4 The Importance of Economic Factors in Setting Combined Environmental Standards

4.1 Scope

Scientific enquiry seeks to reveal the nature and quality of combined exposures from pollutants observing how dose-response functions change for individual pollutants as the dose of other agents is varied. Having obtained values for a range of doses, one can then plot isoboles showing combinations of agents with equal impacts (on human health or on other categories of impact target). With deterministic impact phenomena, experimental and epidemiological studies aim to derive health safety thresholds. With stochastic impact phenomena there are no such thresholds, and efforts are restricted to deriving combined dose-response functions.

Scientific findings of this kind are a necessary input to environmental policy, but are not enough when it comes to setting specific environmental targets. Policy choices are also guided to a large degree by legal, economic and social-science factors. Economic enquiry in this context is chiefly concerned with how to design environmental policies incorporating cost-benefit aspects, or with the economic implications of omitting such aspects from a policy programme. In the following, we aim to show how policies differ between single-pollutant and multiple-pollutant phenomena. The conventional economic approach neglects combined effects, generally assuming a given type of adverse effect to be the result of a single factor. This approach is too simplistic for many environmental phenomena.

From an economics viewpoint, the purpose of environmental policy is to ensure that certain exposure levels for human beings, fauna, flora and the climate are not exceeded, and that this is done in due consideration of cost. Environmental economics accentuates the goal of economic efficiency (cost effectiveness). For single pollutants, it shows that in some circumstances fiscal and market instruments (environmental levies and emission certificates) can deliver environmental protection at lower cost than command-and-control regulation and hence are the appropriate choice in such circumstances (Cansier 1996). But for multiple pollutants, the economic efficiency goal takes on even greater importance. In this scenario, at least two substances are responsible for a given adverse effect. The question then is, given a specific exposure limit, which of the substances should be regulated, and to what extent? Should policies target one substance or all of them, and if the latter, in what combination? Economics provides us with a decision rule in the form of the cost minimisation goal. This favours whichever combination of environmental standards has the least total macroeconomic cost while still ensuring that the applicable exposure limit is not exceeded.

Adverse environmental effects can give rise to health damage, material loss and damage (including lost production), and ecological damage. How impacts are rated depends on what impact levels society finds tolerable – that is, on guideline targets. These are the standard against which limit values set by environmental policy are measured. The prime goal of environmental policy is to protect public health, where health is defined in a specific way and can be measured in terms of the relative frequency or severity of a given illness. The (quantified) guideline targets of environmental policy are taken as given when assessing cost-effectiveness.

Since concentrations of pollutants act on human beings through the environmental media (air, water, soil, etc.), it is necessary to apply empirical findings on environmental impacts to real exposures as observed in the field. We assume this can be done satisfactorily and that the dose-response relationships are of the same type as those obtained in the laboratory. In our analysis, we take the experimental doses as being equal to the maximum permissible exposures.[1]

Environmental economics is also concerned with ascertaining socially desirable levels for guideline targets, and recommends that these levels be decided by comparing costs and benefits. A target should be raised when the benefits (adverse effects avoided) exceed the costs. In this way it is possible to obtain the economic optimum for each guideline target (Cansier 1996). It would be interesting to explore how moving from single to multiple substances changes the way in which these optimum targets are determined. We will not pursue this further in the present study, primarily because no reliable way has yet been found of ascribing a monetary value to adverse effects on human beings, fauna, flora and the climate.

Complying with exposure limits entails limiting emissions. The limits for total permissible emissions constitute the operational targets of environmental policy instruments (emission limits, product standards, environmental levies and emission certificates, voluntary commitments by industry). Limiting emissions means taking action to avoid them, and such action has a cost. This cost is indirectly related to the exposure values.

Our analysis initially assumes that all significant factors are known to decision-makers (government or parliament). In reality, environmental policy decisions are uncertain. This is particularly so for combined effects of pollutants, of which relatively little is yet known. It would thus appear appropriate to conduct an additional analysis for the case of imperfect information.

---

[1] The set of all combinations of doses of different substances which have the same effect on an environmental quality target is termed an "isoquant".

4.2
Cost-effective Environmental Protection with Adequate Information

4.2.1
Adverse Effects

Isoquant curves: Let us take a specific guideline target of environmental policy as given and assume that adverse effects (expressed as the frequency or severity of a disease) can be induced by two substances, A and B. The guideline target must be one that can be stated in quantitative form.[2] We express the impact at alternative doses of A and B with an impact function S(A,B). For our analysis, we need to know what combinations of the two agents remain within the target function. These combinations are plotted by a target isoquant curve. All combinations on the curve meet the target criterion and by this definition are ecologically equivalent. The shape of the curve can vary considerably (see fig. 41):

- With a linear function, the pollutants are mutually substitutable at a constant ratio:

  $$(1)\quad -\frac{dA}{dB} = \frac{\partial S / \partial B}{\partial S / \partial A} = \text{constant}$$

  where $\partial S / \partial SA$ and $\partial S / \partial SB$ are the marginal impacts of the substances.
- With a convex curve, the marginal rate of substitution $|dA / dB|$ rises as the quantity of A increases and the quantity of B decreases. Diminishing quantities of A are needed to cancel the adverse effect of each additional unit of B, and diminishing quantities of B are needed to cancel the adverse effect of each additional unit of A.
- With a concave curve, the marginal rate of substitution $|dA / dB|$ diminishes as the quantity of A increases and the quantity of B decreases. Cancelling the impact of each added A requires increasing reductions in B as more As are added, and vice versa.
- An S-shaped or "sigmoid" curve is made up of two adjacent ranges, one concave and one convex. The marginal rate of substitution rises over the first half of the curve and falls again over the second, or vice versa.

Additivity and independence models: The three scientific reference models for evaluating combined effects on organisms are response additivity, dose additivity and independence (see Section 2.2).[3] Effects that go beyond or fall short of these reference standards are referred to as super-additive or sub-additive effects. There are three possibilities:

---

2 For ease of presentation, we use the same symbol for the type of substance and the quantity.

3 The models do not take into account reactions between agents in the environmental medium (such as the formation of ground-level ozone).

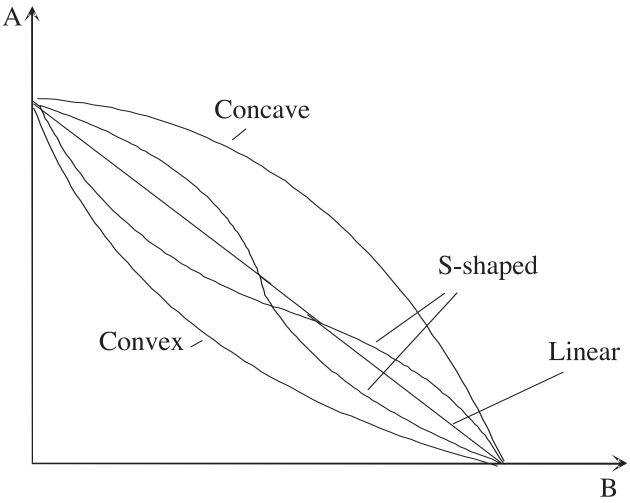

Fig. 4-1 Alternative shapes of the isoquant curve. The isoquant curve for agents A and B is the locus of all pairs of quantities of A and B that are assumed to have the same adverse effect. At a constant level of damage, a given reduction in exposure to A requires a given increase in exposure to B, and vice versa. This rate of substitution is determined by the marginal effects of the agents.

- *Response additivity:* The total effect E of a combination of substances equals the sum of the individual effects E_A and E_B (with the constraint that the sum cannot exceed the maximum adverse effect): $E = E_A(A) + E_B(B)$.
- *Independence:* $E = E_A(A) + E_B(B) - E_A(A) \cdot E_B(B)$
  The total effect is less than that obtained with response additivity.
- *Dose additivity:* Effects are expressed as dose equivalents of a reference substance, for example A. If the dose-response function is linear, the total effect E is defined as $E = E_A(A + u \cdot B)$, where u is the constant coefficient of equivalence. The effect of a given quantity of $\overline{B}$ is measured with reference to the base level of the given quantity of $\overline{A}$ using the dose-response function for $\overline{A}$. If the function is non-linear, the quantity of $\overline{B}$ at a given dose of $\overline{A}$ is obtained by using this formula (in which X is the upper limit to be found for the quantity of A):

$$(2) \quad \int^{\overline{B}} E'_B(B)dB = \int_{\underline{}}^{X} E'_A(A)dA \, .$$

  The total effect of $\overline{A}$ and $\overline{B}$ is then $E = E_A(X)$.

All three approaches allow us to plot an isoquant curve. If the dose-response functions of the component substances are linear, the dose additivity and response additivity models produce identical linear isoquants. In the low-dose range, we obtain approximately the same results as with independence models (see Section 2.2).

- If *both* agents have a linear, super-linear or sub-linear impact function over the relevant dose range (see the sigmoid function in Section 2.2), the isoquant curve will be linear, concave or convex.
- If one agent has a linear and the other a super-linear (sub-linear) impact function over the relevant dose range, the isoquant curve will be concave (convex). If one function is sub-linear and the other super-linear, the isoquant curve can be concave, convex or linear depending on which component effect predominates.

It is a well established hypothesis that dose-response relationships are linear or slightly super-linear for individual substances in the health-relevant low-exposure range (for chemicals and ionizing radiation). This simplifies our analysis, as it is the low-exposure range that interests us. The isoquant curves for additivity will thus be linear or slightly concave.

Combined exposures: Interactions between component substances either weaken (sub-additivity) or amplify (super-additivity) the overall effect compared with the reference models (additivity and independence). The resulting variation in the shape of the isoquant curves concerns us only over the low-dose range. We obtain two cases:

- Linear reference function: If there is unidirectional or reciprocal synergy between the effects of the component substances (super-additivity), the isoquant curve will be convex. If effects are unidirectionally or reciprocally antagonistic (sub-additive), the curve will be concave.
- Concave reference function: Antagonistic effects increase the concavity of the curve. Synergy between the effects reduces the concavity, eventually producing a straight line or a convex curve.

4.2.2 Abatement Costs

Emission reduction has a cost. This cost generally rises in linear or greater-than-linear proportion with the size of the reduction. A greater-than-linear cost increase can generally be expected when the reduction is large. That is, the cost function is likely to be super-linear with high guideline targets and linear with low guideline targets. We intend to establish the relationship between costs and exposures, under the simplifying assumption that exposures are directly proportionate to emissions.

At constant macroeconomic cost K* it is possible to achieve different combinations of reductions in A and B. The set of possible pairs of reductions V_A and V_B is stated by an isocost function. The cost of a further unit reduction in A must be matched by cost savings from decreasing the reduction in B, and vice versa. The condition is $K^* = K_A(V_A) + K_B(V)_B$. From this we derive the marginal rate of substitution between the respective abatement costs of A and B:

$$\text{(3a)} \quad -\frac{dV_A}{dV_B} = \frac{\partial K_B}{dV_B} \bigg/ \frac{\partial K_A}{\partial V_A} .$$

Because the respective quantitative reductions are identical with the difference between the initial emissions $\bar{A}$ and $\bar{B}$ and the residual emissions A and B, the rate of substitution can alternatively be expressed as follows:

$$(3b) \qquad -\frac{dA}{dB} = \frac{\partial K_B}{\partial B} \Big/ \frac{\partial K_A}{\partial A}.$$

If the cost increase with increasing reductions is linear (or super-linear), the isocost function is linear (or convex looking from the origin; see K_1 in fig. 4-2).

4.2.3
Cost-efficient Limit Values

Assume a given environmental quality standard is to be complied with (for example prevention of adverse health effects in a case of non-stochastic impact phenomena). This objective corresponds to a specific target isoquant curve. This curve plots the set of possible target-compliant combinations of A and B. The isocost function, on the other hand, plots the combinations that are attainable in prevention terms. The most cost-effective combination of environmental quality standards for A and B is located at the point where the target isoquant curve is touched by an isocost curve with the least possible cost (scenario M in fig. 4-2).

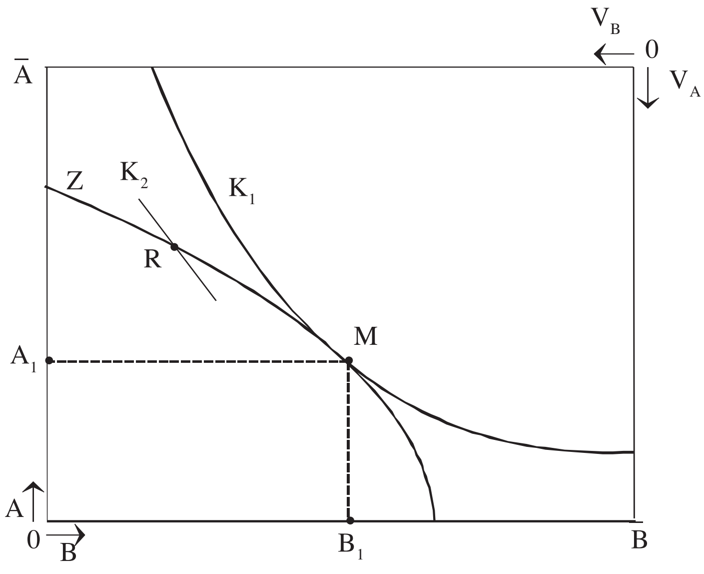

Fig. 4-2 Efficient limit values for agents A and B. In the beginning, exposures to the two agents total $\bar{A}$ and $\bar{B}$. If Z is the target isoquant curve and the cost conditions for preventing A and B are represented by isocost function K_1, the most efficient option is to reduce $\bar{A}$ to exposure level A1 and $\bar{B}$ to exposure level B_1. Selecting a combination such as that at point R would entail a higher cost of environmental protection. B would be reduced too far and A too little relative to the efficient values.

The cost of environmental protection depends on two fundamental conditions: the nature of the relationship between the combined effects and the nature of the relationship between the agents' respective abatement cost and impact per unit.[4]

Significance of relative impacts and abatement cost: Let us begin by taking a given type of combined effect (additive or non-additive). The properties of the efficient solutions vary according to the shape of the effect and the cost curves. The most cost-effective solution may be to regulate both agents (see M in fig. 4-2). The defining condition for an efficient combination is that the agents' relative marginal impacts and relative marginal abatement costs are equal:

$$(4)\quad \frac{\partial S / \partial B}{\partial S / \partial A} = \frac{\partial K_B / \partial B}{\partial K_A / \partial A}.$$

Because the cost rises at a greater-than-linear rate as reductions become larger, achieving cost-effectiveness will generally entail restrictions on both substances (or all substances involved). Policy should not be limited to a single substance even if this is the main cause of the adverse effect in question. With any given pair of pollutants, the reduction in emissions should be greater for the one that has the greater relative adverse effect and the lesser relative abatement cost. All target-compliant combinations other than M in fig. 4-2 would incur a higher macroeconomic cost. If the limit values were to differ in reality – as with scenario R – cost-saving opportunities would go unexploited. The same degree of environmental protection could be reached at lesser cost. In scenario R, the relative impact of B is less (say, 2) than the relative cost of B (say, 4). If we increase B by one unit, we would have to reduce A by two units (assuming A and B are ecologically equivalent). But if we reduce B by one less unit, there is a cost saving that can be reinvested in preventing an additional four units of A, whereas in fact we only need to prevent an additional two units of A. That is, at the specified total cost it is possible to attain a higher target, or the specified target is attainable at lesser macroeconomic cost.

In certain instances the most cost-efficient option may be to regulate a single pollutant. This arises when, at all quantities, the relative marginal abatement cost of the pollutant is less than its relative marginal impact (see also von Ungern-Sternberg 1987). The other factor should only be included in the abatement strategy if the environmental quality target cannot be attained by preventing the main factor alone. The likelihood of this being the case increases as targets become more stringent; that is, with larger reductions in exposures.

In the following we will use the example of carcinogenic effects of radioactivity to illustrate the use of cost-effectiveness principle in determining efficient combined environmental standards. The carcinogenic effects of ionizing radiation can be inhibited or potentiated by other substances (see fig. 4-3 for the possible shapes of isoquant curves). The other substances are not carcinogenic in their own right. Generally, either the effects of ionizing radiation combined with chemotoxic substances act additively or the combined effect is smaller than the sum of the individ-

[4] Where a single agent is more or less the sole cause of an adverse effect and other agents have only a minor contribution to the effect as a whole, regulatory considerations are restricted to the first agent from the outset.

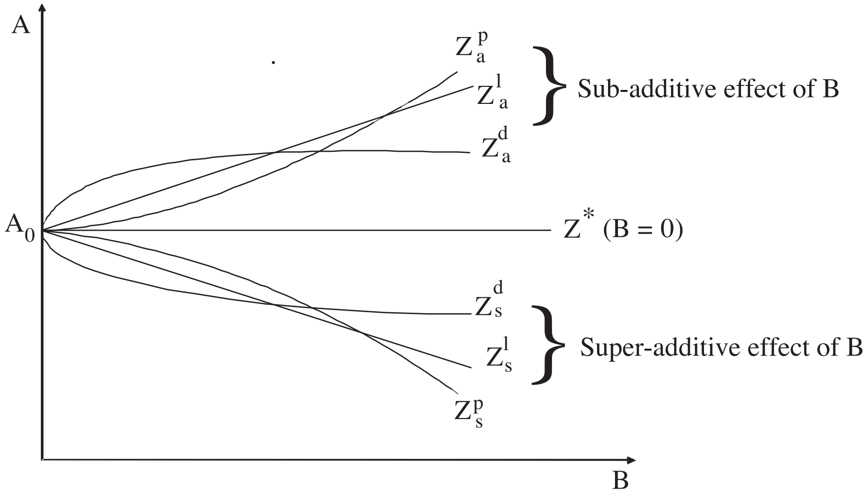

Fig. 4-3 Example of isoquant curves where the adverse effect caused by A is inhibited (potentiated) by agent B. If exposure to B is zero, the adverse effect is assumed to be the limit value A0. With sub-additivity, A must be reduced as the quantity of B increases for the adverse effect to remain constant. The inhibiting effect of B can be linear, sub-linear or super-linear as B increases (see Z_a^I, Z_a^d, Z_a^p). Super-additivity acts analogously.

ual effects. Super-additive effects have been found less frequently (for example with hormones, viruses, and smoking), in most cases at medium to high doses (see section 2.3).

We will look at the case where a second agent inhibits the effect of radiation. What is important is the relationship between the target function and the starting conditions. The target function may only be attainable by reducing A or B. This is assumed to be the case in fig. 4-4. With target function Z_1, the objective is to reduce A (radiation) to A_1; and with target function Z_2, the objective is to reduce B (a chemical) to B_1. Only boundary solutions are efficient. If function Z_1 is typical of cancer risk from ionizing radiation, only the latter ought to be reduced. Taking into consideration the sub-additive effect of the chemical, however, it is enough to reduce radiation to A_1 rather than C (the reduction that would be necessary in the absence of an inhibitor). The chemical reduces the impact of the radiation at zero cost, and this ought to be exploited to the full. If policy were to ignore the interaction and reduce radiation, say, to C, the target would be exceeded, incurring unnecessary cost. We can improve on this policy by relaxing the limits on radiation exposure. (A differentiated policy is needed if the two factors occur together in some places and separately in others.)

A policy may take the interaction into account and choose a combination of environmental standards on the target isoquant curve, for example D, but ignore the cost aspect. At D, the cost is excessive. The inhibiting factor is reduced unnecessarily, which means A is overregulated. It is advantageous to allow the inhibiting factor to take full effect and to reduce the causal factor A to a lesser extent.

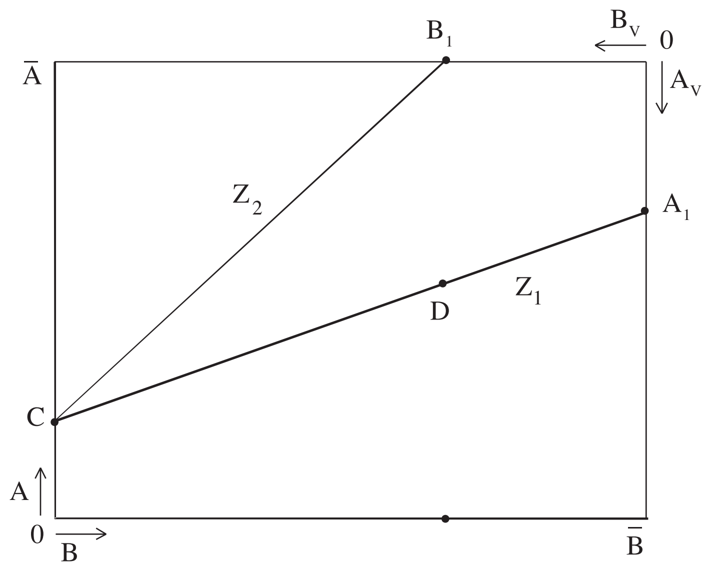

Fig. 4-4 Efficient limit values where the cancer risk triggered by A is inhibited by another chemical B. The sub-additive effect of B reduces the damage caused by A. It would be inefficient not to exploit this mechanism. If effects are sub-additive throughout, exposures to B should be permitted to the extent that they are compatible with the ecological objective. The efficient exposures are determined for target isoquant Z_1 in accordance with A_1 and for target isoquant Z_2 in accordance with B_1. The only cost incurred is from the prevention of A in the former case and from the prevention of B in the latter.

Cost implications of the different types of combined effect: We must distinguish between sub-additive and super-additive effects. The following can be stated:

- The cost of environmental protection is greater with super-additive effects than with additive or independent effects. The potentiation of individual effects is concomitant with an increase in cost. In order to attain the same environmental policy guideline target as would hypothetically be attained with additive or independent effects, policies must stipulate higher limit values for the agents involved. This incurs a higher cost and so tends to make it harder to enforce a given target in the policymaking process. In policy terms it may be appropriate to accept a lower level of environmental protection.
- Where two substances inhibit each other, they effectively deliver abatement at zero cost. If the effects are additive, exposure levels need to be set lower to reap the same benefit as would be obtained if they were sub-additive. Compliance with an environmental target entails less environmental protection and hence lower environmental cost. At the same time, this can give occasion to pursue a more exacting target.

Because super-additive effects tend to produce a convex target isoquant curve, sub-additive effects tend to produce a concave curve and the cost of abatement typically increases at a super-linear rate, there is also another difference: With super-additive effects, the target isoquant curve and the isocost functions can have a similar gradient over a fairly broad range of combinations of exposure values. This facilitates policymaking decisions.

Conclusion: The unnecessarily high cost of environmental protection is a significant barrier to the enforcement of stringent environmental standards. The costs can be reduced with multicausal impact phenomena by combining limit values on the basis of cost-efficiency. Costs are not only a function of the means of abatement; it also depends on the adverse effects of the agents concerned. The two facets should not be considered separately.

The cost-efficiency approach requires policy to have a quantitative focus. Cost comparisons are only possible with quantitative environmental targets (observing acceptable levels of health risk, setting upper bounds for global warming, preventing summer smog, etc.) and exposure values. This supports business and industry demands for policies to make greater use of quantitative as opposed to qualitative targets (such as ones to avoid adverse environmental effects or requiring state-of-the-art preventive environmental management). Operational health risk levels are used, for example, in the Netherlands and the USA (Tegner and Grewing 1996). According to Rehbinder, the accepted view in the debate surrounding hazardous substances in the US is that quantitative risk assessment and ensuing risk/cost analysis is an appropriate means of determining residual risk. The EPA regards risks as tolerable in the range 1:100,000 to 1:1,000,000 (Rehbinder 1991). Under the Clean Air Act, limits for carcinogens must be based on a 1:1,000,000 lifetime probability of contracting cancer (Böhm 1996). In German environmental policy, quantitative targets are still quite rare. They exist in the field of hazard control and for specific phenomena under preventive environmental policies, for example CO_2, SO_2 and NO_X emissions. There are proposals in other areas. The German Council of Environmental Advisors (SRU), for example, has recommended preventing summer smog and reducing the associated cancer risk from currently 1:5,000 in rural districts and 1:1,000 in urban districts to 1:10,000 and 1:1,000,000 respectively (Der Rat von Sachverständigen für Umweltfragen 1994).

4.3 An Empirical Example

The smog problem in urban New York has been investigated using the cost minimization approach (Repetto 1987). Ground-level ozone, which indicates the presence of photochemical oxidants, forms by the interaction of nitrogen oxides (NO_X) and volatile organic compounds (VOCs). The main emitters of VOCs (in Germany) are transport (about 50%) and users of solvents (about 41%). Some 73% of NO_X emissions are produced by transport, 13% by power stations and district heating facilities, and 10% by industry (Umweltbundesamt 1998).

According to Repetto, the isoquants plotting pairs of quantities of NO_X and VOCs that yield equal maximum hourly ozone concentrations have a typical shape that does not significantly change with meteorological conditions. The marginal effect of both precursor pollutants on ozone formation falls with rising concentration. The absolute effect of NO_X on ozone formation diminishes above a certain range. These findings result in convex iso-ozone curves. Repetto observes (for years preceding 1987) that in those US air quality regions where the national ozone standard had been exceeded, there had been efforts to reduce VOCs but not to reduce NO_X. He criticises this policy for failing to consider the cost aspects.

Based on estimates for urban New York, the cost functions for preventing NO_X and VOC emissions have two distinguishing features: an initial range where abatement is possible at negative cost and a sharp rise in cost once straightforward abatement options have been exhausted and more sophisticated and capital-intensive approaches become necessary. Consequently, the isocost functions have a pronounced convex shape (as seen from the origin). The results of the study are as follows:

- If one were to invest, say, \$75 million a year in preventing smog, the most cost-effective option would be to reduce NO_X emissions by 33% and VOC emissions by 40%.
- Because the isoquant and isocost curves have a similar shape, the added cost of deviating from the efficient dose combination is not particularly large over a fairly wide range of combinations.
- Because the individual pollutants have significant side-effects – in particular NO_X in the formation of acid rain – these side-effects ought also to be considered when assessing specific strategies.

A more recent study stresses that iso-ozone curves vary with the composition of the VOC component and hence differ between regions (Hall 1998). This would imply different regional standards for NO_X and VOCs under cost minimisation. A further study develops a decision model by which the optimum mix of NO_X and VOC reductions is arrived at in accordance with the goal of minimizing the overall macroeconomic cost of environmental protection (abatement cost and damage cost of smog in urban areas) and taking into account effect-related uncertainties and learning effects (Chao et al. 1994). Using an example with given starting conditions for ozone concentration and the respective emissions and given estimates for the abatement cost functions, Chao et al. calculate the optimum quantitative reductions in NO_X and VOCs for various alternative hypothetical impact functions.

4.4 Choice of Policy Instruments

Policy instruments are used selectively to attain specific environmental policy goals. Because smoking and alcohol consumption are major potentiators of health damage, policies for these substances might focus on increasing education, restricting advertising and raising taxes on alcohol and tobacco. However, these impact phenomena are not usually placed within the remit of environmental policy, and we will not give them further consideration here. In practice, policies for regulating air, water and soil-polluting emissions primarily focus on prescriptions and prohibitions. Environmental levies rarely come into play (an example is the waste water levy in Germany). Certificate schemes are even rarer and to date have only been used in the US (for SO_2 and NO_X).

Multiple pollutants pose a special challenge in selecting policy instruments to implement an efficient abatement strategy, and different instruments are not all equally suited to the purpose. Certificates have advantages when combined effects show linear additivity. Regulators using emission levies or emission limits cannot avoid setting their targets equal to efficient limit values from the outset. The emission levies on two pollutants A and B, for example, must necessarily equal the pollutants' marginal abatement costs at the identified efficient limit values. Cost-minimizing polluters will then adjust their emissions in such a way that the targets are attained.

As regards regulating different emitters of a given pollutant, cost-minimization by individual polluters tends to result in an efficient abatement strategy across the group as a whole. Polluters with relatively low abatement costs tend to reduce their emissions more than those with relatively high abatement costs. This is macroeconomically efficient. In ignorance of the abatement cost faced by each polluter, regulators thus tend to set uniform emission limits for all polluting facilities. Prescriptions and prohibitions are consequently less cost-efficient than emission levies. This observation is unrelated to the multiple pollutant problem, however. On the other hand, prescriptions and prohibitions have the advantage that they give more precise control over emissions. This is particularly important when it comes to preventing excessive pollution within the catchment area of large or locally concentrated emission sources (hot spots).

The policymaking task is easier with emission certificates and linearly additive combined effects. Regulators do not need to know the abatement cost functions. They need not give any thought to efficient limit values. They need to know two things: a combination of emissions that accords with the protection or prevention objective at hand, and the (constant) impact per unit of each pollutant. The regulators then issue certificates for the permitted total quantity of emissions and stipulate that they can be traded or used as permits to emit the other pollutant, in either case in the ratio of the pollutants' impacts per unit (Bonus 1975; Endres 1985).

The most straightforward approach is to express the pollutants in unit equivalents of a reference substance and only to issue certificates for emissions of that substance (see fig. 4-2). The environment agency issues certificates (in 1:1 units) in quantity A*, which accords with the environmental target (Z_A). The impact functions for agents A and B acting individually are $S_A = a \cdot A$ and $S_B = b \cdot B$. The relationship between permissible combinations of the agents on the target isoquant curve can thus be expressed as $dA = b/a \cdot dB$. Usage of the certificates for emissions of B (Z_B) is stipulated in accordance with this relationship. To emit one unit of B, a polluter must purchase b/a certificates. The certificates are issued by auction. The environment agency sells the permits at the price at which demand equals the supply quantity A*. Demand for the certificates derives from a cost-minimization rule. Each emitter of A, A_i where $i = 1,\ldots,N$ and of B, B_j where $j = 1,\ldots,M$ wishes to minimize its individual cost. This comprises its abatement cost $V_{A,B}$ and its expenditure for the purchase of certificates. The cost of purchasing certificates incurred by each emitter of A equals the certificate price p multiplied by the emitter's residual emissions of A. Emitters of B calculate the cost of their residual emissions in A-equivalents. This cost equals p · b / a · B. We can derive cost functions for each emitter as follows:

$$(5)\ K(V_A^i, Z_A^i) = K(\overline{A}_i - A_i) + p \cdot A_i$$

$$(6)\ K(V_B^j, Z_B^j) = K(\overline{B}_j - B_j) + p \cdot b/a \cdot B_j\,.$$

(where $\overline{A}_i$ and $\overline{B}_j$ are the quantities of emissions produced by each emitter at the outset.)

Taking the certificate prices as given, cost minimization produces:

$$(7)\ p = \frac{\partial K_i}{\partial A_i} \text{ and } p = a/b \cdot \frac{\partial K_j}{\partial B_j}.$$

The emitters attune their emissions (or abatement efforts) to the market price of the certificates until the marginal abatement cost equals the price. The same applies for alternative prices and hence also for the equilibrium price at which there is demand for all certificates issued by the environment agency and hence the condition $A^* = \sum A_i + b/a \sum B$ is met. The resulting structure of demand for the certificates matches the efficient combination of emissions of A and B, which, as we have seen, is governed by:

$$(8)\ \frac{\partial K_j}{\partial B_j} / \frac{\partial K_i}{\partial A_i} = b/a.$$

The market thus ensures that emission certificates issued for A are transformed into the efficient quantity structure for A and B, thus exploiting the advantages of the market mechanism. Policymakers need less information than they do for levies or command-and-control. This is an important advantage of certificates. However, it requires an efficient market. This market efficiency may be assumed to obtain in the domain of global climate protection, for example. Here, the key advantage of (internationally tradable) certificates over a levies scheme would be not needing to know what degree of reduction in CO_2, CH_4 and other greenhouse gases is efficient. This decision can be left to the market. The greenhouse effect per unit of each gas is known. A requirement is that climate policy targets not just CO_2 but all major greenhouse gases, this being necessary for economic reasons (Cansier 1991, Michaelis 1997).[5]

Certificate schemes lose this advantage in the case of non-linearly additive or non-additive combined effects, because in such cases the impact of a unit of each pollutant varies with the quantity of emissions at a given time. It would be asking too much of policymakers to have them adjust the ecologically dictated rate of exchange at which the various certificates are traded to account for changes in unit impacts. Accordingly, we may assume that regulators would have to set efficient

5 On the current state of international climate policy and the chances for and difficulties facing the introduction of tradable permits for greenhouse gases, see Bayer and Cansier, 1999.

limit values for certificates just as with levies and direct regulation. The phenomenon of combined exposures does not suggest a specific set of policy instruments in such cases. How specific protection and abatement objectives are attained then depends, as with single-pollutant phenomena, on the problem at issue. Levies and certificates are particularly well suited for regulating large-area pollution. When it comes to preventing health damage from ionizing radiation, chemicals and the like, on the other hand, command-and-control regulation is the only solution, as protection must be ensured at every exposure site.

4.5 Cost-efficient Limit Values under Imperfect Information

4.5.1 Uncertainty in Decision-making

Relatively little is known about the combined effects of pollutants. Combined effects increase decision-making uncertainty beyond the level that already obtains when determining dose-response curves for individual pollutants and applying laboratory results and epidemiological study findings to real exposures (Der Rat von Sachverständigen für Umweltfragen 1987, Dieter 1995, Hagenah 1996). Yet despite imperfect information, policy decisions on maximum permissible exposures still have to be made, and there is a desire for methodological support from the scientific community. The various models of economic decision theory under uncertain expectations describe the factors that need to be taken into account in complicated circumstances of this nature.

Particular difficulties and uncertainties arise when assessing the nature and intensity of combined effects for two main reasons:

- Because adverse effects can be caused by two or more agents, the effects of many more pairs of quantities must be investigated than with single pollutant phenomena, and this is impracticable for cost reasons. Investigating the effects of only ten different quantities for each of two agents would mean examining a hundred different pairs of quantities. Three agents would mean a thousand experiments. Each of these experiments would require a representative number of fresh laboratory animals. In view of these difficulties, figures for complex mixtures are extrapolated and inferred from results obtained with binary mixtures (see Section 2.4). Hence rather than requiring unequivocal scientific findings, policy must be based on modelling.
- Where the biochemical response mechanisms triggered by individual pollutants are not known, it is not possible to judge the pollutants' overall effects in combination (see Section 2.3). When the various agents operate through different mechanisms, it makes sense to add their *effects*. When the substances operate through identical or similar mechanisms, one should add their *doses*. In such cases, the agents work like dilutions of one and the same substance. Uncertainty about the mechanisms involved would not be a problem here if the dose-

response relationships were linear in the low dose range; however, this is not bound to be the case since linearity is a simplifying assumption made, and reflects the safety margins applied by scientists when extrapolating high doses to lower (realistic) concentrations. Dropping the applied safety margins leaves us with only one "envelope of additivity" (see Section 2.3), and we are forced either to choose one of the alternatives or to take an intermediate solution.

Because response additivity and independence are used as the reference for evaluating interactive combined effects and the latter are not precisely known, the classification of multiple pollutant phenomena as super-linear or sub-linear is likewise uncertain.

These difficulties and uncertainties mean we do not know the precise shape of the target isoquant curves, and are heavily reliant on assumptions. The uncertainty problem is mitigated somewhat in that knowledge of all possible cost and impact-equivalent combinations is not needed in order to make specific recommendations on policy. To satisfy the cost-efficiency principle, it is enough for government decision-makers to obtain (from experts) information on a number of alternative combinations and to make the choice between them on cost criteria.

Scientists endeavour to account for uncertainty by assuming linear and additive dose-response relationships and building a margin for error into the laboratory results:

- For dose-response relationships involving individual substances, figures for low doses are linearly extrapolated from those for high doses. Because the true relationship is probably super-linear, this tends to overestimate rather than underestimate the frequency of adverse effects (see Section 2.2).
- Health safety thresholds for individual pollutants with non-stochastic impact phenomena are arrived at by applying a risk factor to safety levels obtained for animals in the laboratory. The risk factor can exceed five orders of magnitude (see Section 2.4). Only part of the risk factor accounts for uncertainties; the larger remainder reflects the scientifically established greater sensitivity of humans compared with animals and interindividual variations in sensitivity within human populations. The pure safety component is not wholly derived from scientific analysis, but is incorporated by scientists based on their assessment of the risks. Risk factors of this kind are scientifically founded "in principle but not in size". The extreme size of some risk factors is scientifically and politically questionable.
- For combined exposures (where the effects are expected to be the same at the target location in the organism), it is assumed that effects of the substances are additive, corresponding with the effect-additivity, dose-additivity or independence model according to the surmised biochemical process. Additivity ignores possible interactions. If in reality the effects of the agents are mutually antagonistic, the risk is overestimated and relatively high costs of environmental protection are countenanced. If in reality the effects are synergistic, the risk is underestimated.
- With combined pollutants, additional protection is incorporated by using several safety margins simultaneously. For the most part, safety margins do not explic-

itly account for interaction risks, but they may offer some protection in this regard by being set at a generous level. If the effects are sub-additive in reality, safety margins result in (further) overestimation of the actual risk. If the effects are super-additive, it is appropriate to use safety margins in principle, but their size is not made explicit; instead, they are built into the sum total correction factor. A lack of reliable evidence of health impacts through combined exposure given that limits for the various individual substances are observed (medicinal drugs excepted; see Section 2.4) merely implies that the safety margins are not too small. It does not rule out their being larger than needed, in which case a lesser degree of precaution would be acceptable and the cost of environmental protection could be lowered.

These added safety factors for uncertainty are not solely based on scientific analysis. They are assumptions that essentially reflect risk assessments and evaluations. It remains an open question why linearity (in the case of effects of individual pollutants), additivity/independence (in the case of combined effects) and high generic risk factors applied to laboratory findings are considered meaningful conditions for safety – or, alternatively, why other assumptions are not considered sufficient. Besides, it is politically not very realistic to require limits that prevent human exposure beyond all doubt (with 100 per cent probability). People take numerous risks (including health risks) in their everyday lives, and accept a certain probability of suffering harm. Consequently, environmental protection cannot be made absolute. This applies not only for risk prevention in the legal sense of the word, but above all for preventive environmental policies.

The risk factors used in deciding environmental standards under uncertainty should be divulged. Natural scientists can make a fundamental contribution towards determining objective risk factors, but subjective evaluation is outside their domain. The situation is different when it comes to economics, where decisions under uncertainty are a subject of research. Economics shows that in certain circumstances rational decisions are possible with imperfect information. Criteria are derived which ought to be observed when making decisions about risk and which are consequently a useful decision aid when setting limit values. Three situations are distinguished according to the degree of uncertainty:

- The decision-maker has reliable knowledge about possible impact severities and can (subjectively) assign probabilities to them (stochastic model).
- The decision-maker can only state probabilities for the frequency or severity of impacts (fuzzy models).
- The decision-maker only has information about the possible severity of impacts (models subject to uncertainty).

4.5.2 Stochastic Decision Model

In line with the scientific and legal approaches, we assume a policy aim of attaining observance of a specific environmental guideline target with a given probability. The policy stipulates probability $p^* \geq 0$ that impact $S^* \geq 0$ is not exceeded. S^* is the guideline target to be observed. In our example (fig. 4-5), the target isoquant curve

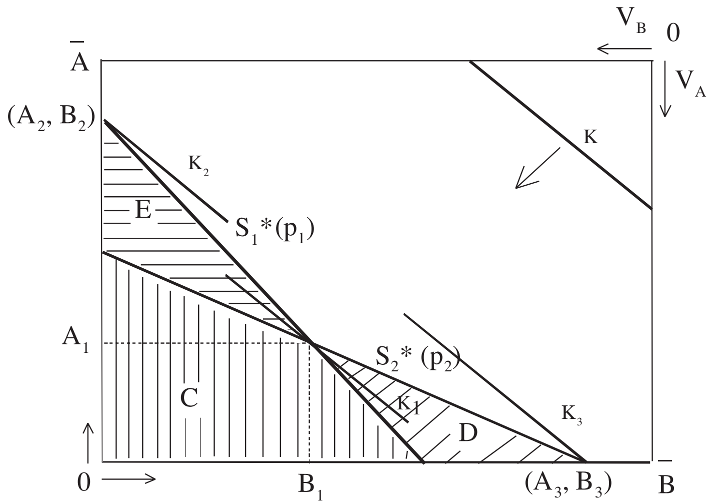

Fig. 4-5 Observance of a protection goal with a specific probability. In the decision-maker's subjective assessment, all combinations of A and B within area C carry a hundred per cent probability that target S* will be observed. If the target is to be observed with certainty ($p_1 + p_2 = 1$), A_1 and B_1 represent the efficient combination of limits for A and B given iso-cost function K_1. All combinations of A and B in area E (D) carry probability p_1 (p_2) of target S* being observed. For example, if we stipulate a probability of p1, the efficient limits are A_2 and B_2 (if $p_1 > p_2$) or A_3 and B_3 (if $p_1 < p_2$).

is expected to be S^*_1 with probability p_1 and S^*_2 with probability p_2. Area C represents all combinations of A and B for which it is certain that S* will be observed ($p_1 + p_2 = 1$). For combinations in area D, the probability that S* will be observed is p_2, for combinations in area E it is p_1.

The environmental policy objective might be to have one hundred per cent (subjective) probability that all impacts will be prevented ($S^* = 0$ and $p^* = 1$). This requirement matches the scientific concept of limit values with deterministic impact phenomena. It is weakened if we merely stipulate a certain probability $p^* < 1$ of all impacts being prevented. This requirement matches the definition of environmental policy targets in German statute and case law. Exposure limits are specified such that there is a "reasonable probability" that health damage will occur if they are exceeded (Cansier 1994). The policy objective is further weakened if we relax the zero impact constraint and permit a certain level of adverse effects $S^* > 0$. This is unavoidable with stochastic impact phenomena unless emissions are to be prohibited entirely – an extreme solution that would be justified in macroeconomic terms only in exceptional cases – and matches the situation in environmental law with preventive environmental policies. As such policies require reductions in emissions through environmental protection to be held in a reasonable relationship to cost (under the principle of proportionality), society must accept a certain amount of residual emissions and hence of adverse affects.

If we require certainty ($p^* = 1$) that $S^* \geq 0$ will be observed, we look for the least-cost combination in area C. This is found at A_1/B_1. Not only is it unnecessary to lower exposures below A_1 and B_1; doing so would incur excessive macroeconomic costs. Note how the safety factors are precisely determined and not explained in terms of exogenously assumed risk margins.

If we vary the environmental policy target and permit a certain impact probability, the solution depends on the ratio of probabilities of occurrence: Let $p_1 < p_2$: To obtain a minimum probability $p^* < 1$ that S^* will be observed, the optimum combination is (A_3, B_3). In area D, the probability of occurrence is greater than p_1. The same result is obtained if the required minimum probability is $p_2^* < 1$. Let $p_1 > p_2$: To obtain a minimum probability $p_1^* < 1$, we look for the least-cost combination in areas C and E. At the assumed slope of the cost function, the optimum combination is (A_2, B_2). If a lower minimum probability $p_2^* < 1$ is required, we look for the least-cost combination in areas C, D and E, and the optimum combination is (A_3, B_3).

4.5.3 Fuzzy Decision Model

Let us assume that our decision-maker knows the set of possible impact severities for a given policy alternative but only has vague assumptions as to their respective probabilities of occurrence. Rather than numeric probabilities, he can only assign them

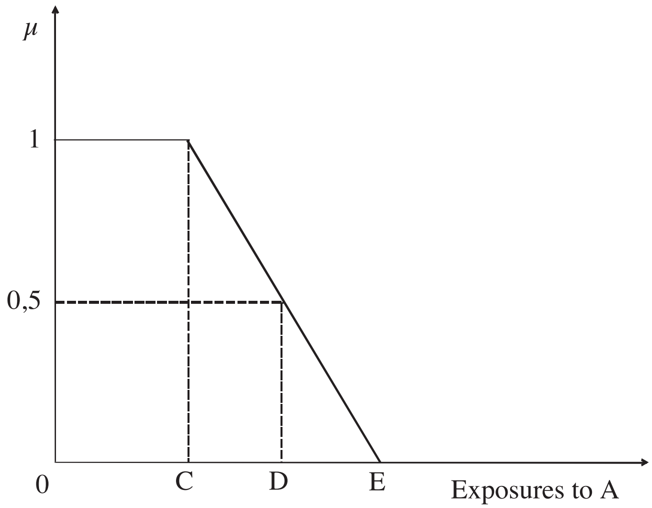

Fig. 4-6 Fuzzy impact interval with a single pollutant. In the exposure range 0 to C, the decision-maker thinks it impossible that a specific severity S* will be exceeded (rating 1). He does think the target could be exceeded at higher levels of exposure, however, at such levels, the relationship between exposure levels and the probability of compliance with targets is (arbitrarily) assumed to be linear. At exposure level 0D, the decision-maker ascribes a value of 0.5 to compliance with the target. At exposure levels of E or greater he considers compliance with the target to be impossible.

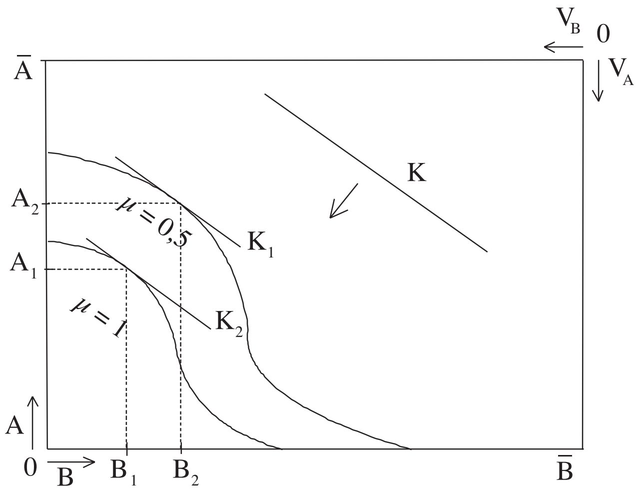

Fig. 4-7 Cost minimisation in a fuzzy model with combined effects. The isoquant curve with $\mu = 1$ plots combinations of agents A and B for which it is subjectively conceivable that impacts will remain at or below the level S*. If the decision-maker requires this degree of certainty, the efficient limit values given isocost function K_2 are A_1 and B_1. If the decision-maker accepts a likelihood of 0.5, on the other hand, the efficient limits are A_2 and B_2.

"degrees of likelihood". We map the problem with an evaluation function $\mu_{S(A,B)}$, which reflects the vague assumptions. In general, one can assign ratings to each impact severity (for more on fuzzy models, see Rommelfanger 1994). The higher the rating assigned to each impact severity, the higher its subjectively adjudged probability of occurrence. Dividing each rating number by the highest rating produces a set of values in the range 0 to 1. We will now proceed to use the values in this range to evaluate the impacts. Severities that are deemed inconceivable are rated 0 and ones the decision-maker considers most likely are rated 1. Imprecise impact ratings of this kind can be represented mathematically in the form of fuzzy impact sets.

By way of example, fig. 4-6 plots a rating function for an individual agent and fig. 4-7 for multiple agents. Ecologically equivalent combinations lie within specific areas. The area containing the set of equivalent combinations is inversely proportional to the stipulated probability that impacts will remain below a specific severity level. If the subjectively greatest possible safety level is required, combinations that come into question are to be found in area $\mu = 1$. On cost criteria, the choice falls to combination (A_1, B_1). This accords with the scientific concept of limit values, which aim to rule out adverse affects given the available knowledge. A policymaker who is satisfied with a probability of occurrence $\mu = 0.5$, on the other hand, would select combination (A_2, B_2).

Fuzzy models can be construed as a weak version of stochastic decision models, requiring less perfect information. It is sufficient for the decision-maker to state "qualitative" severity levels and probabilities of occurrence (such as "low", "medium" and "high"). Qualitative scales of this kind are common, for example, in legal risk theory on preventive environmental policy (Kloepfer 1993). Fuzzy models are thus able to work as a decision aid where stochastic models fail. There is a general demand for such methods since qualitative information is often the only kind available (Munda et al. 1994).

4.5.4 Decision Models Subject to Uncertainty

Let us now assume that the decision-maker can state impact severities but not their probabilities of occurrence. If direct comparison of the policy alternatives fails to yield an optimum combination, we are forced to rely on decision rules like the minimax, minimin and Hurwicz principles. Under the minimax principle, the decision is made in favour of the option with the smallest maximum impact. The decision-maker only considers the worst possible impacts. This brings out a pessimistic or particularly safety-conscious attitude. With the minimin principle the reverse is true. This prefers the alternative with the smallest minimum impact. fig. 4-8 illustrates the two positions. S_1* and S_2* are the conceivable isoquant curves with level S*. Area C contains combinations that are certain to observe S*. Combinations in areas D and E may observe S*. The optimist chooses (under the assumed cost con-

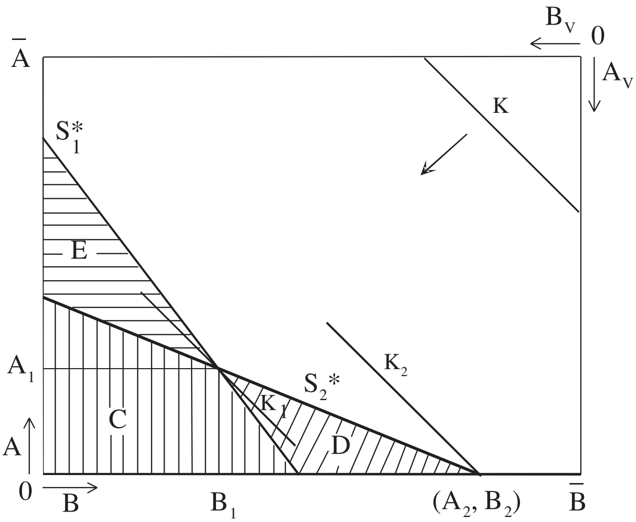

Fig. 4-8 Cost-minimization decision under the minimax and minimin principle. In the example, the decision-maker expects that his ecological objective S* will be observed by combinations of agents A and B either on isoquant S_1* or on isoquant S_2*. He can assign probabilities of occurrence to each of the two estimates. If he is pessimistic, he takes the least favourable result among the alternative combinations of A and B (the combinations in area C) and, assuming the isocost function K_1 and aiming to minimise cost, decides in favour of the combination of limit values A_1, B_1. If he is optimistic, areas C, D and E come into play and the efficient limit values are A_2, B_2.

straints) the least-cost combination in areas C, D and E, i.e. (A_2, B_2); the pessimist chooses the least-cost combination in area C, i.e. (A_1, B_1). This solution accords with the natural science concept of safety standards providing the target $S^* = 0$.

Neither of these decision rules truly lends itself to balanced treatment of environmental risks, since they each take an extreme situation in risk analysis as their starting point. Another approach is the Hurwicz principle, which uses a weighted combination of the greatest and smallest expected impact: $S_H = a \bullet S_{max} + (1 - a) \bullet S_{min}$, where $0 \leq a \leq 1$. The resulting impact function is then confronted with the given protection objective to determine the permissible least-cost environmental standard. This will be closer to solution A_1/B_1 or A_2/B_2 depending on the value assigned to the pessimism parameter (a) or to the optimism parameter (1 – a).

To sum up: in view of the wide variety of uncertainties, decisions about limit values are risk-based decisions. Environmental planning should accordingly be based on systematic risk analysis and make use of economic decision theory. In this point, economic and scientific enquiry should go hand in hand.

The economic decision models clearly show that rational policy decisions are possible with imperfect knowledge. They reveal what factors (relative impacts and relative abatement costs, the extent and measurement of uncertainty, and risk assessment) are important to decisions and how they take effect.

4.6 Conclusions

The cost-efficiency model with multiple pollutants clearly demonstrates that economic methods should be taken into account when setting environmental standards for pollutants in the case of combined exposures and that their relevance is not limited to the selection of policy instruments. The general message is that alternatives should be applied in this decision-making area as well, with the employment of impact-impact and cost-cost comparisons. Since little use is made of such considerations when setting limit values in practice, and since combined effects arise frequently, there is scope to improve current policy in the direction of more cost-effective environmental protection. In the course of reform it may be beneficial (1) to incorporate additional pollutants into policies that have so far focused on individual pollutants (climate protection policy being one example), (2) to exclude all but one pollutant from policies that have so far focused on multiple pollutants (this is likely to be an exception), and (3) to correct the balance of limit values for all agents implicated in a given adverse effect. The economic approach is of benefit even when very little is known regarding ecologically equivalent combinations of agents. For example, if only three ecologically equivalent combinations of a number of implicated substances are known, it is possible to choose between these on the basis of macroeconomic costs. Decision criteria are also available that allow rational decisions to be made about limit values even when there is uncertainty in the data.

Where two or more pollutants act both in combination and separately, separate policies should be followed that apply special limits for combined exposures and otherwise retain the existing limits for individual pollutants (see chapter 5).

The most cost-efficient option in most cases will be to incorporate all significant causal agents into policy. There may be cases, however, in which the most cost-efficient option is to concentrate on a specific pollutant. This may give rise to a conflict with the objective of equal apportionment of burdens, by which all polluters should be called to account. The general applicability rule is integral to the normative legitimization for the "polluter pays" principle. Selective apportionment of burdens causes problems when directly competing enterprises are affected differently, thus distorting free competition. The principle of polluter accountability also applies in liability law. If impacts can be systematically apportioned to polluters, each is held liable for its respective share of the total. If not, all polluters are held jointly and severally liable and any one or any number of them can be called to account for the total (though those who are called to account can claim redress from those who are not). One way of resolving the conflict between efficiency and fairness is to introduce a monetary compensation mechanism between those who come under the policy and those who are unaffected by it. Polluters exempted for efficiency reasons can be induced to share in the cost paid by regulated polluters (the compensation principle; see chapter 5).

4.7 Literature

Bayer S, Cansier D (1999) Kyoto-Mechanismen und globaler Klimaschutz: Die Rolle handelbarer Emissionsrechte, Wirtschaftswissenschaftliche Fakultät der Eberhard-Karls-Universität Tübingen, Tübinger Diskussionsbeitrag 163, March 1999, to be published in: Hamburger Jahrbuch für Wirtschafts- und Gesellschaftspolitik 1999, J.B.C. Mohr, Tübingen 1999
Böhm M (1996) Der Normmensch. Tübingen
Bonus H (1975) Möglichkeiten der Internalisierung externer Effekte als Instrument der Koordination von Unternehmenszielen und gesellschaftlichen Zielen. In: Albach H, Sadowski D Die Bedeutung gesellschaftlicher Veränderungen für die Willensbildung im Unternehmen, Berlin, pp. 207–226
Cansier D (1991) Die Bekämpfung des Treibhauseffektes aus ökonomischer Sicht. Berlin
Cansier D (1994) Gefahrenabwehr und Risikovorsorge im Umweltschutz und der Spielraum für ökonomische Instrumente. In: NVwZ – Neue Zeitschrift für Verwaltungsrecht 7/1994, pp. 642–647
Cansier D (1996) Umweltökonomie 2. edn. Stuttgart
Chao H, Peck SC, Wan YS (1994) Managing Uncertainty: The Tropospheric Ozone Challenge. In: Risk Analysis. Vol.14, pp. 465–475
Der Rat von Sachverständigen für Umweltfragen (1987) Umweltgutachten. Stuttgart
Der Rat von Sachverständigen für Umweltfragen (1994) Umweltgutachten. Stuttgart
Dieter H-H (1985) Risikoquantifizierung: Abschätzungen, Unsicherheiten, Gefahrenbezug In: Bundesgesundheitsblatt. Vol. 38, pp. 250–257
Endres A (1985) Environmental Policy with Pollutant Interactions. In: Pethig R (ed.) (1985) Public Goods and Public Allocation. Bern, pp. 165–199
Hagenah E (1996) Prozeduraler Umweltschutz. Baden-Baden
Hall D-C (1998) Albedo and Vegetation Deman-side Management Options for Warm Climates. In: Ecological Economics. Vol. 24, pp. 31–45
Kloepfer M (1993) Handeln unter Risiko im Umweltstaat. In: Gethmann CF, Kloepfer M (eds.) (1993) Handeln unter Risiko im Umweltstaat. Heidelberg, pp. 55–98
Michaelis P (1997) Effiziente Klimaschutzpolitik. Tübingen
Munda G, Nijkamp P, Rietveld P (1994) Qualitative Multicriteria Evaluation for Environmental Management. In Ecological Economics. Vol.10, pp. 97–112
Rehbinder E (1991) Das Vorsorgeprinzip im internationalen Vergleich. Düsseldorf

Repetto R (1987) The Policy Implications of Non-convex Environmental Damages: A Smog Control Case Study. In: Journal of Environmental Economics and Management. Vol. 14, pp. 13–29

Rommelfanger H (1994) Fuzzy Decision Support-Systeme. 2. edn. Berlin

Tegner H, Grewing D (1996) Haftung und Risikostandards. Strategien im Umgang mit Umweltchemikalien. In: Zeitschrift für Umweltpolitik & Umweltrecht. Vol. 19, 1996, pp. 441–463

Umweltbundesamt (1998) Daten zur Umwelt. Der Zustand der Umwelt in Deutschland. Edition 1998. Erich Schmidt Verlag GmbH & Co., Berlin

von Ungern-Sternberg Th (1987) Environmental Protection with Several Pollutants: On the Divison of Labor Between Natural Scientists and Economists.In: Journal of Institutional and Theoretical Economics (JITE) Vol. 143. pp. 555–567