ELSEVIER

Energy Policy 32 (2004) 481–491

ENERGY POLICY

www.elsevier.com/locate/enpol

The safety valve and climate policy

Henry D. Jacoby*, A. Denny Ellerman

Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

The "safety valve" is a possible addition to a cap-and-trade system of emissions regulation whereby the authority offers to sell permits in unlimited amount at a pre-set price. In this way the cost of meeting the cap can be limited. It was proposed in the US as a way to control perceived high costs of the Kyoto Protocol, and possibly as a way to shift the focus of policy from the quantity targets of the Protocol to emissions price. In international discussions, the idea emerged as a proposal for a compliance penalty. The usefulness of the safety valve depends on the conditions under which it might be introduced. For a time it might tame an overly stringent emissions target. It also can help control the price volatility during the introduction of gradually tightening one, although permit banking can ultimately serve the same function. It is unlikely to serve as a long-term feature of a cap-and-trade system, however, because of the complexity of coordinating price and quantity instruments and because it will interfere with the development of systems of international emissions trade.

Keywords: Safety valve; Cap and trade; Climate

1. Introduction

The safety valve emerged out of discussions in the United States that presumed that the Kyoto Protocol might be implemented by a system of marketable permits or cap and trade, similar to that implemented in the US Acid Rain Program. Under a marketable permit system an emissions constraint or "cap" is applied to a set of emitters, permits are distributed in this amount, and trade is allowed in these emissions rights. The central idea of the safety valve is that the cost of capping emissions at some target level can be limited by an offer from the regulatory authority to sell permits in whatever quantity is demanded at a predetermined price. Thus, if economic growth or other factors were to cause permit prices to be greater than expected, the marginal cost of abatement would be limited to the safety valve price. In the US, the establishment of a safety valve was also seen as a way of raising the likelihood of Protocol's ratification by blunting criticism that the cost of meeting the Kyoto targets would be too high. Although interest in the safety valve has subsided as a result of the Bush Administration's rejection of the Kyoto Protocol, it is likely to resurface in other proposals to place quantitative limits on greenhouse gas emissions.

The safety valve is sometimes introduced as a policy innovation, but in fact it has close and well-established relatives. It is similar to a per-unit penalty found in cap-and-trade systems where the price is set at a high enough level that it is unlikely to be triggered. And, if the price is set sufficiently low that emissions commonly exceed the quantity limit, it resembles an emissions tax. Finally, it is akin to a proposal made by Roberts and Spence (1976) in a similar context that emitters be given the choice of buying permits from the market or from the government at a specified price.

In this note we explain the origins of the safety valve concept in considerations governing the choice of policy instruments for pollution control in general, discuss closely related applications, and trace the evolution of the concept in the climate context. We then consider the role this type of hybrid instrument might play in future domestic policy of the US or other countries, and consider problems it would present within a system of international trade in emissions rights. There are a number of issues of cap-and-trade system design that we do not address, such as whether the cap should be

*Corresponding author. Tel.: +1-617-253-6609; fax: +1-617-253-9845.

E-mail address: hjacoby@mit.edu (H.D. Jacoby).

0301-4215/04/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0301-4215(03)00150-2

sectoral or economy-wide, whether permits should be grandfathered or auctioned (and in the latter case what should be done with revenues from permit sales), and whether the system should be applied upstream or downstream. Also, although the Kyoto Protocol and other proposals to mitigate climate change anticipate the inclusion of a number of greenhouse gases, we will talk mainly in terms of CO_2 denoted in tons of carbon, C. All the points made here can be extended to analysis of carbon-equivalent emissions of multiple gases.

2. Factors in policy design

2.1. The efficiency of price vs. quantity instruments

Much of the support for the safety valve as a component of a cap-and-trade system originates in concern for the economic efficiency of emissions controls, and in particular for the choice between quantity constraints and price penalties when the costs and benefits of emission controls are uncertain. Which instrument is better depends on the relative sensitivity of the costs of emissions reduction, and of the benefits (i.e., climate damage avoided by that reduction) as the level of emission control is varied (Weitzman, 1974). Under uncertainty, the better instrument is the one that is more likely to avoid a big mistake in the stringency of control imposed. For greenhouse gases, it is clear that this is the price mechanism. The argument leading to this conclusion is laid out in graphical form in Appendix A but it is sufficiently important to the safety valve discussion to justify a brief summary here.

If the relationships of costs and benefits to the level of emission control are known with certainty, the difference on efficiency grounds disappears. Either leads to the same control level and the possibility of a large error is thus assumed away. But when these relationships are known only approximately, the key to the choice is whether cost or benefit changes more rapidly as the level of emission control is varied. In the jargon of economists, it depends on whether marginal costs or marginal benefits change more with the level of emissions control. A quantity constraint is indicated if marginal benefits are more sensitive to the control level, but a tax is preferred if marginal costs change more rapidly than marginal benefits.

To demonstrate, consider a case where the damages rise steeply as the level of control is relaxed (perhaps because of some threshold effect) but the marginal control costs do not differ greatly among levels of control. In this circumstance, it is more important to get the quantity right and the price approach is more likely to lead to an inefficient outcome. Fixing the price when costs are uncertain leaves the quantity undetermined, and the response of emitters to the emissions price may turn out to produce emissions that exceed the threshold and trigger large pollution damages.

In the opposite case, damages are not very sensitive to current emissions, but the costs are very much more so. Here it is more important to get the price right and the choice of a quantity constraint is more likely to go wrong. If a quantity target is chosen, the marginal costs of control could be either much higher or much lower than would be optimal. Fixing the price in the range of marginal damages will lead emitters to undertake controls only up to that level, reducing risk of costs being way out of line from the benefits achieved.[1]

Carbon dioxide and most other anthropogenic GHGs have long residence times in the atmosphere, so the climate issue falls into this latter category. Most studies expect climate damage to rise with increasing atmospheric GHG concentrations (i.e., the stock of gases), but emissions in any particular period make only a very small contribution to the existing stock. Thus, unless the additions to the stock in a particular period have the effect of pushing the system over some threshold, as when the water level in a lake rises above the level of a dam, the marginal damage of additional emissions in any single period is essentially constant.[2] Such thresholds are hypothesized for the climate system: for example, changes that would trigger a drastic change in ocean circulations or the disintegration of Antarctic ice sheets. There is no evidence, however, that the climate system is approaching any such threshold at the present time, so the appropriate instrument for limiting GHG emissions is a price penalty on emissions.

This approach is not the one taken in the Kyoto Protocol, which imposes a quantitative emissions constraint at the national level. The safety valve was born from the collision between the desirability of using price instruments for these stock pollutants and the apparent political attractiveness of the quantity approach. If the theoretically less-desirable quantity instrument is to be chosen nonetheless, and the constraint is imposed by means of a cap-and-trade system, policymakers can limit the prospects that the cap will impose costs that are far out of line with benefits by setting a safety valve price at a level that approximates the marginal damage avoided.

[1] Note the focus is on cost uncertainty only. Conventional economic analysis of the prices vs. quantities controversy has held that benefit uncertainty is irrelevant to the choice. Stavins (1996) points out that correlation between benefits and costs can invalidate this result. In the climate case, however, it is unlikely that the correlation will be significant between costs incurred in any one nation and the damages that result from perturbation of the global system.

[2] This argument is frequently misunderstood. The point is not that the marginal benefit of emissions reduction is zero. It is not. But it is not very sensitive to the changes in emissions in any particular period.

2.2. Close relatives of the safety valve

What we have described above can be considered the "pure" safety valve concept. It applies where the regulators try to set the emissions cap at a level where the expected marginal cost of meeting the constraint will turn out to be in line with their beliefs about marginal benefits. Then, to avoid being too far wrong on the high-cost side they include a provision to sell emissions permits at some price near or somewhat above that expected cost level.[3] So, for example, if expectations of cost outcomes were symmetrical about the expected level and the safety valve price was set slightly above expected marginal cost, then the provision would have a slightly less than 50% chance of being triggered. As such, it would be a hybrid price/quantity instrument. In somewhat more than half the possible cost outcomes, marginal cost will be lower than the safety valve price and emissions would be constrained to the quantity limit. In the remainder, the safety valve price will determine marginal cost, emissions will exceed the quantity limit, and a payment will be made for the excess emissions. This pure safety valve has two close relatives that differ from it mainly in the probability that a payment will be required for exceeding the quantity limit.

2.2.1. A financial penalty far above expected marginal cost

Cap-and-trade systems often impose a penalty for uncovered emissions in the form of a per-unit fee set at a level far above expected marginal cost.[4] Although the motivation is different, this penalty is formally analogous to the pure safety valve in that (1) the quantity limit can be exceeded upon payment of the requisite fee, and (2) marginal costs and allowance prices will be no higher than the level set by the "escape" mechanism.[5] The main difference is that the penalty is set at a level that has a low probability of being invoked: the quantity constraint is binding in nearly all instances.

For example, the US Acid Rain Program initially imposed a penalty on SO_2 emissions of \$2000 per short ton:[6] a level far above estimates of expected cost when the system was being designed (Ellerman et al., 2000). Similarly, the European Commission's proposal for a EU-wide emissions trading system would impose a penalty of the higher of (1) twice the average market price in some predetermined period, or (2) 50 euros per metric ton of CO_2 equivalent during 2005–2007 or 100 euros during the First Commitment Period under the Kyoto Protocol and beyond (CEC, 2001). These prices are roughly equivalent to \$155 and \$310 per ton of carbon (tC). With the US not participating in the Kyoto Protocol, the market price within the Kyoto system as refined at Marrakech is expected to be far lower than these levels (Babiker et al., 2002; den Elzen and de Moor, 2001). Refusal of international permit exchange and other elements of flexibility would raise the cost, but even then the likelihood of this penalty being triggered is slight.

2.2.2. Permit sale at a price substantially below expected marginal cost

In contrast to a price set far above expected marginal cost of meeting a target, permits may be offered at a price far below expected marginal cost. The safety valve instrument then becomes, in effect, an emissions tax. A good example is the CO_2 emissions reduction and trading program for the Danish electricity sector (Ellerman, 2000; Pedersen et al., 2000). The essential features of this program are that (1) emissions are capped at a level that is initially about 30% below average annual emissions in 1994–1998, and they can be traded, (2) incumbents are grandfathered, and (3) the penalty for exceeding the limit is 40 Danish kroner per metric ton of CO_2 (roughly \$5.00 per ton CO_2 or \$18 per ton C). There is a quantity limit and what might seem like a penalty for exceeding that limit. However, the penalty is sufficiently low with export levels characteristic of the baseline period that the probability of payment is high.[7] The system is not so much a hard cap with tradable emissions as it is a tax with tradable exemptions. Thus two effects result from a safety valve set far below the expected marginal cost at the level of the cap: it relaxes the target emissions reduction and it effectively changes the control instrument from quantity to price.

---

[3] Logically, concern for fixing marginal cost should also include a provision limiting price variations on the downside, such as a government offer to purchase permits, but this feature has yet to be included in safety valve proposals.

[4] All regulatory systems have some type of penalty or enforcement mechanism, but only some have pre-specified, non-discretionary financial penalties that are automatically invoked when non-compliance occurs. A second level of enforcement is also provided to deal with parties not complying with the quantity limit and who also refuse to pay the penalty. Parking regulations provide a familiar analogue. Nearly all violations are discharged by the payment of pre-specified, non-discretionary fines, but refusal to pay these fines invokes harsher measures.

[5] These penalty provisions often require that allowances equal to the uncovered emissions be deducted from the next compliance period. This requirement distinguishes this application from the pure safety valve and it adds the discounted cost of the deducted permits to the escape price. Still, when the penalty price is set high relative to expected marginal cost, the main factor limiting an increase in marginal cost is the level of the penalty itself.

[6] In 1990 US\$. The penalty is escalated at the rate of inflation and is now approaching \$2800.

[7] The cap is set slightly below the level of domestic electricity consumption so that the tax applies mainly to electricity production for export to Norway and Sweden.

3. The history of the safety valve in the climate debate

3.1. Domestic and international proposals

Proposals for a safety valve in the context of the Kyoto commitments appear to have been aimed at these two targets: both avoiding excessive cost by relaxing the emissions target and moving from quantity target to a price penalty. In the US at least, the Kyoto target was widely viewed as overly stringent, and advocates of the safety valve proposed starting its implementation with "looser goals than are required by the Protocol", and with the suggestion that this loosening could be achieved through the manipulation of a safety valve price (e.g., Pizer, 1999; Kopp et al., 1998). Analyses carried out in the late 1990s estimated that the carbon price required to achieve the Kyoto targets would be in the range of \$50 to over \$200 per ton C, depending on the assumption about Annex I trading (Weyant and Hill, 1999). Meanwhile, other studies taking a longer-term perspective, with a focus on benefit-cost considerations (e.g., Nordhaus and Boyer, 2000) or cost effective approaches to atmospheric stabilization (e.g., Manne and Richels, 1999) indicated that an appropriate near-future period price was much lower, in the range of \$5 to \$14 per ton C. Although the safety valve proposal did not draw explicitly on the latter studies, the proposed safety level, \$25 per ton C (Kopp et al., 1999; Barnes, 2001), would have kept marginal costs close to the range they indicated as appropriate in the early years of greenhouse gas control.

The institution of a safety valve price was also seen as having the additional advantage of moving to a price instrument from the quantity-based approach set in the Kyoto Protocol. In a set of simulation studies considering cost uncertainty, Pizer (1997) had demonstrated the superiority of this hybrid approach to a pure quantitative target on efficiency grounds. He argued that, although the preferred tax-like instrument is not politically acceptable in the US, ". . . [with cap-and-trade and a safety valve] the advantages of a carbon tax can be achieved without the baggage accompanying an actual tax" (Pizer, 1999).

Proposals for a low safety valve price also envisaged increasing its level over time, perhaps at some point to a level sufficient to achieve the original quantity target (Kopp et al., 1998, 1999; Pizer, 1999). A number of arguments were made to highlight the advantages of such a pattern of gradually rising emissions prices, including the value of early price signals in setting expectations of needed change, and the avoidance of restrictions that force premature and expensive turnover of capital stock. Some proponents even anticipated that a safety valve price might eventually converge to a pure quantity system, as cost uncertainties were reduced with time and experience (e.g., Pizer, 1999). Along the way to such a transition the use of a safety valve would essentially serve to redirect the form of emissions control, changing the focus of regulatory decision-making from quantities to prices.

These ideas were carried over into the international discussion in the form of a proposal that compliance with the Kyoto Protocol might be met by paying, say, a \$50 per ton C "compliance penalty" (Kopp et al., 2000; Kopp et al., 2001; Hourcade and Ghersi, 2002). Any revenues collected in this process would be devoted to the purchase of emissions reductions in the second commitment period, in addition to negotiated second-period obligations. The purchases would be made by soliciting offers for project-based credits in an open auction in which all parties could participate. Again, the proposed penalty was small in relation to most estimates of the marginal cost of meeting the Kyoto targets if all of Annex I, including the US, were participating. In effect, acceptance of this proposal would have shifted the negotiations from the quantity targets to the compliance penalty, and so again it can be seen as another, and indeed creative, attempt to change the architecture of the Kyoto agreement. Further, Kopp et al. (2001) argued that this change would encourage ratification of the Protocol by all Annex I countries by overcoming opposition based on the risk of unacceptably high cost.

3.2. Questions and opposition

Despite the efficiency arguments in favor of a price instrument for controlling a stock pollutant, and for a safety valve if a quantity instrument is nonetheless chosen in the political process, the concept has met with questions in some quarters and strong opposition in others. The arguments range from a threat of being forced to do too much too soon to a worry about achieving too little.

A first set of concerns flows logically from a belief that the Kyoto targets were inappropriately stringent. Making an overly tight target more palatable by a lower safety valve price could be seen as a form of "bait and switch." The fear is that, if the stringent target remained in place, the safety valve level would be increased at a rapid pace to meet the target. Discussions of doing just that only strengthened this concern.

From this point of view, the safety valve may be questioned on the same grounds as a pure quantity approach: what logic or analysis lies behind the initial level of restriction and associated cost, and its planned evolution in the future? Economic studies differ in their conclusions about the right level of near-term stringency and especially in the appropriate path over time. The variation originates in differences in views about the net damages of climate change and non-climate benefits associated with emissions reduction, and differences in the handling of the turnover of the capital stock and the contribution of technical change. Also, studies may or may not consider the role of possible future reduction in uncertainty. But most of them agree that for a stock pollutant like the greenhouse gases an economically efficient path would start out at a low to moderate level and rise over time. While adding a safety valve to a quantity constraint embodying the sharp reductions of the Kyoto Protocol would have deferred the costs, the more fundamental objection is whether these costs are warranted at all at this early stage.

By contrast, environmental groups are opposed to the safety valve because of the potential loss of "environmental integrity." Integrity in this sense is a concept that invokes images of the sort of threshold that would justify a quantity constraint, and it is seen as put at risk in several ways. These opponents may have suspected that the safety valve was proposed not so much to support an economically efficient implementation of a target reduction as to relax it. Also, the US public's distaste for any regime that smacks of taxes is well known, and environmental advocates may have believed that such a change of focus would lead to diminished public support and doom the cap-and-trade proposal from the outset. The fear of an inevitable confounding of revenue-raising and environmental objectives is palpable in several of the position papers issued by environmental NGOs at international negotiating sessions. Finally, it is argued that provision of an "easy out", and the concomitant truncation of the distribution of cost outcomes, diminishes the pressure to innovate, and reduces the incentive to early action to build an inventory of emissions permits as a hedge against higher costs in the future.[8]

4. Potential future role of the safety valve

The policy context has changed substantially from the early days of the safety valve proposal when nations were considering full Annex B implementation of the original Kyoto targets and arguing over restrictions on emissions trading. In 2001, the US removed itself from the Kyoto Protocol, and this act greatly lowered the potential demand (and prospective price) for international emissions credits. Moreover, in subsequent negotiations the Parties relaxed the targets of some nations by the liberal allocation of carbon sinks, and abandoned any attempt to place quantitative limits on international purchases of emissions reductions. As a result, if the Annex B nations (less the US) were to make use of all the flexibility now provided in the agreement, the marginal cost of meeting the targets established by the Kyoto Protocol would be very low. In effect, a Kyoto agreement attempting large early emissions reductions, implying emissions prices over a hundred dollars per ton of carbon for some nations (e.g., Weyant and Hill, 1999), has been converted to a gradual-start agreement that could be achieved at prices in the single digits (Babiker et al., 2002).

Having rejected Kyoto, the Bush administration has set a national goal for 2012, in GDP-adjusted terms (White House, 2002), that like Kyoto would require only a small reduction below forecast baseline emissions. Only federal subsidies and voluntary reductions are proposed as yet, but even if the target were imposed through some sort of mandatory program, the carbon equivalent price would need to be only in a range below \$10 per ton C as well (Babiker et al., 2002). Some Congressional proposals would impose mandatory cap-and-trade systems that would lead to a higher price, but the Bush Administration has rejected applying mandatory measures to greenhouse emissions, at least in the period to 2012.

The future of the safety valve depends on the conditions that hold at some future time when cap-and-trade systems might take a substantial role in greenhouse gas control. Will that future be similar to the original Kyoto circumstance, wherein the targets adopted strain political credibility, or will the quantity targets be set on a gradual path of increasing stringency? Further, what will be the pressures for and against a move from a quantity based system (which underlies the adoption of cap and trade) to a regime based on price? We look at each of these possible outcomes, as they may arise in the domestic context in the US, and elsewhere. We then turn to the promise and problems of a quantity constraint with a safety valve if a regime of international emissions trading were to evolve.

4.1. Application in domestic systems

The attraction of a safety valve depends on the dangers from which it is intended to provide "safety". It is conceivable that political conditions could lead again to the setting of an emissions target imposing great short-term stringency, as many observers thought was the case for the US under the Kyoto Protocol. The implied carbon-equivalent price would be far above the range of estimates of the marginal benefit of short-term reductions, and perhaps so high as to be unattainable in

[8] Environmental Defense, a prominent NGO favoring the use of cap-and-trade instruments, has been the most articulate in voicing environmental opposition to the safety valve. The strongest statements are found in various memoranda and talking points distributed at COP-6, when a safety valve feature was being actively discussed. The titles of two of these papers convey the position: "Re: Failure of 'Ceiling Price' on Emissions Permits as a Climate Change Policy Tool," "A 'Cost Cap' Would Cost More—and Harm the Environment: Say 'NO' to the EU and Brazilian Compliance Fund Proposals." See also Aulisi et al. (2000, pp. 32–35) for a similar and more readily accessible current source.

practice. The conditions that helped stimulate the safety valve proposal in the first place would be recreated, and the proposal would have the same advantages and raise the same objections as before. Alternatively, as suggested by the current state of the Kyoto Protocol and US policy, emissions caps could start at less stringent levels, and tighten over time. The inclusion of a safety valve might still be argued, but the danger of unacceptably high prices is decidedly less.

The danger to which the safety valve provides an escape also depends on the presence of another feature common to cap-and-trade systems, permit banking, which also dampens price fluctuations. In both the US Acid Rain Program and the proposed EU Trading System (CEC, 2001), unused allowances can be banked from one period to the next, and this temporal flexibility reduces the extent to which prices will fluctuate in response to changing conditions.[9] When prices are lower than expected in some period, inventory is accumulated causing prices to be higher in that period than they would otherwise be; however, when prices are higher than expected, inventory is drawn down causing prices in that period to be lower than they otherwise would be. While banking will not have as large a price-reducing effect as a pure safety valve when costs are higher than expected, it will have a greater price-supporting effect when costs are lower than expected since safety valve proposals do not include a floor price.

Even so, problems might arise when a cap-and-trade system is first introduced. The initial periods are ones in which cost uncertainty will be greatest and a bank of permits will not yet be available, so price-dampening effects of carry-over will not operate. Moreover, the accumulation of an appropriate inventory would raise costs in the initial periods. To the degree that the initial price is a problem, an even lower initial cap or early action credits can provide protection against an unacceptably high initial price. The inclusion of borrowing would also mitigate a start-up problem, as well as enhance the price-reducing effect that comes with banking alone, although it is not a common feature of cap-and-trade systems.[10] Where these provisions are not sufficient to allay cost worries at start-up, a safety valve can be of help.

For some observers, the issue is not so much the danger of unacceptably high-cost outcomes as it is getting the control instrument right in the first place, that is, converting the control regime from emissions caps to emissions price. The objective is understandable on efficiency grounds, but it is questionable whether the implied fading away of quantity targets is likely even if the safety valve were widely applied. Consider the Kyoto Protocol with all Annex B participating as an example. Emissions reductions in response to a low safety valve price would have fallen substantially short of the original Kyoto target, and this condition might hold over many years. A $25 carbon-equivalent US safety-valve price rising at a sometimes-proposed 7% per year (Morgenstern, 2002) would not, under the cost estimates summarized by Weyant and Hill (1999), have achieved the US Kyoto target for many decades, if ever.[11]

With a hybrid policy made up of such apparently inconsistent parts, which component would be more likely to be abandoned? In our view it is the price component, or at least the idea of a price rising gradually over time. At the very least, the never-achieved quantity target and the low escape valve price would be a continuing source of controversy and conflict over the issue of "integrity". Unless accompanied with some broad agreement about the appropriate long-term path of emissions prices, and acceptance of its "emissions tax" features, the safety valve would be little more than a band-aid on an inappropriate implementation of cap and trade. Perhaps it would be a useful addition in the short run, but its adoption would leave unresolved the more important issue of the appropriate level of emissions control.

The inclusion of a safety valve feature in the recently proposed New Zealand action plan shows that the appeal of this idea is not limited to the US. This proposal would limit the cost in the first commitment period to NZ$25 per ton of CO_2 or US$40 per ton C (NZ, 2002). We have already noted the safety valve character of the Danish penalty fee at about US$18 per ton C and other non-US examples can be found. The Dutch solicitations for joint implementation credits are capped at a price of about $38 per ton C and the Australian state government of New South Wales

[9] In the EU proposal, member states may decide to limit the banking of allowances from the early action period (2005)–(2007) to the First Commitment Period under Kyoto (2008)–(2012), but not within these periods. The Kyoto Protocol allows unlimited banking and borrowing within commitment periods, but only banking between periods.

[10] Borrowing provisions require that the cap for the next period be set beforehand and that some discount rate is applied to the use of the permit in an earlier period. Indeed, cap-and-trade systems having high penalty fees with deduction of next-period allowances for uncovered emissions can be seen as a form of sanctioned borrowing with a very high discount rate. The Kyoto Protocol contains such a provision in its compliance penalty of 1.3 tons of carbon equivalent in the second commitment period for each ton exceeding the target in the first period. This feature can be seen either as borrowing or a safety valve. The marginal cost incurred in the current period need not exceed 30% of the expected marginal cost in the next period, discounted to the present.

[11] Even with unrestricted Annex B trading, the price to achieve the Kyoto targets (the US participating) was variously estimated to be in the neighborhood of $50 to $100 per ton C. By the time these levels were reached by a $25 penalty (set in 2010 and growing at 7% per year), economies would have grown, requiring still higher prices to achieve the target.

recently announced a penalty of A$15 per ton of CO_2 (US$31 per ton C) for failure to meet emission reduction targets (Reuters, 2002). These examples suggest that price rather than physical constraint could emerge as the key policy variable within an individual country. A remaining question, however, is whether such domestic systems are compatible with international trade in emissions permits.

4.2. The safety valve in a system of international permit trade

Under any multinational agreement with quantity targets, the marginal costs of control will differ among the parties. Thus, just as trade among individual emitters can lower the costs of a national emissions cap, cross-border trade can reduce the costs of an international agreement. Such exchange is provided for under the Kyoto Protocol, and the potential value of this device is indicated by the fact that some international transactions are occurring even before the Protocol has gone into force.[12] A question that needs to be addressed, therefore, is whether the existence of a safety valve in one or more of the trading partners may create barriers to international market development.

The concern is the potential for sales under a safety valve to create a new source of international "hot air" that drives out credits based on real reductions, an emissions trading variant of Gresham's Law.[13] Suppose, for example, that the market-clearing price in the international permit market rose above the safety valve price in one of the participating countries. Private agents, whether firms or individuals, could purchase permits at the government "safety valve" window and either sell them directly into the international market or use them in place of other permits transferred abroad. If permits were freely exchangeable, as would be desired for a well-functioning market, the lowest safety valve price among the trading partners would set the international emissions price. Even short of this outcome, difficulties could be created for market trading.

This problem might be controlled in several ways. First, all international transactions could be limited to government-to-government exchange of quotas. This restriction would, of course, sacrifice many of the efficiencies expected from international permit exchange. Second, if permits were devolved to private parties who were allowed to trade internationally, a web of restrictions could be imposed to attempt to prevent exchanges of hot air. All national regulations would need to forbid permit sales from any country whose safety valve level fell below the international price. In a market where the prices of traded permits are likely to fluctuate over time, and banking of permits is allowed, this approach seems clearly infeasible.

Finally, each of the countries participating in the international trading regime could impose its own cap-and-trade system and safety valve, and then agree on a common safety valve price. International permit trade would then occur only in times when the market price was below the level of the global safety valve. This idea is close to that introduced earlier of a globally agreed compliance penalty, to be paid at the time of summing up at the end of a compliance period (Kopp et al., 2001). As in a domestic implementation, evaluation of this proposal depends on the relation of the negotiated safety valve price to the expected marginal cost of meeting the negotiated caps, country by country. If the price is low enough to be triggered frequently, then in effect the agreement is a globally harmonized carbon tax, and if such a multi-nation agreement were achievable (which we think not), the whole system of negotiating quantitative emissions targets would not be needed in the first place.

It is worth noting that these problems of emissions market development in the presence of a safety valve can arise in the domestic context as well, if different systems are applied across sectors of an economy. In the US for example, proposals have been made for the imposition of cap and trade only in the electric power sector alone, perhaps with a safety valve provision (CBO, 2001). Separate proposals have been made for designing a system of trading of permits under the US system of regulating the corporate average fuel economy of motor vehicles, again with a safety valve included.[14] If these two systems, perhaps designed under separate bodies of legislation, were (as anticipated in some proposals) to be joined in a domestic trading system, then the same problems as noted above would arise, and a similar menu of corrections would have to be considered so long as a safety valve was somewhere in use.

5. Conclusions

Several generalizations can be drawn from this exploration of the safety valve concept as it might

[12] E.g., see "International Economy: Companies Agree to First Pollution Permit Swap," *Financial Times*, 7 May 2002.

[13] "Hot air" refers to emissions permits that Russia, Ukraine and others have available to sell under the Kyoto Protocol, even with no emissions reduction, because economic difficulties alone are expected to reduce their emissions far below their Kyoto targets. Gresham's Law concerns the circulation of debased coins and states succinctly that "Bad money drives out the good."

[14] Sweeney (2001) outlines a proposal for applying a safety valve provision to the US regulation of Corporate Average Fuel Economy (CAFE) of the auto fleet. Credits (or deficits) under CAFE might be made tradable among manufacturers, and a government safety valve could be added to ". . . prevent excessive cost . . . in the event that unforeseen market changes or errors in setting targets".

be applied in the climate area. First, if GHG emissions are to be limited (inappropriately in this instance) by the use of a quantity instrument, some feature should be included to avoid the consequences of the complete inflexibility in quantities. The requisite flexibility can be provided by a safety valve or (provided phase-in measures can be taken to deal with start-up problems) by a banking provision, as has been the case with previous cap-and-trade programs. If appropriate phase-in is not feasible, and there is no banking, then a safety valve definitely should be provided.

Two more fundamental issues are, however, often implicit in safety valve proposals. Should the GHG emission limitation goal be achieved by a price or a quantity instrument? And what should that goal be however it is to be achieved? As we have emphasized, the argument for the use of a price instrument for controlling GHG emissions is very strong but, economic reasoning notwithstanding, the dominant choice seems still to be the quantity instrument, not only in the US but also in Europe and elsewhere. That being said, the appropriate goal now and in the future remains an issue. It may be desirable to adopt a safety valve in conjunction with a quantity limit on GHGs, particularly if there is no other way to tame an over-ambitious target. But application of the safety valve proposal will naturally raise objections concerning how these inconsistent components are to be harmonized. Phasing in a target would seem to be a better approach, perhaps with a safety valve to handle anxieties in a start-up period. Once a cap-and-trade system is in place, similar results can be achieved without the safety valve if provision is made for banking, and perhaps borrowing. Finally, assuming it will prove no easier to coordinate a global safety valve than it has been to decide on a global carbon tax, the phasing out of any safety valves in national programs will be required to create a well-functioning international market in emissions permits.

Acknowledgements

We are grateful to John Reilly, David Reiner, Dick Eckaus, Rob Stavins, Dale Heydelauf, Rich Richels, Peter Zapfel, and ZhongXiang Zhang for comments on an earlier version of this paper. All remaining errors are ours alone. The model underlying MIT results was supported by the US Department of Energy, Office of Biological and Environmental Research [BER] (DE-FG02-94ER61937) the US Environmental Protection Agency (X-827703-01-0), the Electric Power Research Institute, and by a consortium of industry and foundation sponsors.

Appendix A. The economics of instrument choice

A recurring issue in environmental economics is whether a quantity instrument, such as a cap, or a price instrument, such as a tax, is more appropriate on efficiency grounds. Under conditions of certainty, the policymaker can be indifferent: either will achieve the environmental goal at least cost. Under uncertainty, however, they are not equivalent. This appendix explains why.

Consider first the equivalence of price and quantity instruments under certainty. The marginal cost of emissions reduction is conveniently expressed in the form of a Marginal Abatement Cost (MAC) curve, shown by the solid line in Fig. 1. Marginal cost in \$/ton is plotted on the vertical axis and total emissions, E, on the horizontal. The upward-sloping, dashed line expresses the Marginal Social Cost (MSC) or damages caused by these emissions. If the MAC and MSC relationships are known with certainty, the economically efficient level of control (indicated by the intersection of the curves) can be achieved equally well by capping emissions at T or by imposing a tax equal to MC(T). The total cost of emissions abatement is the area under the MAC curve between E' and T.

Unfortunately, the cost and benefit relationships are never known with certainty. At best, policymakers have only some rough notion of their placement, and therefore ex ante they can never be sure of the best control level, even though they know enough to warrant some level of control. Most studies of environmental issues indicate that the general shapes of the curves are as portrayed in Fig. 1. That is, as emissions are reduced the marginal cost of abatement increases, and the marginal social cost (or, equivalently, the marginal benefit of further reduction) diminishes. The now classic answer to the question about instrument choice under uncertainty, attributable to Weitzman (1974), focuses on the rates at which the costs rise and benefits fall as emissions are reduced.

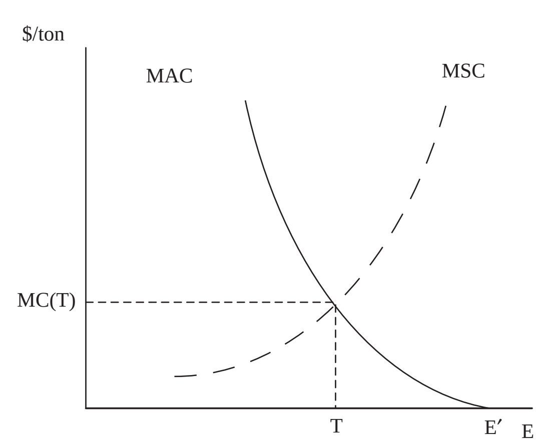

Fig. 1. Quantity target or price under certainty.

Weitzman's point can be illustrated with two variations of Fig. 1. In Fig. 2a the slope of the marginal cost of abatement is relatively flat and the slope of the marginal damage from emissions is relatively steep. As abatement diminishes (or as emissions increase along the horizontal axis), the marginal cost of abatement falls relatively little compared to the marginal damages which rise more rapidly. In the opposite case, shown in Fig. 2b, the slope of the MAC is relatively steep but the slope of the MSC is relatively flat. As an example, assume that the MAC curve is believed to be at some reference level, MAC_R, but it might be higher at MAC_H, or lower at MAC_L. Forced to operate with incomplete information, policymakers might believe that their best choice is represented by the intersection of the MAC_R and MSC curves. But they also know they may be proved wrong: the best level of control may turn out to have been the intersection of the MSC curve with the MAC_H or MAC_L curves. They have no option but to decide on the basis of the limited information, but the choice of instrument to achieve the preferred level of control is no longer a matter of indifference.

Suppose in the world depicted in Fig. 2a they assume reference cost conditions and choose a tax as the instrument of control. The desired level of emissions in this case would be T_R, which regulators would seek to achieve by setting a price of $MC_R(T_R)$, where MC_R denotes marginal cost under Reference conditions. But then suppose the true MAC was later revealed to be

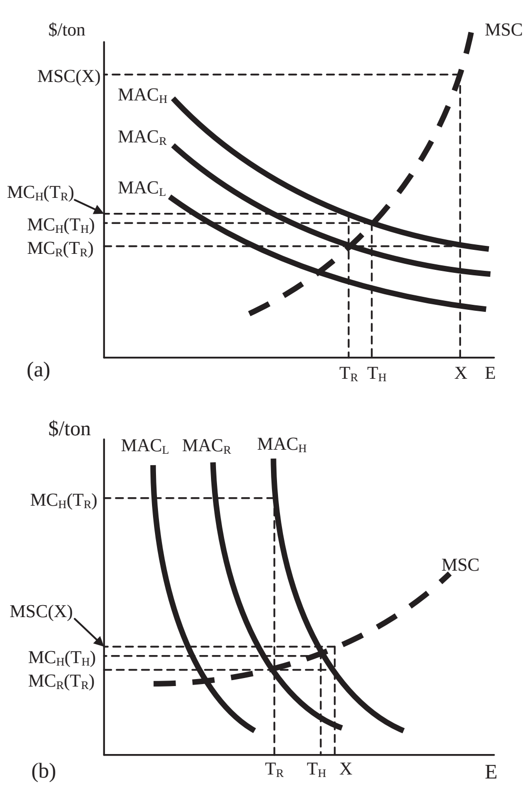

Fig. 2. The choice between quantity and price.

MAC_H. The resulting emissions, shown as point X, would be far above the desired level (T_H) under these conditions, yielding marginal social costs of emissions $MSC(X)$ far above the marginal abatement cost. Regulators would have been better off had they chosen a quantity limit, set at T_R, as the control instrument. Had they done so, emissions would have been much closer to the desired level (T_H) under high cost conditions. While the choice of the "wrong" instrument will always lead to higher marginal cost or higher emissions than would be optimal (or lower if the shift is to MAC_L), the choice of a quantity instrument in this case leads to a much smaller departure from the desired level. This result comes about because the rate of increase in marginal damages as emissions increase is greater in absolute value than the rate at which marginal costs of abatement are falling. Alternatively, if the world is that depicted in Fig. 2b, the choice of the tax at $MC_R(T_R)$ would be the preferred approach because the resulting level of emissions would remain close to the optimal level of control once the uncertainty is resolved. In contrast, application of the quantity target T_R in this latter case would, under high cost conditions, yield a marginal cost $MC_H(T_R)$ far above the marginal social cost at this level of control.

The slopes of these curves are abstractions but they represent real-world alternatives that are relevant to instrument choice in the climate change policy debate. Desired levels of environmental control are often formulated as critical thresholds, a concept that implies that damages increase rapidly as the level of emissions approaches or exceeds some level. This situation is depicted in Fig. 2a. Serious losses of welfare occur if emissions are not kept close to the optimal level. The alternative circumstance, shown in Fig. 2b, is one in which the marginal damages do not change greatly as emissions vary. This latter case characterizes "stock" pollutants, such as the greenhouse gases, whose damages depend on accumulated emissions instead of current releases.

If a quantity instrument is chosen for controlling a stock pollutant like greenhouse gases nonetheless, the adverse consequences of that choice can be reduced by a safety valve, as illustrated in Fig. 3. Assume, for example, that the regulator sets the quantity target at the right level, T_R, believing that MAC_R represents the best existing estimate of costs. Suppose also that the regulator recognizes that the uncertainty surrounding costs may result in the relationship depicted by MAC_H and agrees to sell emission permits without limit at a safety valve price, P_{SV}, at or near the expected marginal cost at the optimal emissions level.[15] If MAC_H is

[15] The logic behind the safety valve would call for the regulatory authority also to offer to buy back any unused permits at price P_{SV}, but this additional provision is typically not included in safety valve

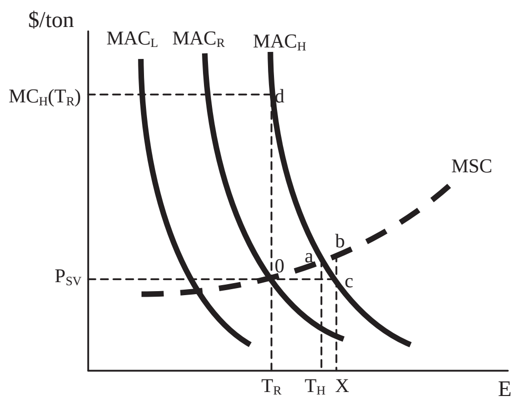

Fig. 3. The safety valve.

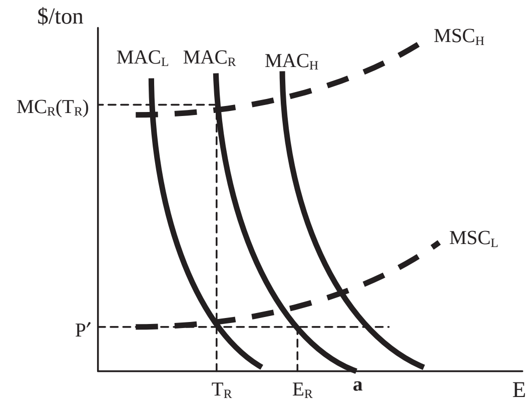

Fig. 4. Safety valve acts as an emissions tax.

realized, then the emitters would abate to level X and buy permits in the amount X–T_R. Emissions will be slightly higher than what is revealed to be the optimal level, T_H, but the loss in welfare is far less that what would have occurred without the safety valve.

The welfare effects of the difference between ex ante expectation and ex post realization are illustrated by the areas beneath the two curves. Because emissions are at level X rather than at the ex post optimal T_H, total costs are lower by the area under the MAC curve between X and T_H, while the social costs of the higher emissions are greater by the area under the MSC curve over this range. With the safety valve, therefore, the welfare loss is the area of the triangle a–b–c. In contrast, if a fixed quantitative target had been imposed at T_R without a safety valve, the total abatement costs would be higher by the area under the MAC between T_H and T_R, while the total social costs of the emissions would be lower by the area under the MSC over the same distance. The welfare loss is the triangle 0–a–d. It is this much greater welfare loss that motivates the argument for price instruments in the climate change policy debate, and for the addition of the safety valve to quantity targets.

Fig. 3 shows what might be called the pure safety valve, where the price is set near to expected marginal cost if emissions are held to the cap. The effect when the safety is set far below this level is illustrated in Fig. 4. Here the quantity limit is set at T$_R$ and the expected cost under reference conditions is MC$_R$(T_R), but the safety valve price for exceeding the quantity limit is set far below, at P'. As before, at a price of P' under reference conditions, emissions will be E_R and the cost to emitters will be the area beneath the MAC curve between the points a and E_R, plus a payment to the regulator of $P'(E_R—T_R)$. It is not obvious why such a low price in relation to expected marginal cost at T_R would be chosen, but it may be that MSC$_L$ is believed to be a more accurate representation of climate damage. If so, then P' is the optimal tax and all that T_R does is to determine the level of exemption from the tax. Whatever the reason, however, if a target such as T_R is agreed and escape mechanisms like P' is included, two fundamental changes have been made: the target reduction has been changed from T_R to E_R and the control instrument has been changed from quantity to price.

(*footnote continued*)
proposals. The concern is (illogically) to avoid the welfare losses associated with significantly non-optimal pricing of the environmental amenity in high cost outcomes, but not in low cost outcomes.

References

Aulisi, A., Dudek, D., Goffman, J., Oppenheimer, M., Petsonk, A., Wade, S., 2000. From obstacle to opportunity: how acid rain emissions trading is delivering cleaner air, at http://www.environmentaldefense.org/pdf.cfm?ContentID = 645&FileName = SO2%2Epdf.

Babiker, M.H., Jacoby, H.D., Reilly, J.M., Reiner, D.M., 2002. The evolution of a climate regime: Kyoto to Marrakech and beyond. Environmental Science and Policy 5, 195–206.

Barnes, P., 2001. Who Owns the Sky: Our Common Assets and the Future of Capitalism. Island Press, Washington, DC.

CBO (Congressional Budget Office, US Congress), 2001. An evaluation of cap-and-trade programs for reducing US carbon emissions, Washington, DC.

CEC (Commission of the European Communities), 2001. Proposal for a Directive of the European Parliament and of the Council establishing a framework for greenhouse gas emissions trading within the European Community and amending Council Directive 96/61/EC, COM 581 dated October 23, 2001, Brussels, Belgium.

Den Elzen, M.G.J., de Moor, A.P.G., 2001. Evaluating the bonn agreement and some key issues. RIVM Report 728001016/2001, Dutch Ministry of Environment, Climate Change Policy Support Project, Bilthoven.

Ellerman, A.D., 2000. Tradable permits for greenhouse gas emissions: a primer with particular reference to Europe. MIT Joint Program on the Science and Policy of Global Change, Report No. 69.

Ellerman, A.D., Joskow, P.L., Schmalensee, R., Montero, J.-P., Bailey, E.M., 2000. Markets for Clean Air: The US Acid Rain Program. Cambridge University Press, Cambridge, UK.

Hourcade, J.-C., Ghersi, F., 2002. The economics of a lost deal: Kyoto–The Hague–Marrakech. The Energy Journal 23 (3), 1–26.

Kopp, R.J., Morgenstern, R.D., Toman, M.A., 1998. The Kyoto protocol: the realities of implementation, resources for the future, Washington, DC (http://www.weathervane.rff.org/features/feature027.html).

Kopp, R.J., Morgenstern, R.D., Pizer, W.A., Toman, M.A., 1999. A proposal for credible early action in US climate policy, resources for the future, Washington, D.C. (http://www.weathervane.rff.org/features/feature060.html).

Kopp, R., Morgenstern, R.D., Pizer, W.A., 2000. Limiting cost, assuring effort, and encouraging ratification: compliance under the Kyoto Protocol. CIRED/RFF Workshop on Compliance and Supplemental Framework (http://www.weathervane.rff.org/features/parisconf0721/summary.html).

Kopp, R.J., Morgenstern, R.D., Pizer, W., Ghersi, F., 2001. Reducing cost uncertainty and encouraging ratification of the Kyoto Protocol. 2000 Taipei Conference for Greenhouse Gases/Reduction, and Pollution Control in Asian–Pacific Countries, Taipei, Taiwan.

Manne, A.S., Richels, R.G., 1999. The Kyoto protocol: a cost-effective strategy for meeting environmental objectives? The Energy Journal, Special Issue, 1–23.

Morgenstern, R.D., 2002. Reducing carbon emissions and limiting costs. Resources for the Future, Washington, DC (http://www.rff.org/climatechangemorgenstern.pdf).

NZ (Department of Prime Minister and Cabinet, Government of New Zealand), 2002. Climate change: the government's preferred policy package. A discussion document, April.

Nordhaus, W.D., Boyer, J., 2000. Warming the World: Economic Models of Global Warming. MIT Press, Cambridge, MA.

Pedersen, Sigurd Lauge, 2000. The Danish CO_2 emissions trading system. RECIEL 9(3). Available also at http://www.ens.dk/graphics/ENS_Forsyning/Kvoter/Danish_CO2_cap_Final.pdf.

Pizer, W.A., 1997. Prices vs. quantities revisited: the case of climate change. Discussion Paper 98-02, Resources for the Future, Washington, DC.

Pizer, W.A., 1999. Choosing price or quantity controls for greenhouse gases. Climate Issues Brief No. 17, Resources for the Future, Washington, DC.

Reuters News Service, 2002. Fines for NSW power retailers greenhouse failures. By Michelle Nichols, May 9.

Roberts, M.J., Spence, M., 1976. Effluent charges and licenses under uncertainty. Journal of Public Economics 5, 193–208.

Stavins, R.N., 1996. Correlated uncertainty and policy instrument choice. Journal of Environmental Economics and Management 30, 218–232.

Sweeney, J.L., 2001. Tradable fuel economy credits, analysis and text extended from Chapter 5 of impact and effectiveness of corporate average fuel economy (CAFÉ) standards, US National Research Council, 2002.

Weitzman, M.L., 1974. Prices vs. quantities. Review of Economic Studies XLI, 477–491.

Weyant, J.P., Hill, J.N., 1999. Introduction and overview, to the costs of the Kyoto Protocol: a multi-model evaluation. The Energy Journal, Special Issue, vii–xliv.

White House, 2002. US. Climate Strategy: A New Approach. Policy Briefing Book, Washington, DC, 14 February.