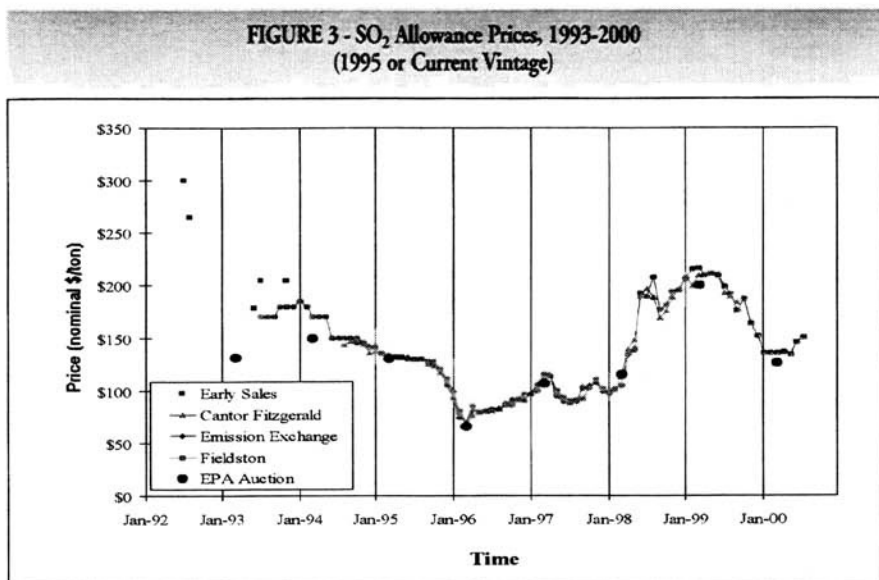
(3) Emissions must be measured so that there are other requirements pertaining to the monitoring and reporting of emissions. In the U.S. SO<sub>2</sub> program, each stack was required to install a continuous emissions monitoring system and to report the results to EPA. It should be noted that the measurement of emissions as the basis of compliance is unusual in environmental regulation where more typically the requirement is to install a particular piece of equipment or to engage in certain practices.

which a market for allowances developed. Proponents of the program were not optimistic about the development of an allowance market mostly because the recipients of allowances, electric utilities, were viewed as unlikely traders. Since the cost-based regulation under which utility executives operated did not reward the risk-taking involved in trading allowances on an external market, the general expectation was that utility traders would avoid such risk and limit their trading to what could be conducted internally between units owned by the utility.

As shown in Figure 3, early allowance price indications showed a high degree of dispersion in keeping with this expectation; however, by mid-1994, about six months before the program went into effect, price indications from various sources converged. Thereafter, the law of one price, indicating a workable market, has existed. The price has displayed a fair degree of volatility over time, but there has always been one price at any one point in time.

Participants in this market were not limited to the electric utilities receiving allowances. The differences in marginal cost between utilities caused brokers to attempt to arbitrage these spreads, and the inevitable volatility created incentives for market makers and other speculators to provide hedging instruments to risk-averse utilities and to profit from the price differences through time. Other unexpected participants were upstream coal suppliers, who often bundled allowances with coal supplies, and environmental organizations that retired a small number of allowances from circulation.

Experience with the U.S. SO<sub>2</sub> emissions trading program encourages the belief that markets will emerge when the need exists; however, the nature of the allowance and the rules of the system also make a difference. In the U.S. SO<sub>2</sub> program, the absence of any requirement that the regulator review individual trades greatly facilita-



ted trading and the emergence of an external market. It is also important that the permits be viewed as reasonably secure and durable property rights, the value of which will not be altered arbitrarily by government fiat. In the U.S. SO<sub>2</sub> program, the allocation of allowances to firms up to thirty years in advance of when they could be used for compliance reinforced the sense that allowances were property rights and provided much more liquidity to the market than if the government had placed the allowances in accounts only a year or two ahead of time. This feature of the allocation also encouraged investment in capital-intensive, deep reduction technology (scrubbers) the cost of which would be recovered over a number of years. Investors knew up front how many allowances they would have available for use elsewhere or for sale to others over the life of the scrubber. When sold, the sale of such streams of allowances provided cash to reduce the financing needs for the initial capital investment while also fostering the development of the market. Finally, the ability and incentive to make such investments contributed to the over-compliance in Phase I that has made the program so attractive from an environmental point of view.

#### D – The politics of allowance allocation can be helpful

Quite aside from the environmental and economic merits of cap-and-trade emissions trading, the U.S. SO<sub>2</sub> experience suggests that the property rights created by these systems facilitate agreement on the enactment of the environmental programs. The proposal for allowance-based emissions trading broke what had been a decade-long stalemate on acid rain legislation during which environmental advocates of the conventional command-and-control measures had been defeated. The support of a new Republican administration was probably the decisive factor in breaking the stalemate, but the creation and grandfathering of the allowances to those on whom the new emission reduction requirement was to be imposed held an undeniable attraction. The granting of allowances was never called compensation and the incumbents would have received rights to emit under conventional regulation, but these new rights were more explicit, more secure, and perhaps most importantly tradable, which implied that the rights were no longer inseparable from and capitalized in the value of the emitting asset. A further advantage was that the allowances were easily divisible and allo-

cable to address equitable concerns raised in the legislative debate and thereby to gain votes for enacting the legislation.

The legislative allocation of allowances was also important in achieving a degree of finality that would not have otherwise occurred. In earlier environmental legislation, Congress had indicated that certain goals were to be achieved and had instructed the executive, or appropriate expert agency, to figure out how to implement the program to achieve those goals. The executive agency's rulings concerning the conditions for continuing to emit (the command-and-control equivalent of allowance allocation) were usually viewed by those on the receiving end as arbitrary and capricious with the result that, at least in the American system, implementation is tied up in litigation for years. The process by which allowances were allocated by the legislature was no more inspiring than is the case for tax or other legislation, but finality was achieved, if for no other reason that in the American system the grounds for suing Congress are many fewer than for a mere executive agency. Moreover, to the extent that legislatures are better at resolving genuine issues of equitable treatment than bureaucracies, the grounds for contesting implementation of the program were lessened.

Finally, despite all the bargaining with allowances in the legislature, there is no indication that the resulting allocations affected the efficiency of the market outcome. A nearly perfect Coasian separation of equity and efficiency appears to have been achieved. The process of allocation may not have been elegant, but it did deal with all the politics up front and thereby let implementation proceed more or less as a technical matter. And, having been endowed by valuable rights in the process, those most affected had a distinct interest in making sure that the system worked and that its integrity was maintained so that the value of the newly created rights could be realized.

## E – Opt-in provisions are tricky

The U.S. SO<sub>2</sub> emissions trading program has also revealed a troubling but not overwhelming problem with a frequent feature of cap-and-trade programs, opt-in provisions. An opt-in provision allows a source of emissions outside of the cap to volunteer to receive allowances and to become part of the cap. In theory, firms outside the cap with relatively low abatement costs would seek to join in order to exploit their cost advantage in abatement, thereby reducing costs and extending the cap. In practice, the U.S. experience has demonstrated that opt-in provisions contain an element of moral hazard.

The problem arises from the practical impossibility of setting a baseline for allowance allocation that will correspond exactly with what the volunteer's emissions would have been if it were not to opt-in. If the baseline is too stringent, low cost abaters will be discouraged from joining since doing so will impose some uncompensated costs upon them, and the purpose of the provision will be in part defeated. If the baseline is too lax, excess permits (« hot air » in Kyoto-speak) will be created, sources offering no low cost abatement will join to obtain the freebies, and the cap will be inflated to the detriment of the environmental objective.

Moral hazard enters as a result of the unavoidable lag between allocation and implementation and the effects of continuing change in the economy on the opt-in candidate. Allowances cannot be allocated simultaneously with implementation, when the regulator would know the demand placed on the unit and its immediately preceding emission rate. Instead, allowances must be issued several years ahead based either on some historical basis or a prediction of what is expected. In the interval, the economy will change and affect opt-in candidates in ways that will cause the proposed allocation to any given unit to be either too

stringent or too lax. Those for which intervening change has worked to create excess allowances will opt-in, even though they may offer little low cost abatement, while those for which intervening change has made the proposed allocation more stringent will be discouraged, even though they could offer relatively low cost abatement. The same factors work to tighten and loosen the effective requirement on mandated units, but with no opting out or in, the opposing effects offset each other.

In the case of the U.S. emissions trading program, the number of generating units that volunteered for the Phase I cap was unexpectedly large. Depending on the year, the number of voluntary units was half to three-quarters as many as those mandated to be subject to Phase I, and their entry expanded the scope of the cap by about 20 % from what it would have been without the opt-in provision. More importantly, the allowances distributed to these units were about 23 % *above* a reasonable *ex post* estimate of what aggregate emissions from the voluntary units would have been. In contrast, the number of allowances distributed to the units mandated to be in Phase I was about 24 % *below ex post* estimates of their aggregate emissions. In all, about one million tons of excess allowances were distributed to opt-in units during Phase I or about 3 % of the 37 million allowances issued to all units during the five years from 1995 through 1999. As a result, the aggregate, cumulative cap on SO<sub>2</sub> emissions has been inflated slightly. At the same time, these opt-in units reduced emissions by about 1.2 million tons in response to the incentives provided by participation in the cap-and-trade program so that there has been some savings in abatement costs.

The implications of this aspect of the experience with the US SO<sub>2</sub> emissions trading program are mixed. Purists will likely argue that the moral hazard is unavoidable and the risks of cap inflation too great to allow inclusion of opt-in provisions. Pragmatists

will argue that rules can be devised to minimize the selection bias and that the benefits of reducing costs and expanding the cap outweigh the small adverse effects of moral hazard in a well-designed opt-in provision (4). In the U.S. program, it can be argued both that little was gained by the opt-in provisions since virtually all the participants would soon become subject to the law's provisions in Phase II and that the damage was slight when the million ton inflation of the cumulative ceiling is compared to the 90 million tons that will be allowed during the first ten years of Phase II, when these excess allowances will be used.

In the case of GHG emissions, where caps will be placed on CO<sub>2</sub> first because of measurement problems and on industrialized nations first because of their greater ability and willingness to pay, extending the cap to cover other gases and other countries is far more important in achieving the ultimate environmental goal than was extending the cap during Phase I of the US SO<sub>2</sub> emissions

trading program. In fact, the hot air embedded in the Kyoto Protocol can be seen as an inducement for the countries of Eastern Europe and the former Soviet Union to develop acceptable inventories and measurement protocols that are the necessary precondition for voluntarily undertaking GHG emissions abatement and participating in global emissions trading.

Architects of GHG emissions trading systems, whether at the global or national level, will not have the luxury of deciding to forego opt-ins because of the moral hazard that will be encountered, as could be argued for the more limited environmental objectives of the U.S. SO<sub>2</sub> emissions trading program. Voluntary accession or opt-ins will be an essential part in achieving the environmental goal of GHG emissions cap-and-trade systems. The U.S. experience would suggest both that moral hazard cannot be ignored and that the damages in a well-designed system are not great. In the end, the damage will have to be weighed against the

benefits of extending the cap to bring in other gases and other countries ■

#### REFERENCES

ELLERMAN, A. Denny, Paul L. JOSKOW, Richard SCHMALENSEE, Juan Pablo MONTERO and Elizabeth M. BAILEY (2000), « Markets for Clean Air : The U.S. Acid Rain Program », Cambridge University Press, 362 pp.

(4) For instance, the U.S. program allowed substitution units to enter and exit from year to year. Requiring a unit to stay in once it opted in would have reduced the selection bias but not eliminated it.