Final Report

Potential Use of Economic Instruments Under the HSNO Act

for
Environmental Risk
Management Authority

16 June 2004

ER-OP-04-01 07/04

ISBN 0-478-21527-4

covec limited 66 wyndham street po box 3224 shortland street auckland

t: (09) 916-1970 f: (09) 916-1971 www.covec.co.nz

Executive Summary

This report examines the potential use of economic instruments under the Hazardous Substances and New Organisms (HSNO) Act. Economic or market-based instruments are policy tools that use market incentives to achieve objectives. This report is responding to the specific provisions under the HSNO Act that refer to the potential use of two instruments—transferable permits and environmental user charges.

Current Regulations

The current regulatory system ensures identification, evaluation and management of risks, so that risk levels are reduced. In doing so, it imposes costs on introducers of hazardous substances that increase the costs of entering the market, eg by importing a new substance. This is particularly significant and potentially problematic because it applies differently to new and existing substances. It is likely to be limiting the introduction of some beneficial new products.

The current costs of authorisation are an up-front fixed cost and, in a competitive market, may not be being passed on in product prices, because they have little or no impact on variable costs. This means the regulation provides no incentive to limit use of substances that are approved for use. And, despite the controls to limit risk, there are residual risks associated with hazardous substances.

Economic Instruments

Economic instruments can provide ongoing incentives for risk reduction and environmental improvement by increasing the costs of every unit of substance used. To correct one of the problems of the current system, economic instruments need to be applied to all products within a category, and not just new applications.

The economic effects of taxes or charges and tradable permits are broadly equivalent. They both introduce a cost at the margin, ie for every unit of product. This affects prices and levels of consumption of a product, or provides incentives for actions to limit the liability to the charge or the requirements for allowances.

Environmental Charges

The theoretically ideal instrument to address environmental externalities is a charge equal to marginal damage. However, there are considerable practical difficulties in implementing such a charge, and very few practical examples anywhere. There is no current data set that would allow either the quantification of relative damage in a suitable metric (eg monetary terms), nor are there data relating to the probability of a risk event. And levying charges on different products simply on the basis of intrinsic hazardousness of the substance does not necessarily bear any relationship to the relative probability of an effect.

In other countries, charges have been levied in a more arbitrary way to target an environmental improvement. Such an approach could be used, and in the absence of any information about likely costs, a low level charge would most likely yield an efficiency improvement by bringing private costs closer to real resource costs. However, although it would increase the prices of hazardous substances relative to other goods, there is the potential to change relative prices amongst hazardous substances in a way that does not reflect relative probability of damage.

Environmental charges raise revenues and might be used as an alternative revenue source for ERMA. They could be designed to meet criteria for good revenue raising mechanisms—they could both correct a market failure by charging an existing externality (environmental damage from hazardous substances) and could remove an existing distortionary revenue raising mechanism (a charge on first introducers of new substances that can discourage market entry of beneficial products).

Tradable permits

Cap-and-trade systems start with a binding limit on some activity or substance. Individual and tradable rights to some proportion of this aggregate limit are allocated to firms and can be exchanged. From the perspective of the firm, if there is a liquid market, tradable permits function as a price mechanism, rather than a binding physical limit on activity or production. They provide incentives for firms to find the lowest cost means of meeting the aggregate limit.

A key component of a cap-and-trade system is the cap. Currently limits to the use, import, manufacture or even total allowable effect of hazardous substances have not been introduced. Such an approach would be required to make a cap and trade system practicable. Other important design elements include approaches to allocation, decisions on the use of trading over time (as well as between firms) and compliance penalties.

Instrument Comparison

The choice between the instruments depends largely on the objectives. If targeting quantifiable physical improvements, tradable permits can provide greater certainty of outcome, but with some uncertainty over prices and cost impacts. In contrast, charges provide greater price and cost certainty, while having uncertainty over environmental outcome.

Next Steps

Introducing economic instruments would provide a positive addition to the existing regulatory system. They could yield environmental improvements consistent with the objectives of the HSNO Act, and environmental charges could raise revenue in a more economically efficient way than existing mechanisms. To design an instrument that achieves these improvements in an economically efficient way requires analysis to quantify current damage effects of hazardous substances, and the relationship to quantities of individual substances introduced to the economy. To understand expected outcomes, it also requires an assessment of the expected effects of the instruments on prices, demand and supply of specific substances.

Introducing a simple (eg per kg of pesticide) low-level charge is an option that could yield efficiency improvements as a revenue raising tool and would be likely to have positive but limited incentive effects. There is a slight risk that relative prices of substances are changed in a way that does not relate to risk levels.

Contents

1. Introduction

1.1. Objectives and Scope of the Report

This report examines the potential use of economic instruments under the Hazardous Substances and New Organisms (HSNO) Act. It examines the powers under the Act and the types of economic instrument that might be employed. It discusses the theoretical basis for use of economic instruments and examples used elsewhere, particularly those applied to hazardous substances.

The report sets out the advantages and disadvantages of using different types of economic instruments, and the market and regulatory requirements for establishment. It does not quantify the costs and benefits of introducing these instruments.

1.2. Economic Instruments and the HSNO Act

The purpose of the HSNO Act is to protect the environment, and the health and safety of people and communities, by preventing or managing the adverse effects of hazardous substances and new organisms (Section 4).

This is a balancing act. In achieving this purpose, environmental bottom-lines are defined as safeguarding the life-supporting capacity of air, water, soil and ecosystems, while, decision makers under the Act need to consider *inter alia* the economic and related benefits and costs of using a particular substance or new organism.

Currently the approach to managing effects is via a system of risk analysis and prescribed controls that apply to packaging, labelling, disposal, management of emergencies, tracking and qualifications of handlers.

Economic or market-based instruments are potential alternative or additional policy measures. They are policy tools that use private market mechanisms to achieve policy objectives. The possibility to use them arises from specific provisions in Sections 87-95 (transferable permits) and Section 96 (environmental user charges) of the Act. In addition, recent provisions under Section 77A of the Act, enable ERMA to introduce additional or substitute controls that it thinks fit, provided that the controls are more effective, more cost-effective or more likely to achieve the purpose of the control—which is controlling the adverse effects on people or the environment[1].

[1] See the definition of "Controls" under Section 2 of the Act.

1.3. Limitations of the Report

This report provides an overview of the potential use of economic instruments but does not discuss their potential application under the HSNO Act in any detail. To do so would require a much more detailed assessment of the nature of current markets for hazardous substances and a greater understanding of the existing impacts of the use of substances.

The report has also limited the discussion of use of the instruments to hazardous substances. This approach was agreed with ERMA following some initial scoping work. Although, it is possible to extend the approaches to new organisms there are limitations, particularly the absence of competitive markets.

1.4. Structure of the Report

This report provides background information that would:

- Inform a decision on the introduction of a transferable permit scheme;
- Provide material as the basis for a report under Section 96 of the Act.

Section 2 of this report describes economic instruments including the theoretical underpinnings and the way in which they work. Section 3 discusses a number of international examples of the use of economic instruments. Sections 4 and 5 examine some of the practical issues involved in implementing economic instruments under the HSNO Act.

2. Economic Instruments—What Are They and How Do They Work?

Economic or market-based instruments are policy tools that use market incentives to achieve objectives. Examples include taxes and charges, tradable permits, liability regimes and subsidies. This report is responding to the specific provisions under the HSNO Act that refer to the potential use of transferable permits and environmental user charges.

In this section we describe the different types of economic instrument in generic terms. Some introductory economics concepts are included because of their importance to discussing the effects of the instruments on prices, costs and quantities of production and use, and decisions to enter a market.

2.1. Pricing under Competitive Markets

Economic instruments function by changing costs and thus prices; understanding price-setting and consumption is important to understanding their effects. We discuss pricing issues initially under ideal market conditions. In Section 4 we describe actual markets for hazardous substances, and how market structure may alter the functioning of the economic instruments.

The first concept that is useful to understand is marginal cost pricing. Under competitive markets, prices of goods and services are expected to be defined in the market through the interaction of costs of production (supply) and consumers' willingness to pay for products (demand), and to equal the marginal costs of supply—defined as the cost of supplying one more unit of product[2]. When prices equal marginal costs, goods are allocated in an economically efficient way, because successive units are consumed only if the value to the consumer is greater than or equal to the cost of supply, and producers will continue to produce so long as consumers' willingness to pay is at least equal to the costs of production.

Ideally charges and transferable permits are specified in a way that changes marginal costs. They can ensure that market prices equal the true marginal cost of supply, taking account of wider costs to society, or they can ensure that some desired quantity of consumption is achieved at least cost.

[2] However, where there are economies of scale, short run marginal cost pricing will not be sufficient to cover a firm's fixed costs.

Figure 1a represents pricing under a competitive market. The supply curve (S) is upward sloping and it represents the price at which firms are willing to supply another (marginal) unit of a specific product—this can be interpreted in the short and long run.

Figure 1a: Competitive Pricing

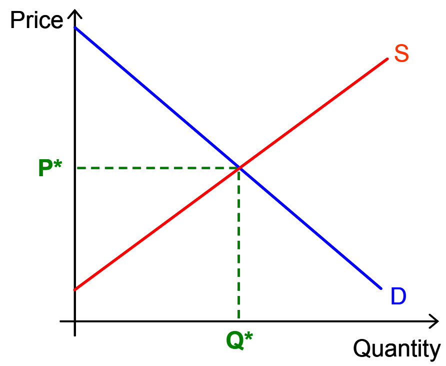
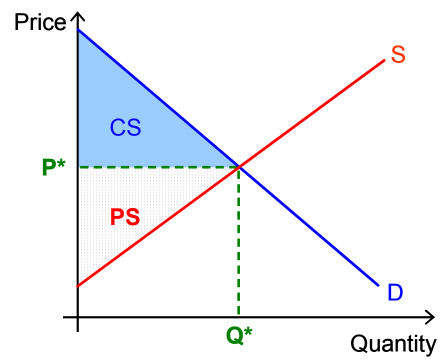


Figure 1b: Surpluses



Short-Run Interpretation

Short run marginal costs (SRMC) just include the variable costs of production. For example, the marginal costs of producing a chemical such as chlorine would include only the costs of the raw materials—salt and electricity[3].

Within individual firms or plants, marginal costs of supply may be constant over large volumes of output. However, costs will often increase as output approaches capacity but, more importantly, marginal costs increase as output shifts to the next (less efficient) producer[4]. Under competition, firms are willing to increase output, provided that the sales price covers their marginal costs of production. In making this decision, fixed costs, including costs of capital, are ignored in the short run, because they are unavoidable. Thus:

- short run marginal costs determine the level of output or production of a product from an existing plant or process; and
- firms compete on the basis of their marginal production costs.

The demand curve (D) represents the quantity consumers are willing to purchase at various prices. Some consumers will be willing to pay higher prices than others; the demand curve represents the marginal willingness to pay as we move from consumers willing to pay the most to those willing to pay the least for a product.

[3] In practice, estimating the marginal cost of chlorine production is complicated because of the production of co-products, caustic soda (sodium hydroxide) and hydrogen.
[4] The supply curve assumes that, if demand for a product was much less, it would be supplied by the lowest cost plant. As demand increases, the next cheapest plant provides supply, and so on.

Under competitive markets, an equilibrium is reached—a price (P^*) and a quantity (Q^*) of output at which demand is exactly equal to supply. At this point, no firm is willing to supply any more units unless the price increases, and no consumer is willing to purchase any more unless the price falls.

From this we can estimate the total national benefits of this market (Figure 1b). It is made up of two elements:

- consumer surplus (CS) equal to the difference between the price that consumers would be willing to pay for a product (D) and what they actually had to pay (P^*); and
- producer surplus (PS) equal to the difference between the costs to producers of producing a unit of product (S) and what they received (P^*).

The sum of consumer and producer surpluses represents the total national value of that quantity of production and consumption, over and above the costs of production. Competitive markets can ensure that the combined size of these two surpluses is maximised.

Long-Run Interpretation

These concepts can also be used to explain longer term effects. The long run marginal costs (LRMC) of supply can be thought of as the costs of supply from the next plant or product to come on line to meet demand; long run marginal costs include short run marginal costs plus an amount to cover annual fixed and capital costs over the lifetime of the plant; it is equivalent to an average incremental cost of supply. Firms are willing to enter a market, or to introduce a new product, provided that the expected sales price, times expected output, covers the costs of entry and production; this would include the costs of meeting the regulatory requirements of the HSNO Act.

Thus market entry is determined by the average costs of production. And competitive market prices will generally trend to long run marginal costs, at least on average[5].

[5] Where there is excess current production capacity there is competition at the margin and prices may fall to SRMC, but as demand approaches potential production, there is less competition and firms can raise prices; they are limited by the costs of new entry, as raising and sustaining prices above that point will result in new firms entering the market. If demand exceeds supply (or is close to doing so), firms enter and can price to recover their costs, again limited by the potential for additional firms to enter. If new entrants have lower (average) costs than the SRMC of existing suppliers, these firms will enter the market, even when there is excess current capacity.

2.2. Externalities

Competitive markets produce economically efficient outcomes. But they fail when some costs (or benefits) borne by society are not taken into account.

An externality is a cost or a benefit affecting someone other than the buyer or seller in a market, such as the environmental or health effects of hazardous substances.

Where externalities are negative, we can think of the effect as an additional external or social cost of supply[6]. For every unit of product, the external cost would equal, say, the probability of a hazardous event associated with that product, times the cost of the hazardous event.

Figure 2a shows the effects of internalising external costs as a higher supply curve—the marginal social cost of supply. Whereas the private market price was P^* resulting in consumption of Q^*, the socially optimal price is P^s. This is the equilibrium market price that results if the marginal costs of supply include the damage costs associated with the production of an additional unit; it results in a socially optimal level of consumption of Q^s.

Figure 2a: Social Cost of Supply

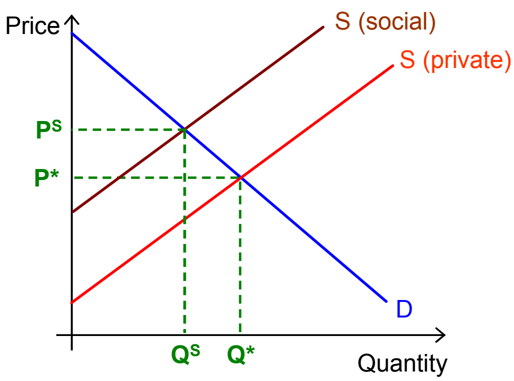


Figure 2b: Effects on total surplus

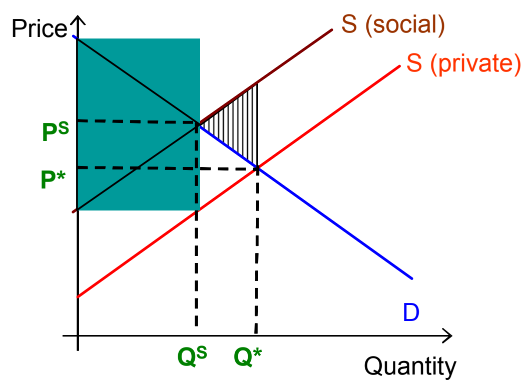


Figure 2b illustrates the true social value of production—the shaded triangle to the left of Q^s, rather than the larger triangle pictured in Figure 1b. And consuming Q^* (which is what results when prices do not reflect social costs) rather than Q^s, has a cost for society equal to the small shaded triangle to the right of Q^s; this is the difference between the (higher) social costs of producing Q^* and the value to consumers, represented by D.

[6] We use the term "social" to represent the effects on society as a whole

2.3. Environmental Charges

We first discuss the functioning of environmental charges in theory; we discuss some examples of their use in Section 3, before examining the way in which they might function under HSNO in Section 4.

2.3.1. Internalising Externalities

Environmental charges operate by increasing the cost of a unit of production. Theory would suggest that the optimal environmental policy instrument is a charge on each unit of production (or consumption) equal in size to the marginal damage cost[7]. For a charge on a production externality, this would result in the effect illustrated in Figure 2a above. Costs of supply would increase to equal the social cost of supply, prices would rise and the higher equilibrium price would result in reduced levels of consumption. In theory, if the damage costs can be measured accurately, this is the best outcome for society and results in an optimal level of environmental damage (see Box 1).

At Q^s, the value to the marginal consumer of the product is exactly equal to the marginal cost of supply, including the damage costs.

The UK landfill tax was originally set this way[8]. A study of the environmental damage associated with landfilling different types of waste[9] was used to develop a charge with two levels—one for standard waste (originally set at £7 per tonne) and a lower rate for inert waste (originally set at £2 per tonne). The scheme is described in more detail in Section 3.3.

A UK tax on aggregates was also developed using an analysis of environmental damage costs, and similar approaches were used in analysing a possible water pollution tax[10], although this has not been adopted.

---

[7] Baumol WJ and Oates WE (1988) The theory of environmental policy. 2nd Ed. Cambridge.
[8] Prime Minister's Strategy Unit (2002) Waste Not Want Not. A strategy for tackling the waste problem in England. www.number-10.gov.uk/su/waste/report/ downloads/wastenot.pdf
[9] CSERGE, EFTEC and Warren Spring Laboratories, (1993). *Externalities from Landfill and Incineration*, HMSO, London.
[10] Environmental Resources Management (1999) *Economic Instruments for Water Pollution Discharges*. Department for Environment, Food & Rural Affairs. www.defra.gov.uk/environment/water/quality/econinst2/contents.htm

Box 1: Policy Objectives—Economic Efficiency

Economics treats the optimal point that society can achieve as an economically efficient outcome. This is based on the original definition posed by Italian economist Vilfredo Pareto as a point at which no other distribution of goods could be achieved that would make someone in society better off without making someone else worse off. But this pareto optimality criterion is difficult to achieve—changes usually make some people better off while making others worse off, eg pollution associated with production or consumption of a commodity. The criterion is modified by an additional test—can the winners compensate the losers? From the perspective of society as a whole, it does not matter if there are losers, so long as the winners would be able to compensate them. This can happen if the benefits to the winners are, in aggregate, greater than the costs faced by the losers. And this is the result if we can measure the damage costs of a pollutant, charge this cost to users of a product, and consumers are willing to pay this additional cost. The value of the product to consumers must be at least as high as the costs of production, plus the environmental damage.

Note, under this definition (know as Kaldor-Hicks efficiency), it does not matter if compensation is not actually paid to the losers. What is important is that it could have been paid. This then makes society as a whole better off and better able to tackle any distributional issues, such as occur when certain communities bear greater proportions of environmental damage.

In contexts where these redistributions are never actually made, policy makers have sometimes concerned themselves with the distributional impacts of environmental policy. Hence environmental justice (EJ) is an increasing concern particularly in the US, where market approaches to environmental policy have resulted in some concentration of environmental effects in identifiable communities.

If markets are competitive, and the level of the charge truly reflects the environmental damage cost, policy makers might be indifferent to the outcome. In other words, if the imposition of the charge leads to very little change in behaviour and little change in environmental effect, it suggests that the commodity is highly valued; its value to society is sufficiently high that the community is willing to accept some level of environmental damage in order to continue consumption of the product. In making this statement, we are using a specific approach to measuring what is optimal for society (see Box 1).

However, the examples of efficient taxes (equal to damage cost) given above are atypical. Very often, marginal environmental damage costs are difficult to estimate and more arbitrary methods are used to select the size of a charge, or they are set with reference to analysis of levels required to have a desired effect.

2.3.2. Incentive Effects

Environmental charges are often used to achieve pre-specified outcomes rather than to achieve some notional optimal level of environmental improvement. Environmental charges can have a number of outcomes that include:

- reduced consumption, because of the increased market prices that result from increased marginal costs of supply—the effect depends on the price elasticity of demand (see below);
- investments or management changes that result in reduced charge liability and thus reduced environmental damage, eg using cleaner production methods or changing substance ingredients;
- reduced market entry because of increased average costs; and
- market exit—firms ceasing production because they can no longer cover costs while selling at a competitive price.

An example of an incentive charge is the pesticides charge introduced in Denmark as part of a government strategy to reduce pesticide use. The charge level was set in order to have an effect. It is discussed in more detail in Section 3.2.

In Sweden the level of a tax on nitrous oxide (NOx) emissions was set at a level expected to result in plants fitting low-NOx burners (LNBs) [11]— control devices that result in lower emission rates at combustion plants. The decision to install an LNB is based on the comparison of total costs of control, discounted over the expected lifetime of the equipment, with the costs of paying the charge over the same period.

To ensure targeted outcomes, charges have also been set in tandem with other instruments. For example, the UK Landfill Tax was set on the basis of estimates of damage cost, which in theory, should result in the optimal rate of final disposal, of recycling and of product redesign. However, the government has taken additional steps to achieve recycling targets, particularly through a set of regulations covering the packaging industry.

Elasticities

The environmental effects of a charge depend on the price elasticity of demand for the goods and the competitiveness of the market. The price elasticity of demand is a measure of the change in consumption that results from a change in price.

Figures 3a and 3b show the market outcome of an environmental charge on two goods with different demand elasticities. In Figure 3a, demand is relatively elastic—a change in price results in a significant shift in demand. Because of the demand elasticity, not all of the

[11] Anderson RC and Lohof AQ (1997) The United States Experience with Economic Incentives in Environmental Pollution Control Policy. US Environmental Protection Agency

charge is passed on to consumers—the difference between P^0 (equilibrium price without the charge) and P^1 (equilibrium price with the charge) is less than the size of the charge.

The price increase that is passed on leads to a reduction in the quantity consumed (from Q_0 to Q_1), and thus reduced environmental damage.

Figure 3a: Elastic demand

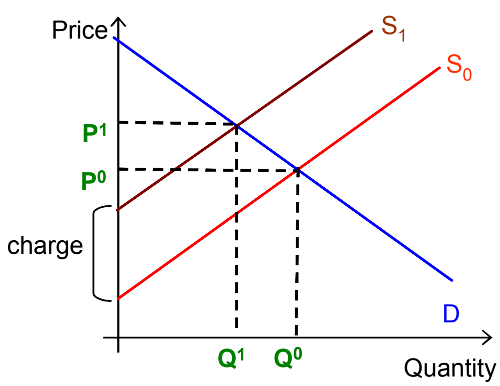
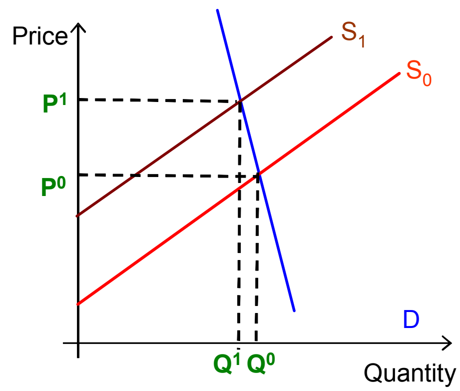


Figure 3b: Inelastic demand



In contrast, Figure 3b illustrates the case with inelastic demand. Here more of the charge is passed on to consumers but the resulting reduction in quantity consumed, and thus in environmental damage, is small.

Long run elasticities can be different from short run elasticities[12]. In the short run, price increases can only alter the way in which existing assets are used; in the long run, price increases change the types of assets that are invested in.

Uncompetitive Markets

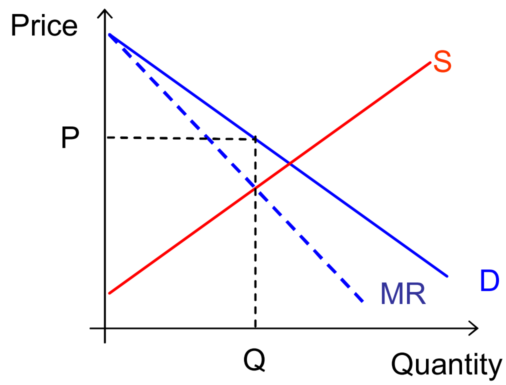
The other factor that will result in differences in the effects of an environmental charge is the competitiveness of the market. To this point we have discussed the way prices are set and the relationship to consumption in competitive markets, ie those with numerous buyers and sellers, and no barriers to trade. Markets for many hazardous substances do not meet this description.

There are some generic products that come close to fully competitive markets but many products are introduced under patent, by monopoly suppliers or for very limited demand by one or two companies.

[12] In the MED's energy model, the long run elasticity of demand for petrol is approximately three times as high as the short run elasticity. Ref: Fuels & Energy Management Group Ltd (2000) Road transport sector energy demand & CO_2 output. Projections & Analysis of Reduction Strategies. Appendix C: Energy Demand Projection Modelling Procedures.

Price setting under monopoly or near monopoly markets differs from competitive markets. Quantities of production will be determined by the interaction of the marginal revenue curve and the marginal cost of supply. Marginal revenue is the change in revenue with a change in sales. It falls with quantity produced because of the lower willingness to pay for the product by the marginal consumer, and because the lower price is spread across all sales. Thus the marginal revenue curve is increasingly lower than the demand curve as output increases (Figure 4). Price is set at a level that limits consumption to a quantity at which marginal revenue = marginal costs of supply(S).

Figure 4: Pricing in uncompetitive markets



Under such markets, if supply costs are increased via an environmental charge, this will result in a smaller change in price than under a competitive market (Figures 5a and 5b). However, it may mean that prices are already above the socially optimal and the policy problem is not an environmental one; too little of the commodity is being consumed, not too much. But the change in environmental quality attributable to the instrument can be considerably reduced[13].

Figure 5a: Competitive market

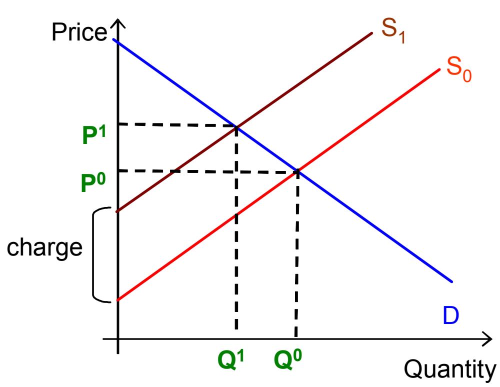
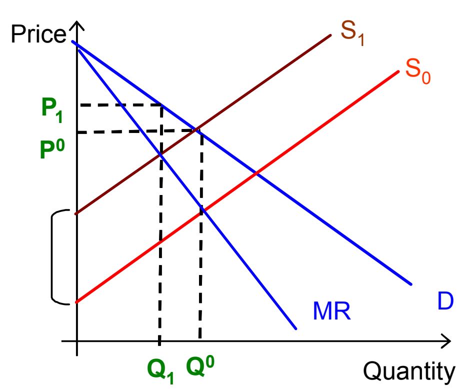


Figure 5b: Monopoly Market



[13] Note this is not the same as a reduced environmental outcome, because the counter-factual is different.

2.3.3. Revenues

Environmental charges produce revenues. The use that is made of the revenues involves a separate decision from the decision to introduce the charge. In practice, often some, at least, of the revenues are retained to achieve environmental objectives related to the objectives of the charge, eg some portion of the UK's landfill and aggregates tax revenues are allocated to related projects.

- Landfill site operators who contribute to organisations "with objects concerned with the environment, enrolled under the Landfill Tax Credit Scheme", may claim a credit of up to 6.8% against their annual landfill tax liability[14].
- Some of the revenue from the Aggregates Levy is contributed to a Sustainability Fund used to finance programmes to minimise demand for primary aggregates, promote environmentally friendly extraction and transport, and reduce the local effects of aggregate extraction[15].

Such targeted use of funds, referred to as hypothecation, can remove some of the economic efficiency gains from using the instrument.

As noted above, if a charge is imposed at a level equal to the marginal damage cost, the resulting market behaviour—supply and consumption of the good on which the charge is levied—is optimised. The theoretical best use of the revenue collected is to correct some other market distortion, either in the form of a subsidy, where this is the best market correcting approach[16], or more likely, to replace income raised through a distortionary tax, eg a tax that has been levied to raise revenues rather than to correct a market failure/externality. Taxes are distortionary when they are levied on goods or services, or on income, in a way that changes behaviour from what it would have been in the absence of the tax. For example, taxes on labour mean that, to attract workers, wage rates need to be higher than they would be otherwise, and firms employ less labour and use more other resources. Using revenue from environmental taxes to reduce such taxes results in the so-called double-dividend of corrective taxes or charges. There is one social dividend (benefit) from correcting the externality; there is a second social dividend from reducing other taxes (or from correcting another market failure). This approach has been championed in a number of countries, most notably via the call in European Community President Jacques Delors' 1993 White Paper on growth, competitiveness, and employment, for a shift in the burden of taxation within the EU from "goods" to "bads"[17].

[14] HM Customs & Excise. www.hmce.gov.uk/news/bb-0904.htm
[15] Environment Protection Economics Division (2003) Mid Term Evaluation of the Aggregates Levy Sustainability Fund. Department for Environment, Food and Rural Affairs. www.defra.gov.uk/environment/waste/aggregates/pdf/mte.pdf
[16] For example, where there are positive externalities, or benefits for the wider community. Fireworks would be a trivial but illustrative example of a hazardous substance with positive externalities.
[17] White Paper on growth, competitiveness, and employment: The challenges and ways forward into the 21st century. COM(93) 700 final. Brussels, 5 December 1993

Consistent with this, when the landfill tax was introduced in the UK, it was accompanied by a 0.2% reduction in employer national insurance contributions (NICs); and the aggregates levy was accompanied by a further 0.1% cut[18].

In contrast, when, as a result of the additional revenue raised, the government spends more, the test, from an economics perspective, is—is this additional expenditure well-being enhancing? In other words, do the benefits to society outweigh the costs?[19] Even where additional spend is justified, there are arguments for separating the expenditure from the revenue raising tools. More specifically, revenue raising is uncertain, which means too much, or too little can be spent on the new activities. Similarly, reductions in distortionary taxes can be done to too great or too little an extent. The reduction in national insurance contributions associated with the UK's landfill tax, for example, has consistently been greater in value than the amount raised by the tax[20].

The unpredictability of revenues from environmental taxes is compounded by the incentive effects that they have. The principles of good taxation for revenue raising purposes, which relate to the low levels of distortion (eg by taxing goods with low price elasticity of demand) can be at odds with the objectives for environmental taxation, which can be to distort, ie to change patterns of consumption.

Hypothecation has been suggested as a means to limit the impacts of the economic instrument on specific sectors. For example, the tax reduces firm revenue and profitability and this can be partly compensated through lump-sum or some other compensation. The efficiency (and environmental) impact of the instrument remains, provided the compensation is not on the same (marginal) basis as the original instrument. For example, if there is a charge on each kilogramme of production/import of a hazardous substance, and pricing reflects marginal costs, lump sum compensation would not reduce the price impact, or the incentive effect of the instrument. It would simply enable levied firms to retain profits[21]. In Denmark, simultaneously with the introduction of the pesticides tax, property taxes were lowered for agriculture properties.

[18] Employees pay 11% of income between £4,615 and £30,940 per year, and 1% of income above this limit. Employers pay 12.8 % on all earnings above £4,615 a year. The link is somewhat confused over time as NICs have subsequently been raised to pay for additional government expenditure on the National Health Service. There is an efficiency gain if, because of the Landfill Tax and Aggregates Levy, other taxes are lower than otherwise they would be. This is almost impossible to test.
[19] This is a necessary but not sufficient reason for intervention which should also be justified on the basis of market failure.
[20] HM Treasury and Department for Environment, Food & Rural Affairs (2002) Possible changes to the Landfill Tax Credit Scheme: Consultation Paper.
[21] It does over-reward the industry and over-encourage investment, if continued over the long term.

2.3.4. Cost Recovery

Cost recovery is a specific example of a tied charge, ie the charge is levied specifically to recover the costs of a regulation, eg to pay for assessment of a hazardous substance or new organism.

Charges are prescribed in this way under Section 21 of the HSNO Act.

Here, we are assuming a prior choice is made to implement the regulation and thus to incur the costs. The important economic consideration here is whether the charge is an efficient way to raise revenue.

Tax theory suggests that the best way to raise revenue is to cause least distortion. In the absence of tax, people will be making choices about their expenditure in a way that maximises their well-being, taking account of market prices (that reflect costs of delivery) and the benefits that they obtain. Levying taxes changes effective prices, changes patterns of expenditure and reduces well-being. Changes to patterns of expenditure are minimised when tax levels are small and/or when the goods or services taxed have low price elasticity of demand; here changing the price via tax has minimal impacts on expenditure patterns and well-being.

In addition, taxes that correct externalities are efficient, because, although they may distort behaviour, they distort it towards a pattern of expenditure that is desirable—it is the pattern of expenditure that results from prices fully reflecting social costs so the new pattern is more optimal (using the definition outlined in Box 1 above).

2.3.5. Transaction Costs

Environmental charges can have significant transaction costs, ie administrative and other costs involved in implementation, particularly because of greater monitoring requirements. Emission charges have required much more sophisticated systems for measuring emission rates, for example.

Some approaches to the development of charging systems for hazardous substances would require more detailed approaches to the definition and measurement of hazard associated with individual substances. The level of transaction costs will often determine who a charge is levied on. For example, charging at a very disaggregated level, eg individual farms to monitor use of a product rather than the importer, will have much more significant transaction costs per unit of tax or damage. These additional costs reduce the available revenue from the instrument that might be used to reduce other market distortions.

2.4. Transferable Permits

Whereas environmental charges give some price certainty but often uncertainty of outcome, transferable or tradable permits provide certainty of outcome but uncertainty of price. They come in two basic forms:

- Cap and trade, as used in the US acid rain programme;
- Credit-based systems, as used in the UK packaging and renewables systems.

2.4.1. Cap and Trade

Where there is some overall limit on emissions, damage, industrial activity, production or use of a substance, tradable permits or allowances[22] can provide market incentives for discovery of the least cost means to meet this limit. A binding cap is converted into individual allowances to emit, damage, import, produce or use smaller units (eg one tonne) of the product.

Where there are fewer allowances available than aggregate emissions/imports would be under business as usual, allowances have a monetary value. For the US SO_2 emissions trading system, in comparison with their own costs of limiting emissions, plant owners are willing to buy allowances at a price equal to or below this cost, and are willing to sell allowances at a price equal to or above this cost. Plants that reduce their emissions can free up allowances that they can sell, and reduce their own and aggregate demand for allowances.

Prices of allowances are set in the market in the same way as for other commodities. They are expected to equal the marginal costs of supply. In the case of the US SO_2 allowance market, this is the marginal cost of emission controls. For an overall limit on hazardous substances, it would be expected to equal the additional costs of an alternative but inferior or more costly substance, or the costs of doing without the substance.

Firms within the same industry will value allowances differently, reflecting their specific costs of using substitutes, and they may require different quantities. This is the basis of trade. Allowances will be bought and/or held by those firms that value them the most.

The required elements of a cap and trade system are listed in Box 2.

Other elements that can be introduced include the ability to trade over time as well as space. For example, an allowance initially allocated for this year might be unused and held over to use next year. This is termed banking. Banking can ensure a smoother price path over time—there is no end of year price spike or trough because of under or over-supply—and it improves compliance; holding too many allowances has a low cost as the allowances

[22] 'Allowances' is the term used in the US trading systems. It differentiates between a 'permit', which allows a plant to emit at a certain rate annually, and an 'allowance' to emit a single tonne.

have a value past the compliance date. Borrowing involves using this year an allowance initially allocated for use in a future time period. Borrowing can similarly help to smooth the price path, but does not have the same compliance benefits—it can enable firms to avoid the actions that the policy measure is targeting; this is problematic especially where there is limited liability.

Box 2: Elements of a Cap & Trade System

The Cap
A cap is essential to the cap and trade system. For hazardous substances, a limit would need to be set on the total level of damage risk, or quantity of import, production and/or use of a specific substance.

Unit of Trade
The aggregate limit, eg an overall tonnage limit on imports, is converted into individual rights to import a single tonne. These rights or allowances can then be owned, bought or sold. The allowances will generally be unique, numbered commodities and will be specific to a time period, eg a single year.

Allocation
Individual allowances can be given away initially or sold (eg by auction). This has distributional effects but is not expected to affect the final market price of allowances. The US SO_2 system included a mix of both approaches; auctions were useful initially in helping firms to identify the market price.

Non-compliance penalties
The market value of allowances is determined by the penalty for not holding sufficient allowances. Under the US SO_2 system penalties have been set at a level approximately ten times the market price of allowances.

Compliance period
The compliance period is the time period for which the holding (or surrendering) of allowances must at least equal the emissions or import levels. Typically this will be a single year, although inter-year trading may be possible also. For international greenhouse gas emissions trading, the compliance period is five years, as defined under the Kyoto Protocol.

Tracking System
To achieve compliance, and the integrity of the system, cap and trade schemes maintain records of the individual allowances and who owns them. Allowance registries are used to register trades and current holdings.

2.4.2. Credit-based systems

A credit-based system works to encourage desired activities or outcomes rather than placing a limit on undesirable outcomes. Allowances are not distributed initially. Rather, they are created. The UK packaging recycling system is based around tradable recovery notes, generated when a volume of waste is recycled. These are used to demonstrate compliance with an obligation to achieve targeted rates of recycling. Similarly, the UK's Renewables Obligation enables companies that generate a MWh of electricity from renewable sources to produce a Renewable Obligation Certificate (ROC) that can be purchased and used to demonstrate compliance by local electricity suppliers with renewable supply targets.

Some greenhouse gas trading schemes, such as the means to comply with the provisions of the Kyoto Protocol, include a mixture of cap & trade (emissions trading) and credit-based systems (joint implementation and the clean development mechanism, both of which involve the calculation of emission reductions associated with specific projects).

Credit-based systems can be highly complicated because of the need to establish a counter-factual, ie a scenario of what would have happened otherwise. This has led to very significant transaction costs in defining levels of reduction from greenhouse gas projects. Here credit-based systems are used to reward emission reductions. But to define a reduction requires, for example, the definition of levels of expected energy efficiency improvement in an industry alongside the specific achievements of an individual firm.

In contrast, examples such as the recycling and renewables schemes noted above, are introduced far more readily, because the counter-factual is assumed to be no recycling and no renewables generation.

2.5. Economic Effects of Charges and Tradable Permits

The economic effects of taxes/charges and tradable permits are broadly equivalent.

They both introduce a cost at the margin. This affects prices and levels of consumption of a product, or provides incentives for actions to limit the liability to the charge or the requirements for allowances.

2.5.1. Efficiency Improvement

The theoretical basis for efficiency improvement is clear. For charges equal to damage costs, private decisions are expected to lead to efficient allocations of resources in the economy. For tradable permit systems, the flexibility provided to firms ensures that least cost means can be discovered for achieving a given government policy requirement. In practice, some of the theoretical advantages of economic instruments such as tradable permits can be lost through poor design, including following political interference to protect individual firms and industries[23]. However, ex-post analysis of the functioning of tradable permits has demonstrated that, when well designed and with low barriers to trade, a very high proportion of the predicted efficiency gains can be achieved[24]

[23] Stavins RN (1995) Transactions Costs and Tradeable Permits. Journal of Environmental Economics and Management 29:133-148.

[24] Kerr S and Maré D (1997) Transaction costs and tradable permit markets: the United States Lead Phasedown. College Park: University of Maryland. In: Newell and Rogers (op cit)

Analyses of the costs of environmental policy have confirmed the cost reductions associated with the use of economic instruments[25]. It is estimated, for example, that the cost saving attributable to formal trading under the US SO_2 system versus a uniform emissions rate standard is approximately $250m annually in 1995–2000, and that it would be approximately $784m annually from 2000 onwards, when the cap is tighter.[26]

2.5.2. Private Costs

While an incentive-based charge can be effective, it can result in higher costs for industry than traditional regulatory means. Taking the NOx charge example discussed above; under a regulatory requirement to meet environmental standards, plant owners face the costs of installation and running of low-NOx burners or other control devices. Under an environmental charge devised to provide this incentive, plant owners face these same costs—the Swedish charge is set at a level that ensures that firms install this equipment—but they also face a charge on the residual emissions.

This is still an efficient outcome where the charge is equal to damage cost. If the plant could not afford to pay the charge from its residual profit, it means that the value to society of this plant is less than the damage it causes.

2.5.3. Competitiveness

Impacts on competitiveness are frequently a concern of policy makers. This can be misplaced. The objective of economic instruments is to reduce the competitiveness of firms that cause more damage. Firms compete on the basis of their marginal costs of production; economic instruments increase marginal costs because there is damage associated with marginal production and/or consumption.

The legitimate concern of government is with fair competition, ie that competition among firms is based on the true economic costs of production[27]. The issue of fair competition arises in the case of industries for which regulation in one country is more stringent than in another, eg an environmental charge in new Zealand increases costs for agricultural production above those in other countries. Here not imposing an environmental charge results in over-production if comparing true social costs with the market price (value) of the product. But where that price is set in international markets in which there are no environmental charges, imposing one in New Zealand may result in unfair competition and under-production.

---

[25] See, for example, Winston Harrington, Richard D. Morgenstern, and Peter Nelson (1999) On the Accuracy of Regulatory Cost Estimates. Resources for the Future Discussion Paper 99-18.

[26] Carlson C., Burtraw, D., Cropper, M. and Palmer, K.L. (2000), 'Sulfur Dioxide Control by Electric Utilities: What Are the Gains from Trade?', *Journal of Political Economy*, **108**:6, 1292–326.

[27] Kerr S (2000) Fair Competition and Annex B Trading. In: Kerr S (Ed) Global Emissions Trading. Key Issues for Industrialized Countries. Edward Elgar. pp:157-184.

3. Examples of the Use of Economic Instruments

Rather than a comprehensive review of their use, this section describes a number of examples of economic instruments to illustrate the way in which they might be developed and used. It uses examples from the most relevant applications of these instruments.

3.1. UK Pesticides Taxes or Charge

In the UK a pesticides tax or charge was considered as a means to provide incentives for environmental improvement, but was not introduced. Instead a voluntary agreement has been developed with the pesticides industry. The case for a pesticides charge is being kept under review, particularly with consideration of the effectiveness of the voluntary scheme. In this section we describe the scheme as designed for the Department for Environment, Food and Rural Affairs[28].

3.1.1. Objective

The analysis of the instrument noted that approvals and enforcement systems are incapable of eliminating all environmental impacts of pesticides. The charge was investigated as part of a government pesticide minimisation policy, with the objectives of:

1. securing reductions in pesticide use;
2. achieving substitution by less harmful pesticides or alternative techniques;
3. increasing efficiency of use;
4. providing incentives for environmental protection measures, eg buffer strips by watercourses.

3.1.2. Relating the charge to damage

In the UK, pesticide use is governed by an approval process which balances risks and benefits; there is already a levy to pay for the approval process.

Six charge options were analysed. These were banded (different rates for different intrinsic hazard levels) and non-banded variants of ad valorem, per kg and per dose charges (Table 1).

[28] ECOTEC (1999) Design of a Tax or Charge Scheme for Pesticides. Department for Environment, Food & Rural Affairs. www.defra.gov.uk/environment/pesticidestax

Table 1: Tax/charge options included in impact assessment

Non-banded	Banded
Option 1 — ad valorem	Option 4 — ad valorem
Option 2 — per kg	Option 5 — per kg
Option 3 — per dose	Option 6 — per dose

Two approaches were considered for ranking pesticides to allocate them to bands. The first was a compound indicator using:

- WHO toxicity class to represent potential human/mammalian health impacts;
- risk to aquatic species represented by LC_{50} for trout and LC_{50} algae;
- risk to soil beneficials using the EC_{50} worms;
- risk to bees using the LC_{50} bees;
- risks to birds using the LD_{50} quail/mallard ducks[29]; and
- an index representing the hazard posed to groundwater.

Each parameter was classified into high, medium or low categories with scores of 5 for high, 3 for medium and 1 for low[30]. The second approach used expert opinion and a Delphi process.

These approaches were used to classify pesticides by potential impact; they measured the intrinsic hazard of the chemical, independent of quantity. The hazard rankings could then be used alongside a tax or charge established on a weight (per kg) basis. A per dose tax was considered also, ie the tax rate for a given product was estimated from the number of recommended doses. However, it was argued that damage was less likely to be related to dose levels than to weight of product used, given the way that the damage categories were defined.

3.1.3. Taxes or charges

Charges were defined as payments for specific purposes, whereas taxes were defined as means to raise general revenue with no related service. Two approaches were considered:

- a statutory charge with a licensing system; and
- a product tax

[29] the lethal dose or concentration of the product required to have a lethal effect on 50% of the quail/ducks under examination

[30] This was based on a system developed by Kovachs J, Petzoldt C, Degni J and Tette J (1992) A Method To Measure The Environmental Impact Of Pesticides. *New York's Food and Life Sciences Bulletin* No 139, 1992. In ECOTEC (1999) Design of a Tax or Charge Scheme for Pesticides. Department for Environment, Food & Rural Affairs

The statutory charge is efficient only if the level of service (licensing the pesticide) was related to the rate of use of the pesticide (and thus damage); it was thought that this might be possible. However, a product tax is likely to be more efficient because it directly targets rate of use.

3.1.4. Point of Collection

Options considered included:

- at the point of manufacture, the simplest approach but it does not include imports and would need to be modified to exclude exports;
- final sale, which has the potential to have disproportionally significant impacts on small retailers because of assumed scale economies in compliance;
- point of use by farmers and others, thought to have very significant administrative costs and inadequate data.

The taxable event should minimise evasion and facilitate enforcement. The scope for manufacturers and retailers to avoid the tax is less than for final users, because of visibility, potential for cross-checking through other sales data and the smaller number of entities. The tax on manufacturers was the favoured approach, but it was suggested that the tax event be worded carefully to ensure that imports were included. The suggested wording was to define the point of tax collection as the first point of sale or use in the UK. This would include both manufactured and imported products, while excluding those exported.

3.1.5. Impacts of the tax

The analysis required data on hazard, price and quantity for individual pesticides; this was available for 42 pesticides, or approximately half of all sales in the UK.

The estimated response to pesticide price changes was complicated because of how decisions about pesticide use are made. ECOTEC quoted a study[31] which suggested that only 20% of expenditure on pesticides in the UK was made on the basis of decisions taken by the farmer; the largest proportion (66%) was down to the agricultural advisor[32] and the remainder to other farm staff. And 99% of farmers who claim to make their own decision, consulted with advisors prior to making the decision. These factors may be part of the reason for estimated low price elasticities of demand for pesticides, ie there is a small change in consumption associated with a change in price.

---

[31] Hearn S (1997) Communicating with Farmers, *Synopsis of a Paper Prepared for the Pesticides Forum*, February 5 1997, PF/15 In: ECOTEC (1999) Design of a Tax or Charge Scheme for Pesticides. Department for Environment, Food & Rural Affairs

[32] And more than 50% of these are employed by distributors of pesticides

The effects of a tax or charge were further complicated by the characteristics of the markets, which are not particularly competitive. Markets exist for:

- **products sold under patent**, which compete with un-like products (which may or may not be close substitutes). These are the domain of larger R&D oriented companies;
- **other products which may be sold only by one company but may be off patent**; the company concerned may have sufficient expertise in manufacture and be able to maintain the product concerned as a core product for many years (often also the larger R&D companies); and
- **generic products**; these are products that are off-patent and are sold by different companies, and for which price competition may be more likely. Smaller companies that are less R&D intensive play a (relatively) greater role in this area.

The market share of generics is expected to increase over time, but in general, this suggests that this is not a market under equilibrium in which the effects of a tax can be easily understood by conventional micro-economic analysis of competitive markets.

Nevertheless, ECOTEC estimated on the basis of published price elasticities of demand, that in the most part, the tax would be passed on from manufacturers to farmers. It was assumed that farmers would face 70-85% of the level of the tax and that a smaller amount would subsequently be passed on to consumers in product prices.

ECOTEC estimated the effects of a tax for a range of scenarios as a means of summarising the vast array of possible responses. A tax equivalent to 30% of the value of the product is estimated to increase pesticide prices by 21-26% (estimated as £4-5/kg) and lead to:

- a 3.7-4.4% reduction in use under a relatively inelastic demand; or
- a 9.1-10.7% reduction in use under less inelastic demand.

A tax equivalent to 100% of the value of the product is estimated to increase pesticide prices by 70-85% and lead to:

- a 10.1-11.6% reduction in use under a relatively inelastic demand; or
- a 23.3-26.5% reduction in use under less inelastic demand.

To measure environmental effectiveness of the tax, ECOTEC estimated the total number of months, summed over all catchments, in which an ambient standard is breached. ECOTEC estimates that, if a 20% reduction in pesticide use could be achieved, this would lead to a 32% reduction in exceedances. Of note, they found relatively little difference in effect between a tax set up to vary with the intrinsic hazardousness of the product (ie a banded approach) and one that was simply an across the board tax per kg of product.

ECOTEC went on to consider the implications for water pollution control expenditure, but were unable to quantify this benefit.

3.1.6. Uses of the Revenue

ECOTEC argues that assisting users to minimise pesticide use would increase the effectiveness of the instrument. They argue for the use of some of the tax revenue to fund this. There is a strong theoretical argument against hypothecation of tax revenues in this way. ECOTEC suggests that this example is weak hypothecation, in which revenues and expenditure are not rigidly tied together. On practical grounds this might makes sense, however, it must be remembered that this is an additional cost of the instrument which should not be ignored in analysis.

3.2. Danish Pesticides Tax

The Danish tax on pesticides was implemented on January 1, 1996 with the objective to reduce pesticides use by 10%. It was motivated by the 1986 Pesticides Action Plan which aimed to halve the pesticides use in 1997 compared to 1981-85[33].

The tax is imposed at the level of manufacturing and import. This reduces the costs of control and administration, because the number of companies registered at this level is considerably smaller than at the retail level. Enterprises that produce or import the dutiable pesticides must register with the customs authorities.

A tax increase took effect November 1, 1998. The purpose of the increase was to achieve a further 10% reduction. The tax currently applies to three groups of pesticides and is applied on an ad valorem basis at 33.33% of the retail price exclusive of taxes; for insecticides the tax is 53.85% of the retail price because of the lower price of these products, and in order to achieve a significant demand reduction. Additionally, the tax applies to wood preservatives, where it is set at 3% of the product's gross value[34].

The fact that the tax is based on retail prices requires a strong monitoring of retail prices. This is achieved through a price label system, which indicates a maximum selling price. Pesticides may not be sold at prices that exceed the maximum price, and retailers who sells at a price lower than the maximum price, or who offer a rebate, are not entitled to corresponding tax refunds. Tax liability is proportional to the maximum retail price.

[33] COWI Consulting Engineers and Planners AS (2000) Economic Instruments in Environmental Protection in Denmark. Danish Environmental Protection Agency

[34] Before November 1, 1998, the tax rates for insecticides and fungicides were 35% and 27% respectively, whereas the tax on wood preservatives was the same as now.

Simultaneously with the tax, property taxes were reduced for agricultural properties. The tax and the reduced property taxes are legally and financially independent. A scheme to provide financial support for the development of organic farming was introduced simultaneously with the increased pesticides tax on November 1, 1998.

The tax has led to a significant reduction in treatment frequencies. Treatment frequency was around 2.7 in 1995/1996, and by 1997 it had dropped to 2.45. The Pesticides Action Plan aims to reduce treatment frequency down to a level of 1.34.

The achieved reductions are attributable not only to the tax, but also to other initiatives taken to reduce pesticides use in Denmark. These include stricter environmental standards for the use of pesticides, education schemes for farmers, and supervision of pesticides application equipment.

3.3. UK Landfill Tax

The UK introduced a landfill tax in 1996. When the landfill tax was introduced the rates were based on estimates of the environmental externalities associated with disposing of waste at landfill. There are two tax rates: a standard rate, originally set at £7 per tonne, for "active" wastes; and a lower rate of £2/t for "inactive" wastes. While the lower rate has remained at £2/t since inception, the standard rate was increased to £10/t in 1999. Current policy for the standard rate is based on an escalator announced in 1999, under which there were a series of five annual £1/t increases from April 2000 to April 2004. The standard rate for active wastes rose to £15/t in April 2004. Through introducing the escalator, the landfill tax has become more of an incentive-based or "behavioural" tax, designed to reduce landfill disposal[35].

Annual tax revenue is £502 million (2001/02), net of contributions to the Landfill Tax Credit Scheme, of which 95% is 'active' waste revenue. This is offset by a 0.2% reduction in employer National Insurance Contributions.

Since the introduction of the tax, there has been a 60% reduction in the volumes of 'inactive' waste sent to landfill sites, whilst the volume of 'active' waste sent to landfill has remained broadly unchanged. The latter is explained by the fact that the costs of landfill, including landfill tax, remain low compared to other alternative methods of treatment/disposal. Moreover, landfill disposal costs represent a relatively small proportion of business operating expenses.

[35] Strategy Unit (2002) Waste Not, Want Not. A strategy for tackling the waste problem in England. www.number-10.gov.uk/su/waste/report/downloads/wastenot.pdf

Of those countries which have a landfill tax, the UK currently has the lowest tax rates for active waste (Table 2), apart from France (which has an escalator of 1 Euro per tonne per annum) and Finland which is proposing large increases. As the UK also has relatively low gate fees, the overall cost of landfill remains low compared to other countries.

Table 2: Landfill Tax Rates and Prices

Country	Tax rate (£/tonne)	Landfill prices (£/tonne)
Austria	18 - 54	36 – 82
Belgium	3 – 14	43 – 51
Denmark	28	13 – 21
Finland	9	-
France	4	-
Italy	0.6 – 16	-
Netherlands	8 – 40	47
Sweden	17	-
UK	2 – 13	13 – 23
Czech Republic	11	-
Norway	25 – 50	-
Switzerland	6 - 20	56 - 65

Source: Strategy Unit (2002) Waste Not, Want Not. A strategy for tackling the waste problem in England

3.4. US SO_2 Trading System

A cap and trade system for SO_2 emissions began in the US in 1995 under Title IV of the 1990 Clean Air Act Amendments that established the Acid Rain Program[36]. It is a long-running instrument that illustrates many of the important elements of a tradable permit system.

3.4.1. The cap

The goal of the Acid Rain Program was to reduce SO_2 emissions to approximately 50% of 1980 emission rates, a cut of approximately 10million tons. It sets a total cap on SO_2 emissions from electric power plants nationwide and was introduced in two phases. Phase I (1995–99) limits emissions by approximately 40% and Phase II (2000 forward) by 50%.[37]

[36] For a description, see www.epa.gov/airmarkets

[37] US Environmental Protection Agency (EPA) (2000), 'Analysis of the Acid Deposition and Ozone Control Act (S.172)', prepared for The Senate Committee on Clean Air, Wetlands, Private Property, and Nuclear Safety. US Environmental Protection Agency, Office of Air and Radiation, Clean Air Markets Division.

3.4.2. Sources covered

The SO_2 trading system applies to electricity generating plants only, although other industrial plants can enter on a voluntary basis. Phase I started in 1995 and applied to the largest coal- and oil-fired power plants (greater than 100 MW_e and with emission rates greater than 2.5lbs/mmBtu—ie, 1,075g/GJ). It originally covered 263 units at 110 power plants, but other units opted voluntarily into the Program; at the end of Phase I in 1999, there were a further 135 units in the Program, totalling 398 units in aggregate.
Phase II started in 2000 and extended the coverage to all power plants with capacity above $25MW_e$, and involves over 2,300 units.

3.4.3. Allowance distribution

Allowances under the US SO_2 trading system are allocated in perpetuity. In other words, a plant will be given the same number of allowances every year for ever, or until government policy is changed. This means plants continue to be given allowances on closure, and sales of allowances can be of the annual allowance or the right to a stream of future allowances.

In Phase I, the EPA gave allowances to units on the basis of a formula that used an emission rate of 2.5lbs SO_2/mmBtu of energy input, multiplied by the unit's baseline energy input (the average amount of fossil fuel consumed from 1985 to 1987). There were some additional allowances distributed to certain states reflecting agreements made during the design phase[38]. In Phase II, the EPA allocated allowances to each unit at an emission rate of 1.2lbs/mmBtu (516g/GJ) of heat input, multiplied by the unit's baseline energy input.

In addition to annual allocations, allowances are also available upon application to three EPA reserves.

- During Phase I, units could apply for, and receive, additional allowances by installing flue-gas desulphurisation (FGD), or by reassigning their reduction requirements among other units with FGD.
- A second reserve provides allowances as incentives for units achieving SO_2 emissions reductions through customer-oriented conservation measures or renewable energy generation.
- The third reserve contains allowances set aside for auctions, which are sponsored annually by the EPA.

The annual auction, which distributes approximately 3% of allowances, was very important in the early stages and was responsible for a large proportion of the initial trades. It helped to establish and reveal prices, ensuring price transparency for market players. In the absence of the auction, it is likely that there would have been a relatively

[38] affected units in Illinois, Indiana, and Ohio were allocated a pro-rata share of 200,000 additional allowances each year from 1995 to 1999

long period of bilateral trading with no public information, and that this would have tended to continue the initial perception of high prices, and the greater likelihood of more investment in Flue-gas desulphurisation (FGD). The auction is reducing in importance over time now that a number of brokers have set up, which provide a simple means for exchange and reliable price data.[39]

3.4.4. Treatment of new entrants

There is no special treatment of new entrants. Units that began operating in 1996 or later are not given allowances, but must purchase them, either from the market (eg, from brokers or directly from other generators), or from the EPA auctions and direct sales. Because no new coal plants are expected to be built in the USA, this rule has not been an important issue. For the US NOx emissions trading schemes (see below), where new entrants are an issue, an alternative approach has been adopted, enabling new entrants to obtain allowances.

3.4.5. Voluntary entry

Plants not originally covered, including other electricity generators and industrial plants, can opt into the trading system. In Phase I, this included power plants that would be included under Phase II that could opt in as a substitution or a compensating unit during Phase I. Generators can make emission reductions at the substitution unit rather than the unit included compulsorily (Table 1 units)[40]. In addition, compulsory units that reduce their load factor below their baseline[41] may designate a compensating unit to provide compensating generation.

The Opt-in Program gives units the opportunity to participate voluntarily, even if they are not required to participate in the Acid Rain Program under either phase. These plants are given allowances on the basis of historical fuel input multiplied by an emission rate, in the same way as the compulsory units. It is potentially of interest to plants that can reduce their emissions below historical levels at low cost and sell allowances. By 1999 only 10 units had entered under the Opt-in Program. In total, their emissions were 18% of their allocation of allowances (ie, they could sell or bank 82% of their allocation), compared with an average of 71% (selling or banking 29%) across all sources in the programme. Thus, although the approach to allocation was very generous, there has been little voluntary entry.

---

[39] Ellerman, A.D., Schmalensee, R., Joskow, P.L., Montero, J.P. and Bailey, E.M. (1997), 'Emissions Trading under the US Acid Rain Program: Evaluation of Compliance Costs and Allowance Market Performance', Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology.

[40] The units listed in Table 1 of the Clean Air Act Amendments 1990.

[41] The baseline is defined on the basis of the thermal value of fuel input averaged over 1985–87.

3.4.6. Banking

Allowances can be banked. Where emission levels are lower than the number of allowances distributed, these excess allowances can be kept for use in future years. There are no limits on the extent of banking.

3.4.7. Compliance

At the end of each year, units must hold a quantity of allowances equal to or greater than the amount of SO_2 emitted during that year. There is a 60-day (2-month) grace period at the end of the year during which units must bring themselves into compliance.

If a unit's emissions exceed its allowances, the unit must pay a penalty and surrender the necessary allowances for the following year. The penalty for non-compliance was set initially at $2,000 per excess ton of emissions. This amount has increased each year from 1995 at a rate equal to the consumer price index. It is approximately ten times the cost of allowances.

3.4.8. Monitoring

Emission sources under the trading programme must install continuous emissions monitors (CEMs) to measure stack emissions. All CEM systems must be in continuous operation and must be able to sample, analyse and record data at least every 15 minutes. The EPA describes this as applying a gold standard to back up the paper currency of emissions allowances, with CEMs instilling confidence in the market-based approach.[42]

The regulations use a conservative approach to substitute for missing data, which gives an incentive to keep monitor downtime to a minimum. There are additional regulations referring to performance criteria of monitors.

The CEM rule includes requirements for notification, record-keeping, and reporting. The owner or operator must report the data in a standard electronic format.

[42] http://www.epa.gov/airmarkets/monitoring/factsheet.html

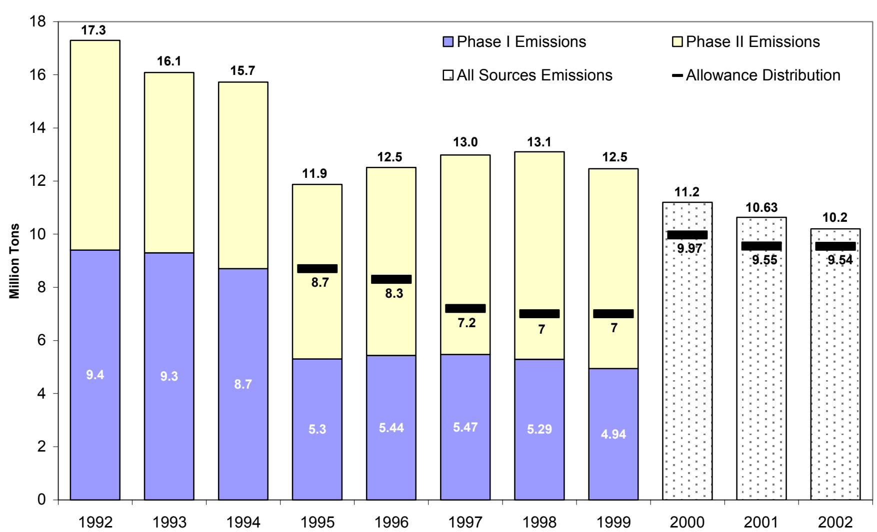
3.4.9. Performance

The trading scheme saw over-compliance throughout Phase I, for two reasons:

- the level of the financial penalty for non-compliance—this provides an incentive for plants to over-comply to some extent, because of the risk of being out of compliance and the high costs of doing so; and
- the reduced costs of over-complying—because allowances can be banked, excess allowances can be retained for use or sale in future time periods. Thus, they retain a value at the end of the compliance period. In the absence of banking, excess allowances would have no value and plants would seek to minimise their holdings of allowances consistent with limiting the risks of non-compliance. Banking therefore increases the probability of compliance because it limits the cost of over-compliance.

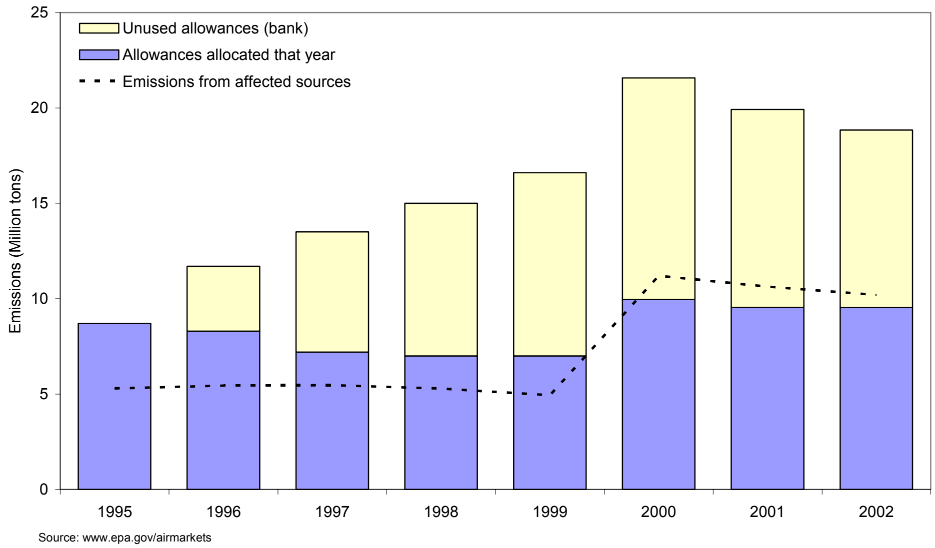
Figure 5 shows the results of the trading programme through to the end of 2002. The solid black lines are the allowances allocated in that year. From 1995 there are compulsory units (Phase I) and others (Phase II) that are included from 2000 onwards and could choose voluntarily to opt in from 1995. From 1995 to 2000, emissions were below allocations; this led to banking of allowances to be used in the future. This has meant from 2000 onwards emissions have been above the annual allocation as the banked allowances are used. Emission rates are well below total allowable emissions, made up of allowance allocation plus the bank (Figure 6).

Figure 5: SO_2 emissions versus allowance allocations



Source: www.epa.gov/airmarkets

Figure 6: Banked Allowances and Allowable Emissions



3.4.10. Allowance prices and price volatility

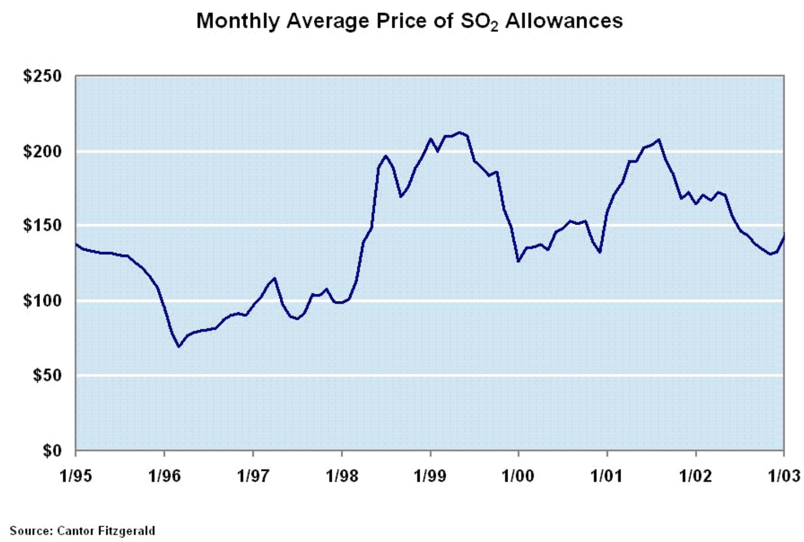
Initially, SO_2 allowance price expectations were estimated at the long-run marginal cost (LRMC) of fitting FGD. This led to what, in retrospect, is seen as FGD over-capacity. FGD costs have fallen over time, and many plants found other lower-cost means for achieving emission reductions, particularly the use of low-sulphur coal.

Prices have not followed a smooth path (see Figure 7). Fluctuations over time have reflected a wide range of factors, including:

- shortfalls in availability of low-sulphur coal;
- fluctuations in electricity demand and dispatch reflecting weather and other demand factors;
- additional policy pressures that have increased the supply, or perceived future supply, of FGD (eg, the EPA's pursuit of companies for non-compliance with New Source Review).[43]

[43] New Source Review requires plants that have undergone significant modification or increases in load factor to install best available control technologies (or achieve the lowest achievable emission rate in areas in non-compliance with national ambient air quality standards). Many plants have avoided this through claims that changes were routine maintenance rather than modifications.

Figure 7: SO_2 allowance prices



3.5. NOx Trading

3.5.1. Ozone Transport Commission

A NOx cap and trade system started in the north-eastern USA in 1999. It was established to assist the states in the Ozone Transport Commission (OTC)[44] to achieve their ambient air quality standards for summertime ozone, and was motivated by the fact that the states in the OTC were unable to meet standards because of transfer of ozone from upwind states. NOx is an important precursor for ozone.

The programme caps summertime NOx emissions at 219,000 tons in 1999 and 143,000 tons in 2003, less than half of the 1990 baseline emission level of 490,000 tons. The main design features of note are:

- it applies to summertime (May–September) emissions only, because of its objectives in controlling ozone, a summertime phenomenon;
- the decision on allocation of allowances is made at the state rather than federal level. States have used a variety of grandfathering (gifts on the basis of historical activity) formulae;

[44] The OTC comprises the states of Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Pennsylvania, Maryland, Delaware, the northern counties of Virginia, and the District of Columbia.

- sources with large NOx emissions must monitor using CEM systems, but those with lower NOx emissions may use simpler estimation methods employing predictive emission monitors. Regardless of the method used to determine emissions, the data supporting these determinations must be reported electronically to the EPA;
- there is scope for voluntary opt-in by plants not otherwise required to comply.

There is no auction of allowances because this is regarded as unnecessary. Unlike when the SO_2 system was established, when the NOx cap and trade system began there was already a well-functioning system of brokerage services with transparent pricing of allowances. It was believed, correctly as it turned out, that such price transparency would rapidly establish for NOx allowances. This was assisted by two years' allowance trading prior to the start of the first budget or compliance period.

Limited banking of permits is allowed. Sources can bank indefinitely and can use allowances from the current or previous years. However, if the bank reaches 10% or more of the total annual budget, a system of flow control is used to increase the costs of using the bank. After the bank has breached the 10% trigger, the ratio of banked allowances to the allowed bank size is used to devalue all banked allowances. This ratio is used to define how many allowances in each unit's bank can be surrendered 1:1 against emissions, with the others required to be surrendered at a 2:1 ratio. For example, if the total cap is 100,000 tonnes (10% = 10,000) and there are 20,000 tonnes in the bank then the flow control ratio is 10,000:20,000 = 0.5. The number of banked allowances in each company's account is multiplied by this ratio to determine how many allowances can be traded at a 1:1 ratio. The implications of flow control are that allowances of different vintages have different prices.

Firms have two months to complete final trades at the end of each budget period to ensure compliance. Because the NOx system is administered by states rather than the federal government, compliance penalties have been set in terms of offset requirements. If a company is out of compliance, it must surrender three allowances in the next compliance period for every ton of emissions in excess of allowances that have been surrendered in this period. This does not affect the liability of owners or operators under state law, and ultimately the sanction of removal of the right to operate.

3.5.2. Ozone Transport Assessment Group

The Ozone Transport Assessment Group (OTAG)[45] proposals extended the OTC trading approach to a much larger number of eastern states. The programme starts in 2004, although allowances were distributed from 2001 to allow for trading and price discovery prior to the commencement of the first compliance period. A cap is set on NOx emissions on the basis of an average emissions rate of 0.15lbs/mmBtu, equivalent to less than 200mg/Nm^3.

The main departure from other trading systems is the use of a rate-based approach with updating to the distribution of allowances. The total emissions cap is divided by the activity data from plants included in the programme to yield an emissions rate in lbs/mmBtu (the energy value of the fuel input) (input-based) or lbs/MWh generated (output-based). This rate is the basis for annual allocations to any plant that is included in the programme, both existing plants and new entrants. The EPA has developed a model rule as guidance for the states using an input-based approach; three states (Connecticut, Massachusetts and New Jersey) are planning to use an output-based approach. Significantly, output-based approaches would include allocations to plants with no emissions (ie, renewables and nuclear facilities).

In the EPA's model rule, the rate used for allocation is updated every five years, but individual states are considering options varying from annual to ten-year updating. Under a five-year approach, activity data are collected for each unit (the EPA proposed an average of the activity data from the two highest of four previous years). The cap is divided by the aggregate activity data to yield a rate that is used as the basis for the allocation to individual units. The allocations are then reduced proportionally so that a total of 95% of allowances are allocated initially, the remainder being retained for new entrants. For each unit, its activity data in the base year(s) times the agreed rate are used to calculate a number of allowances. This same number of allowances is given to that unit each year for the next five years. For example, under output-based allocation, if the rate for allocation was 1.5kg/MWh, a unit with an average output of 2,200 GWh in the base years would be given 3,300 1-tonne allowances in each of the next five years. For subsequent control periods, a new set of average activity data is used as the basis for allocation. The suggested approach in the US EPA's model rule is noted in Table 3.

[45] A partnership between the US EPA, the Environmental Council of the States and various industry and environmental groups

Table 3: Years used for activity data in US NOx trading

Control period	Years used for activity data
2003–07	Two highest years over 1995–98
2008–12	Average over 2002–04
2013–18	Average over 2005–09
Subsequent five-year periods	Average in 4–8 years prior to first year of control period

Source: US EPA.

The rate-based approach is similar to historical grandfathering, but it results in plants with emission rates below the rate-based standard being given more allowances than they require, which provides them with an additional source of revenue.

New entrants are allocated allowances from the 5% set-aside, using the emissions rate and projected activity rates. At the end of the year, the allocation is adjusted to take account of the actual activity data and allowances removed (or extras given) as appropriate. If the allowance pool for new entrants is exhausted then new entrants will be required to purchase allowances in the market.

The new-entrants issue has been the main reason for the shift towards updating rate-based allocation methods. It is more important for NOx emissions than for SO_2 because new entrants, expected to be largely combined cycle gas turbines (CCGTs) and open-cycle gas turbines (OCGTs), will produce NOx emissions. In contrast, these plants have, effectively, zero SO_2 emissions.

There is some variation among states in their plans for this new NOx trading programme, including:

- the size of plants included in the system;
- the use of early-reduction credits—ie, allowing plants that make early reductions before the start of the first compliance period to earn additional allowances for use in the compliance period. (This is a source of current conflict with the EPA and some states);
- the frequency of updating.

The compliance approach is the same as used for the OTC NOx programme—ie, using an offset requirement.

The system of flow control to limit banking is also the same as that used in the OTC system.

3.5.3. RECLAIM

The Regional Clean Air Incentives Market (RECLAIM) is a cap and trade programme that has been adopted for the Los Angeles air basin.[46]

A large number of concerns were raised during its design regarding the use of a grandfathering approach to allowance distribution, including:

- unfairness to plants that had made prior reductions;
- the lack of representativeness of the year chosen for setting baselines;
- unfairness to firms that were growing.

In response to these concerns, the final approach chosen involved three sets of allocations for each facility, in 1994, 2000 and 2003. The basic ton allocation used an emission rate times the maximum activity rate during 1989–92. In addition, for 1994, there are additional allowances based on offsets purchased by plants under New Source Review.[47] Plants were also given non-tradable allowances for the first three years of the programme if their 1987, 1988 or 1993 emissions were greater than their initial allowances.

The highly complex set of rules for allocation was developed in order to meet equity and fairness concerns:

- allowing plants to use peak activity over a four-year period avoided the use of an unrepresentative year;
- basing allocations on emission rates under independent rules provided different rates of reduction for different facilities based on their prior controls;
- adding offsets gave credits for previous emission reductions;
- providing three years of non-tradable allowances if current emissions are greater than 1994 allocations avoided facilities 'starting in the hole'.

The RECLAIM programme also has features that relate to environmental concerns.[48]

- It is divided into coastal (upwind) and inland (downwind) zones. Inland sources can obtain allowances from either zone, while coastal sources can only obtain and use allowances that originated in the coastal zone.

[46] Harrison, D. (1999), 'Turning Theory into Practice for Emissions Trading in the Los Angeles Air Basin', in S. Sorrell and J. Skea (eds.), *Pollution for Sale: Emissions Trading and Joint Implementation*, Edward Elgar, pp. 63–79.

[47] In areas out of attainment with national ambient air quality standards, new plants that wish to establish, or old plants undergoing substantial modification, are subject to New Source Review. They need both to introduce stringent emission control equipment and to purchase offsets in the form of reduced emissions at other plants.

[48] Harrison, D. (1999), op cit.

- Banking is not permitted in the programme but facilities are divided into two compliance cycles: one from January 1st to December 31st and the other from July 1st to June 30th. Transactions can be conducted with facilities in either cycle. The dual-compliance-period approach was adopted to avoid the possibility of allowance price volatility, particularly at the end of a compliance period, depending on whether there was an allowance shortage or excess.
- There is an annual budget, although emissions and allowances have to be reconciled on a quarterly basis, including a one-month reconciliation period.

There is no provision for new entrants.

3.6. Mercury Trading

Mercury is a pollutant emitted from a number of industrial sources, including coal-fired electricity generation. The success of cap and trade schemes for SO_2 and NOx applied to electricity generators has seen interest in extending the approach to other pollutants, notably mercury and CO_2.

In December 2003, the US EPA signed the Utility Mercury Reductions proposal to substantially cut mercury emissions from coal-fired power plants—by nearly 70 percent when fully implemented[49]. The proposed Utility Mercury Reductions rule[50] would permanently cap emissions from coal-fired power plants and provide companies with flexibility to achieve early reductions of mercury. The EPA proposed two alternatives for controlling mercury.

- One approach would require power plants to install controls known as maximum achievable control technology (MACT) under section 112 of the Clean Air Act. If implemented, this proposal would reduce nationwide mercury by 14 tons or about 30 percent by early 2008.
- A second approach would create a cap and trade program that, if implemented, would reduce nationwide utility emissions of mercury in two phases. When fully implemented mercury emissions would be reduced by 33 tons (nearly 70 percent).

In February 2004, EPA signed a proposed cap-and trade rule that supplements its 2003 proposal. States may choose to adopt the cap-and-trade program to achieve and maintain the necessary emission budgets.

[49] www.epa.gov/mercury/
[50] US EPA (2004) Proposed National Emission Standards for Hazardous Air Pollutants; and, in the Alternative, Proposed Standards of Performance for New and Existing Stationary Sources: Electric Utility Steam Generating Units. Federal Register / Vol. 69, No. 20

The proposed cap-and-trade system would be based on the Acid Rain Program. The EPA would allocate to each state specified amounts of emission “allowances” for mercury. The states would subsequently allocate those allowances to utilities, which would trade them.

3.7. Lead Phase-Down

3.7.1. Description

A different use of tradable permits was in the US phasedown of lead in gasoline. It was accomplished partly through a tradable permit system among refineries[51].

The EPA established performance standards for individual refineries and compliance deadlines extending over several years, starting in 1979.

- Large refiners[52] were to produce a quarterly average of no more than 0.8grams per gallon (gpg) for the first year and 0.5gpg in the next two years.
- Small refiners faced a scale of five different standards, ranging from 2.65gpg for the smallest to 0.8gpg for the largest.

The averaging method encouraged increased production of unleaded gasoline rather than, necessarily, reducing the lead content of leaded gasoline. By the early 1980s lead levels in gasoline had reduced by about 80%; in 1982, revised regulations narrowed the definition of a small refinery and recalculated lead limits as an average of lead in leaded gasoline only. The new standard was 1.1 grams per leaded gallon (gplg); the EPA further proposed a reduction to 0.1gplg by 1986.

Until 1982 the regulations had been prescriptive—they included technology standards and individually binding refinery performance standards for lead content. This was causing difficulties and significant costs, especially for small refineries. At the same time, some of the larger refineries were able to make significant and low cost reductions.

[51] Newell RG and Rogers K (2003) The Market-based Lead Phasedown. Resources for the Future Discussion Paper 03-37.

[52] Production capacity of over 50,000 barrels per day (bpd) and/or those owned or controlled by a refiner with total capacity of over 137,500 bpd.

The tradable permit system allowed inter-refinery lead averaging from late 1982—some refiners could produce higher concentrations than others, while the average across all refineries met the EPA's standard. The EPA allowed the market to develop on its own. Refineries were required to provide the following information[53]:

- Total grams of lead that the reporting refinery allocated (sold) to other refineries, and the names and addresses of such other refineries (A);
- Total grams of lead that the reporting refinery was allocated (bought) from other refineries, and the names and addresses of such other refineries (B);
- Total grams of lead "constructively used" by reporting refinery (C=actual lead usage –A+B);
- "Constructive average" lead content of each gallon of leaded gasoline produced by the reporting refinery during the compliance period (C/total gallons produced); and
- If compliance was demonstrated through averaging with more than one other refiner, supporting documentation showing that all parties agreed to the constructive allocation.

Banking was another flexible component of the regulations. It allowed refiners with low concentrations in 1985 (less than 0.5gplg but more than 0.1gplg) to use this same amount between 1985 and 1988.

The lead trading program finished in 1988; the EPA was concerned about introducing incentives to reduce lead concentrations below 0.1gplg, which it believed was required for engines designed to use leaded gasoline to operate properly.

3.7.2. Impacts

The phasedown program was estimated to have been responsible for 36% of the total gasoline lead reduction up to 1988, relative to that resulting from market shifts, and the tradable permit scheme to have saved hundreds of millions of dollars[54]. The scheme was regarded as being highly efficient, with few barriers to trade

3.8. VOCs Trading

The Illinois Emissions Reduction Market System (ERMS), established in 2000, is the first cap and trade program for volatile organic compounds (VOCs)[55]. The program was developed as a means for Illinois to meet the federal one-hour standard for ground-level ozone, with which the State was in non-attainment.

---

[53] Newell and Rogers (op cit)
[54] Hahn, Robert W., and Gordon L. Hester. 1989. Marketable Permits: Lessons for Theory and Practice. *Ecology Law Quarterly* 16: 380–391 In: Newell and Rogers (op cit)
[55] http://www.epa.state.il.us/environmental-progress/v28/n4/erms.html

The ERMS is a cap-and-trade system in which participating sources must hold "trading units" for their actual volatile organic matter (VOM) emissions. The ERMS includes major point sources in the Chicago non-attainment area that are required to obtain a Clean Air Act Permit Program (CAAPP) permits and with baseline VOM emissions of 10 tons/season.

Each year, participating sources are issued trading units based on historical or baseline emissions[56], lowered proportionally to provide the emission reduction required from stationary sources—the allotment is generally set at the source's baseline emissions, less 12%. At the end of each ozone season (May 1 through September 30), sources must hold sufficient trading units to cover their actual VOM emissions during the season.

Surplus allotment trading units (ATUs) may be sold to other sources or banked for use in the following season.

Because the requirement to hold ATUs applies only to the ozone season, when ambient ozone air quality is a problem, methods of complying with the ERMS include shifting production outside the ozone season, or scheduling maintenance outages during the ozone season, in addition to emission reduction methods.

The ERMS allows sources to go "outside" the ERMS to obtain ATUs. A source may have ATUs issued to it based on VOM emission reductions at a permitted source that is not required to participate in ERMS. ATUs may also be issued for reductions from mobile (vehicle) or area (non-point) sources.

---

[56] The "baseline emissions" are a participating source's average VOM emissions during two ozone seasons based on historical operation as adjusted for voluntary overcompliance or non-compliance. A source may select the two seasons to be used from the 1994, 1995, and 1996 seasons, or if these are not representative of typical operations, it may propose alternative emission data from 1990-1993 and 1997 seasons.
Adjustments must be made to account for emission units that were not in compliance with the applicable regulations effective in 1996.
Additionally, sources may claim adjustments to their baseline emissions to take credit for voluntary overcompliance. This allows sources to take credit for measures taken after 1990 that result in a reduction of VOM emissions beyond that required by applicable 1996 regulations. For example, if a source had an emission unit that was allowed to emit 20 tons/season of VOM emissions based on the applicable 1996 rule, but the source added control equipment in 1992 that resulted in the unit only having VOM emissions of 15 tons/season, the source can add the additional 5 tons/season of VOM emissions above their actual emissions into the baseline. As a result, a source that added control after 1990 above and beyond what was required by applicable 1996 regulations will not be "penalized" for their efforts towards environmental control as compared to a source that takes such action after the ERMS is in place. The adjustment for voluntary overcompliance assures that there will be surplus ATUs upon startup of the ERMS in 2000.

The EPA has established an Alternative Compliance Market Account (ACMA). This is a reserve of ATUs managed by the Agency which starts with ATUs equivalent to 1% of the entire program's allotment of ATUs to participating sources. If no ATUs available through the market, this becomes a safety net.

Sources that do not hold sufficient ATUs are required to make monetary compensation. The monetary compensation for emission excursions relates to the ACMA. A source must buy ATUs from ACMA at the rate of 1.2 times the emission excursion for a first time excursion, and 1.5 times if it has two consecutive emission excursions. ATUs from ACMA are priced at the lesser of $1,000 or 2 times the average market price of an ATU.

Participating sources in the ERMS program reduced VOC emissions by 38% below their allotted emissions in 2000, 47% in 2001 and 48% in 2002.

3.9. UK Renewables Obligation Certificates

The UK Renewables Obligation requires electricity suppliers to supply 10% of demand from approved renewable sources in 2010, with lower annual targets in years prior to this date.[57] Proof of compliance is via holdings of Renewable Obligation Certificates (ROCs), which are allocated to renewable generators for every MWh generated from an approved source.

There are a number of elements of this trading programme that are different from most:

- it introduces a price cap on ROCs. Suppliers have the option of purchasing ROCs or paying a buy-out price of £30/MWh. This is introduced because the government is not willing to meet the renewables target at any cost;
- the revenues from the buy-out mechanism are returned to suppliers in proportion to their holdings of ROCs. This means that, while there is an overall under-supply, purchases of ROCs are worth more than the £30/MWh avoided cost of the buy-out because the value includes, in addition, the returned revenue from the buy-out[58]. ROCs are currently trading at just under £50/MWh[59]

The use of a buy-out is equivalent to a hybrid tradable permit/emissions tax system. Normally, in contrast to taxes that provide price certainty, but emissions uncertainty, tradable permits provide emissions certainty but price uncertainty. Such a hybrid, in which firms have an option of purchasing allowances or paying a charge, provides certainty in

[57] Department of Trade and Industry (2001), 'New and Renewable Energy: Prospects for the 21st Century. The Renewables Obligation Statutory Consultation.'
[58] At the extreme, if there was only one ROC, the holder would receive all the buy-out revenue, equal to £30/MWh times 10% of total UK electricity supply (less one MWh).
[59] www.platts.com/Electric%20Power/Resources/News%20Features/roc/index.xml

achieving an emissions objective only if the cost is below a certain amount. Above this, emissions are allowed to increase.

3.10. UK Packaging Recovery Notes

The UK's Producer Responsibility Obligations (Packaging Waste) Regulations require specified businesses to recover and recycle specified tonnages of packaging waste. To demonstrate their compliance with these obligations they must hold evidence in the form of Packaging Recovery Notes (PRNs).

Obligated parties must discharge their responsibilities using registered recyclers who can produce PRNs. Although it was not the original intention of the regulations, PRNs have become tradable commodities. Rather than becoming directly involved in recycling or joining a compliance scheme, PRNs can now be purchased online[60].

The price of PRNs reflects the costs of recycling, but because the system has no banking and borrowing, it is subject to end-of-year price spikes.

3.11. Lessons Learned

The theoretically ideal instrument to address environmental externalities—a charge equal to marginal damage—has been very difficult to achieve in practice and there are few examples of attempts to do so. Where this approach has been used, modifications have generally been introduced to ensure achievement of targeted environmental improvements, eg the escalator in the UK landfill tax and targeted use of revenues raised to achieve additional environmental improvement.

This experience raises questions relevant to the application of economic instrument under the HSNO Act—to what extent is the instrument to achieve a balance between costs and benefits or to achieve targeted improvements? This issue has wider relevance under the Act than just the use of economic instruments, ie the way in which the requirement to take account of the benefits and costs of using a particular hazardous substance or new organism modifies the purpose of the Act, which is to protect the environment, and the health and safety of people and communities.

Where an environmental improvement is targeted, tradable permits can be a better option. In the US, in particular, there is considerable enthusiasm for wider application of this instrument, because of the success of the acid rain programme and because it can achieve low cost improvement while allowing government to avoid raising taxes.

[60] http://www.t2e.co.uk/

4. Potential Use of Environmental User Charges Under HSNO

There are a number of potential uses of environmental user charges under the HSNO Act. This section addresses the options. In doing so it makes reference to a number of examples of the use of environmental user charges in Section 3.

The way in which the potential performance of these economic instruments is assessed is with respect to the question posed in the HSNO Act: will they be as cost-effective as alternative controls? We have also assessed them in terms of the overall efficiency improvement, ie will they result in a greater aggregate of consumer and producer surplus?

We start from the theoretically optimum instrument. This is a tax equal to the marginal level of damage. In the absence of any current set of regulatory controls, marginal damage would be estimated using the probability of a hazardous event, or of damage associated with a new organism, times the magnitude of the risk event.

Current regulatory controls result in management practices that reduce risks. However, they do not eliminate risk. The efficient instrument would be charged on the level of marginal residual risk. We start by examining whether this is definable.

Secondly, we examine whether a charge might be defined that could replace some or all of the existing set of controls. This might function if the level of risk reduction could be defined as associated with individual components of the current regulatory system.

Thirdly, we examine what kind of charge might be used to have an incentive effect to reduce risk levels or to reduce use of hazardous substances.

Lastly we examine the potential for environmental charges as revenue raising techniques.

Before addressing these options in detail we examine:

- the current regulatory system; and
- the nature of current markets for hazardous substances.

4.1. Regulatory Controls

Under the HSNO Act, existing controls conform to the basic pattern of:

- Risk identification—what sort of risks to people or the environment are associated with this substance or organism?
- Risk characterisation—how great is the risk?
- Risk treatment

The classification systems for the HSNO hazardous properties are set out in Schedules 1 to 6 of the *Hazardous Substances (Classification) Regulations 2001*. The classification systems comprise numbered classes (eg Class 6) indicating the intrinsic hazardous property, numbered subclasses (eg subclass 6.1) indicating the type of hazard, and lettered categories (eg Category A) indicating the degree of hazard.

On the basis of the risk characterisation or classification, a set of controls are prescribed. The prescribed controls are defined in regulations and cover (Section 76):

(a) requirements for packages or containers for hazardous substances;
(b) requirements for specifying the identification, labelling, or advertising of hazardous substances:
(c) requirements for disposal of hazardous substances:
(d) requirements to manage any emergency involving a hazardous substance:
(e) systems for tracking hazardous substances which may include requirements that—
  (i) the whereabouts of the substance be recorded at all times or from time to time:
  (ii) the quantity of the substance be recorded:
  (iii) a person be identified as being in charge of the substance:
  (iv) any person handling the substance holds prescribed qualifications:
(f) qualifications, including requirements that a person be a member of any specified professional body or organisation, for any person handling a hazardous substance.

There is currently a transition phase as substances controlled under previous legislation are moved over to the new set of controls. In the meantime, new applications are subject to the new controls. This provides poor incentives as the new controls are more stringent, and more costly in many instances, whereas many of the new substances are less hazardous than existing ones.

Economic Effects

The control regime has some impact on variable costs, eg every product must have a label, but much of the cost is a once-off cost of authorisation. As such, regulation increases average costs and the costs of entry but may not alter significantly the marginal costs of production that determine price in competitive markets. Thus the regulatory approach reduces risk but, once the authorisation hurdle is overcome, may provide little incentive for reduction in use rates. This provisional conclusion, however, depends on the competitiveness of the individual markets and elasticities of supply and demand. Markets are discussed in Section 4.2 below.

4.2. Markets for Hazardous Substances

The nature of the markets is crucial to understanding the impacts of economic instruments. However, there are tens of thousands of hazardous substances; they include commodity chemicals with numerous potential users and specialist products imported for a single process or plant.

It is beyond the scope of this study and analysis to review the nature of all of these markets. Rather we discuss them in more generic terms. The key elements are whether there are market participants—buyers or sellers—that can exercise market power and thus determine price away from the competitive equilibrium. Where the market is not competitive, economic instruments will not have the same impact on price and will not result in the same incentive effects.

Chemical markets are often separated into:

- generics;
- products which are off-patent and sold by several companies; and
- those under patent for which the seller can have 100% market share and can exercise greater control over price.

However, it is not clear that these necessarily represent separate markets. There may be considerable potential substitution between different types of product within any product category, eg between generics and in-patent products. Thus despite the significant market share of an in-patent product, this may not necessarily allow the firm to introduce a small but significant non-transitory increase in price (SSNIP), the test used to define a separate market.

Were they separate markets, generics would be more likely to show characteristics of a competitive market in which prices were reflective of marginal costs and in which the environmental charge would be passed on. It is not clear what proportion of the market for substances covered by the Act would fit the generics category, although petroleum products—petrol and diesel—is a very large sale item that does.

The agricultural chemicals market is dominated by large multinational companies, often with large R&D spend, that bring new products to the market[61]. In contrast, generics are more typically sold by small local companies; by their nature these products will tend to be older and already being used. They will thus also be those that are in transition to the new HSNO control regime.

A UK study estimated that generic products there comprise only approximately 25% of the pesticides market[62]. In contrast NZIER suggests that the bulk of agricultural chemicals (animal remedy products and pesticides) in the NZ market are generics[63].

If there are monopoly suppliers, ie one or very few imports or producers, but numerous users, prices are likely to be above competitive levels. This is likely to be the case, to a greater or lesser extent, for the non-generics commodities. As discussed in Section 2.3, prices will tend to be set at a level that maximises producer surplus (profit).

For the purposes of this study, we note that markets are likely to vary from:

- generic markets with characteristics similar to the competitive ideal; to
- those with monopoly characteristics in which producers and/or imports have considerable control over price.

There may also be markets in which there are very few buyers in which these firms exercise considerable market power, depending on the number of sellers.

Thus there will be no uniform effect of an environmental charge or tradable permit system. It is beyond the scope of this study to analyse the current market structures for all hazardous substances.

4.3. Charge on Residual Risk

Current regulatory controls reduce the risks associated with hazardous substances (and new organisms) but do not eliminate them. A charge might be introduced to target this residual risk. Below we assess whether residual risk is measurable and if a charge could be applied.

As noted above, a charge equal to marginal damage is an economically efficient way to address an environmental problem. But the important question here is whether it

---

[61] New Zealand Institute of Economic Research (2001) Cost of Compliance. Report to the Ministry of Agriculture and Forestry.
[62] ECOTEC (1999) Design of a Tax or Charge Scheme for Pesticides. Department for Environment, Food & Rural Affairs
[63] New Zealand Institute of Economic Research (2001) Cost of Compliance. Report to the Ministry of Agriculture and Forestry.

represents an efficiency improvement if added to an existing regulatory control? In other words, do the environmental benefits justify the additional cost burden?

There are two relevant aspects of efficiency.

- Productive efficiency refers to the efficient use of existing resources. It is enhanced if a charge is levied equal to marginal damage. Marginal costs would then equal the full social costs of production; this will be reflected in prices and can ensure efficient levels of production and consumption.
- Allocative efficiency refers to efficient levels of allocation of resources to an industry or product. This may not be enhanced. The total costs of introducing a new substance include the costs of the existing regulations and the costs of the charge. If total costs are higher than damage would have been in the absence of any regulation, there will be under-investment in new products[64]. This is an issue if the costs of existing regulations are greater than the benefits—that issue is not the subject of this study and has not been addressed in detail by other studies[65].

The other issue of note is the problem of the current regulatory system that applies to new products and is only being transferred slowly to existing products. Here the marginal effects of a new product may be net positive, ie they cause environmental damage but may displace an existing product that causes greater damage. Charging the new product while not charging the existing one would not be an efficient outcome under either definition. If environmental user charges are to be introduced they need to apply both to existing and new products.

4.3.1. Can Risk Levels be Measured?

Measuring risk levels associated with individual substances would require either:

- evaluation at the point that the substance enters the environment, or affects human health—this is so highly disaggregated as to be impractical; or, more likely
- an assessment of the intrinsic riskiness of an individual substance, adjusted for the estimated effectiveness of the current set of regulatory controls.

Current controls have been put in place as best practice measures to reduce levels of risk. There is no quantified pre- and post-control measure of risk levels. An ex-ante analysis of

[64] As New Zealand is largely an importer of products developed in other countries, investment in this sense refers to the process of introducing a new product to the market through the approval system.

[65] The current set of controls under HSNO have been criticised as being costly and time consuming (New Zealand Institute of Economic Research (2003) HSNO Act: Impact on costs and innovation. Report to Ministry of Economic Development) but there has been no detailed assessment of costs and benefits. Tonkin & Taylor's cost benefit work is not sufficiently detailed to allow any real comparison (Tonkin & Taylor (1998) Benefits and Costs of Proposed HSNO Regulations. Report to Minstry for the Environment)

the benefits and costs of the proposed HSNO regulations[66], for example, provides some estimates of costs and benefits for a number of example products but these are largely based on assumptions made for illustrative purposes rather than any empirical evidence.

Defining Initial Levels of Risk

Currently risk assessment is used to estimate whether a substance is classified as hazardous or not[67]. Thresholds define the amount or concentration of a substance that is likely to cause an adverse effect on people or the environment. It is used as a trigger level for an effect that may require controls on the substance to meet the purpose of the HSNO Act. The threshold effect level is the minimum point on the classification ladder.

For use in a charge system, this set of classifications, and specifically the lettered categories indicating the degree of hazard, would need to be converted into some classification of monetary damage. Currently there is no such basis for defining damage within any category, eg substances that are corrosive to dermal tissues, let alone between categories. Taking the dermal tissue example, the classification of degree of hazard takes account of the length of time of exposure and the time before damage occurs. To define monetary damage would require assessment of, for example, the costs of hospitalisation or estimates of willingness to pay of potential sufferers to avoid this damage. Such calculations could be made, but there is a substantial amount of work that would be required to define these costs.

Even were this undertaken, the assessment would require an estimate of the likelihood (probability) of an event that would cause damage, eg a substance may cause dermal tissue damage after short exposure, but what is the probability that this will occur in the absence of (or with) regulatory controls? Such an assessment also has not been made.

Effects of Current Controls

The effectiveness of existing controls in reducing levels of potential risk below the initial level would need to be defined using one of the following approaches:

- expert opinion on the effectiveness of the current set of controls—this would require the identification of neutral experts as there could be strong incentives to influence the outcome of the assessment given the financial implications. It may also require some system of at least spot checks of opinion against real-world risk reduction;
- empirical analysis, eg of the number of reported events such as hospitalisations or industrial accidents associated with hazardous substances; or
- a more arbitrary approach that might, for example, assume that the current set of controls reduce damage levels by 70% below the intrinsic level of risk.

[66] Tonkin & Taylor (1998) Benefits and Costs of Proposed HSNO Regulations. Report to Minstry for the Environment.
[67] ERMA (2001) User Guide to the HSNO Thresholds and Classifications of Hazardous Substances

In general this approach requires considerable additional data to what is currently available to policy makers. This applies both to assessing the intrinsic or initial level of risk or potential damage associated with an individual substance, and the effectiveness of the control regime.

4.4. Charge to Replace Existing Controls

We can imagine a substance for which, in the absence of any controls, the risks to people and the environment is 100 units of risk. As a result of the set of controls, this risk is reduced to 20, but if only part of the control regime was employed, eg the labelling requirements but not the set of handling/protective clothing requirements, the level of risk was only reduced to 50. Under such an approach a charge might vary, depending on which aspects of the control regime were adopted.

However, such a system requires the same set of data as discussed above for a charge on residual risk, ie it requires firstly that we define the level of uncontrolled risk (probability of damage and damage level) and the effectiveness of the control regime. Further it requires that the effectiveness of individual aspects of the control regime can be defined with respect to these parameters.

Currently the data do not exist to define such a charge system. The implications of more arbitrary approaches to setting charge levels are addressed under Incentive Charges below.

4.5. An Incentive Charge

4.5.1. Practicality

In contrast to the discussion above of more idealised charge systems based on measurements of residual damage, some of the examples discussed in Section 3 above use simpler approaches. In the case of the Danish pesticides tax, a simple ad valorem system was used, with the tax level defined with reference to the absolute value of the product, ie the lower cost products had a larger percentage charge applied.

The analysis of a potential pesticides charge in the UK defined a system that used an evaluation of the hazardous nature of the pesticide substance.
The analysis suggested that a banded system that varied the charge according to risk levels, rather than a simple charge per kg, regardless of the hazardous nature, made very little difference to environmental effect. The analysis gives little information on the modelling approach used to obtain this result. But the conclusions are consistent with the Danish experience, ie that a simple charge can provide incentives for reduced use.

There are different ways in which a charge might be applied, most likely either on a per kg basis or as an ad valorem charge.

The efficiency effects of an incentive charge are uncertain. In the absence of a monetary valuation of damage costs, an arbitrary choice of charge level might either be too high or too low, relative to social costs. Simplifications, including using rates similar to those applied in other countries, eg using the Danish tax rate, provide no guarantee either way. An analysis of the impacts of a 120% tax on pesticides in Denmark[68], suggested that pesticide use would reduce by 40% and agricultural production by 0.5-1.5%; however, we have no information on whether the benefits justify these costs.

Also, although an arbitrary charge can change prices of hazardous substances relative to other goods, it can also change the relative prices of individual hazardous substances. Where these are market substitutes this might have perverse incentives if the approach to charging does not reflect relative damage levels. Where the price elasticity of substitution is greater (more elastic) than the price elasticity of overall demand, the environmental effect may be negative.

Modelling work could be undertaken in New Zealand to estimate the likely effects of a charge. This would need to be done at the level of individual substances, and to include all substances that were substitutes.

An alternative approach is trial and error. This is essentially what the UK Landfill Tax has transformed into. An initial charge level set using an analysis of the damage costs has been increased, taking account of the effectiveness of the tax in shifting waste management practices. Tax levels have increased to achieve greater levels of reuse and recycling. Such an approach might be used with hazardous substances also, starting low (defined eg with reference to current price levels of products) and changing over time following monitoring of responses.

The current classification system allows for some degree of differentiation in setting charge levels. Despite the modelling results of the UK pesticides study, which may be a product of the particular model employed, it is difficult to see why a differentiated charge, with higher levels of charge for products measured to be more hazardous, would not have a greater positive environmental incentive effect.

[68] Jensen, Thomas and Poul Erik Stryg (1996) The Regional Economic Consequences of Environmental Taxes in Agriculture. SØM publication no. 12, AKF, Institute of Local Government Studies

4.5.2. Effects

Incentive charges based on modelling of the expected effects can ensure that there is some certainty of effect and some certainty to business of cost levels. In addition, modelling can enable assessment of the likely impacts on individual industries and the economy as a whole.

In contrast, a trial and error approach to setting an environmental charge can introduce considerable uncertainty for business, because the way in which price might change over time cannot be easily predicted.

4.6. Revenue Raising

This section examines the current schedule of fees and charges as a way to raise revenue. It also examines how a system of environmental user charges might be used as an alternative approach.

4.6.1. Current Fees and Charges

Currently ERMA's revenues include core Crown funding, interest and income from fees and charges. Fees and charges constitute approximately 10% of total revenues.

Charges are set under Section 21 of the HSNO Act and are based on ERMA staff time involved in "exercising or performing any function, power, or duty" under the Act so as to recover the "actual and reasonable costs incurred". A scale of charges is published which include hourly rates for ERMA staff, and for external advisers and some other costs to be charged at cost. This means that the costs are up-front and, to an extent, uncertain to applicants at the outset.

NZIER contrasts this approach to the Australian regulatory system in which there is more cost certainty and the system involves a two-part tariff including an up-front fixed cost and a variable cost[69]. NZIER suggests that the current system, by charging only new entrants, indicates that there is no public benefit from the introduction of new products. This is more a problem of an absence of a charge or control on existing products, rather than incorrect passing on of costs that are faced by government as a result of having to decide on the merits of new products. But the charging system certainly has the potential to have perverse effects. As noted above, the marginal impact of introducing new products may be positive. Charging only new products would therefore introduce completely wrong incentives.

[69] New Zealand Institute of Economic Research (2001) Cost of Compliance. Report to the Ministry of Agriculture and Forestry

Introducing charges as an up-front fee makes sense if the current regulatory system is an efficient way to tackle hazardous substances; the charges would reflect the structure of existing costs. But this "if" has not been tested.

4.6.2. Environmental User Charges for Revenue Raising

Charges per unit of product could be used to raise revenue for ERMA.

As noted in Section 4.1 above, if this is a charge based on unit damage then it would be an efficient tax and a good way to raise revenue. In addition, the current system of charging for applications to introduce a new substance, levies the new applicant for the costs of risk assessment prior to approval. Subsequent introducers of the same substance do not pay these costs. These costs can be significant, and the system represents an entry barrier that can have perverse effects—disincentivising the introduction of new products which may be environmentally beneficial.

Shifting revenue raising to an environmental charge on all products within specified categories, ie existing and new products, would be a less distortionary tool than current revenue raising mechanisms.

As noted above, currently there are no means of identifying unit damage costs to set the charge at the efficient level, but introducing a low level charge would be likely to represent an efficiency gain over current revenue raising means.

4.7. Introducing an Environmental User Charge

This section discusses the issues that would need to be addressed in introducing an environmental user charge for pesticides as an example of a hazardous substance.

4.7.1. The Objective

The objective of the charge would need to be clearly defined. The discussion of theory and practice included a number of different options and specifically:

- Internalising external costs via a charge equal to marginal damage costs—here ERMA would be indifferent to the effectiveness of the charge in achieving changes in environmental effects.
- Incentivising environmental improvement;
- Raising revenue to fund actions that would achieve the purposes of the Act.

4.7.2. Charge Level

The different approaches defined by the objectives require different types of analysis to define the appropriate charge level.

The internalisation approach would require analysis of the damage costs associated with different hazardous substances and the probability of damage. Currently data for such analysis are not readily at hand.

The incentivising or behavioural charge requires, ideally, an assessment of the costs and effectiveness of options available to firms to respond to a charge and thus the expected behavioural response to different charge levels. In the absence of information required to undertake such analysis, it is possible to introduce an initial low-level charge and to vary the charge level over time. This approach is not ideal, and can introduce considerable uncertainty for business.

With a low level charge, it may be possible to demonstrate, with fairly limited information, that the costs imposed are not greater than the benefits, and thus that the charge will both have some incentive effect and represent an efficiency improvement. However, there is a risk of perverse incentives—because a charge changes not only the price of hazardous substances relative to other goods, but also the relative prices of individual hazardous substances, the resulting price changes may not reflect the differences in effect. Where the charge is set in an arbitrary way, eg a blanket per kg charge, without taking account of the way in which a substance is used, product substitution at the margin may lead to negative environmental impacts.

A revenue raising charge benefits from analysis of the likely response to the charge, in a similar way to the analysis of behavioural change. This enables projections of revenues that might be raised. Such analysis also enables an assessment of the charge in terms of tax theory—are hazardous substances inelastic goods?

4.7.3. Who to Charge

In competitive markets the expectation is that, regardless of where the charge is imposed, the effect in the market is the same. Market participants will respond in the same way, regardless of whether it is included in the price of a good or is paid in proportion to their use of a good. Considerations of the best place to levy the charge are best assessed in terms of practicality or transaction costs.

The UK analysis of a potential pesticides tax, examined options that included importers/manufacturers, retailers and final users. The large number of end users, and the disproportionate impacts on small retailers, led that analysis to suggest importers/manufacturers as the best option, despite the complications because of the need to reimburse exporters.

This approach would appear to be the simplest approach for adoption in New Zealand also.

4.7.4. Use of the Revenue

The revenue from an environmental charge can be used to replace existing taxes or to enable additional expenditure, either within ERMA or enabling the government to replace some central funding, releasing funds for additional expenditure elsewhere. The options depend, to a great extent, on the likely charge-take, ie how much revenue will be available to redistribute.

Environmental charges, especially those introduced to have an incentive effect, have some revenue uncertainty. This limits the extent to which plans can be drawn up for use of the revenue, at least until there is some experience with its use or until there has been more detailed modelling of expected effects.

For ERMA, a charge represents an opportunity to obtain additional funding. Within government, this would need to be justified on the basis of the benefits obtained.

5. Potential Use of Transferable Permits Under HSNO

This section analyses the potential use of transferable permits under HSNO.

In Section 2.4 we discussed two main types of tradable permit system—cap & trade, and credit-based systems. Credit-based systems apply to situations where a positive action is being rewarded. Examples include targeted levels of renewable generation, levels of recycling or emission reductions. For hazardous substances, the cap & trade approach is the most obvious approach. This might apply to an aggregate cap on:

- total risk or damage; or
- total production or importation of individual products or classes of products.

Where aggregate limits exist, transferable (tradable) permits represent the least cost means to achieve these limits.

The key design elements of a tradable permit system are addressed below.

5.1. Design Elements

5.1.1. The Cap

A binding limit would need to be established on either levels of damage or on levels of production or import of products.

The discussion in Section 4.1 above illustrates the difficulties in determining measurable levels of risk or damage. This is likely to be impractical to introduce.

A cap on total imports or production of a substance would be possible; it would need to be set in regulation as a binding limit.

5.1.2. Unit of Trade

The aggregate quantity of imports/production needs to be divided into separate tradable units. This can be sorted by market participants but is most readily pre-defined by the regulator. The unit needs to be sufficiently small to encourage a liquid market in which there are numerous trades. However, this needs to be balanced by the transaction costs of tracking numerous individual holdings of tradable units.

Liquidity, the measure of the quantity and rate of trading, is important because the more liquid the market, the more easily a firm can find a buyer or a seller when it needs one (and is therefore more willing to buy or to sell), and the greater the number of revealed prices. This allows market participants to compare more directly their costs with the costs of others, and helps to ensure that allowance prices are equal to marginal costs.

5.1.3. Compliance Period

The compliance or budget period is the length of time at the end of which emitters must reconcile their total emissions with allowance holdings. The choice of budget period will reflect the environmental limits on trading. Trading systems have quite different budget periods:

- proposals for international greenhouse gas emissions trading have a budget period reflecting the compliance period under the Kyoto Protocol (ie, five years from 2008–12). Parties to the Protocol must limit their emissions aggregated over the whole five-year period;
- NOx trading in the north-east US and VOC trading in Illinois applies to the summer (May–September) months because they are aimed at reducing ambient concentrations of tropospheric ozone, a largely summertime phenomenon;
- In California, the RECLAIM NO_X trading system uses overlapping annual compliance periods[70].

5.1.4. The nature of the legal obligation

Allowances provide a right to import/produce a single tonne (or kg), rather than a right to import/produce a single tonne annually—once used, the allowance cannot be used again. The obligation on firms is thus to surrender allowances in equal number to their activity levels within the budget period.

5.1.5. Allocation

Allocation of units is one of the most contentious elements of tradable permit systems. In economic terms, largely it does not matter—the price of allowances, who ends up holding them, and the actions taken to restrict activity to below the cap, should not change under different allocation systems. However, it has distributional effects. If allowances are given away this represents a transfer of wealth to holders. And for practical reasons, particularly the reduction of industry objections, it is very often the favoured approach.

[70] Allowances are issued either for a January to December year or a July to June year. Allowances can be traded between the different systems during these overlapping periods thus enabling some banking and borrowing (see below).

There have been arguments over the efficiency benefits of auctioning[71] and this is more important if markets in allowances are unlikely to be fluid, eg if there are few players or there are no simple means for exchange. But in general, the advantages are likely to be minor. One issue for which auctions have been useful is in price discovery. In the US SO_2 market, the EPA's auctions of a small (3%) percentage of available allowances allowed early discovery of expected prices of allowances which could have taken a long time under bilateral trading. Auctions also represent a means for using tradable permits as a revenue raising mechanism.

5.1.6. Non-compliance penalties

There are a number of elements of the trading system that might be covered by the compliance regime, but chiefly insufficient holding of allowances and inadequate monitoring. Effective compliance penalties are an essential component of the trading system. The compliance penalty creates the value of the allowance. It therefore needs to be sufficiently high to deter non-compliance and certainly in excess of any expected allowance price.

In the economics literature, non-compliance is viewed as a rational act that weighs the costs and benefits of compliance and non-compliance—a firm will comply if the probability of being caught times the fine is greater than the value of non-compliance.[72] The implications are that the penalty for non-compliance must be at least as great as the marginal costs of compliance, but must also reflect an additional amount because of the possibility of not being caught. The US SO_2 trading system has established a penalty regime in which the fine for non-compliance is ten times the expected costs of coming in to compliance (ie, the costs of allowances).[73]

A number of other elements of the compliance system improve its effectiveness, including certainty of the size of the penalty and automaticity (ie, firms know that if they do not comply they will face the penalty).

Penalties will also be required for failure to have a good monitoring system, or to report production levels to an adequate standard. Provided that there was an adequate compliance system for over-production or over-import, non-compliance with data requirements could be achieved through adjustments to the monitoring data such that an inflated estimate of production/imports was used instead of the missing or inadequate data. This would then require the firm to purchase additional allowances to meet emission

[71] Crampton, P. and Kerr, S. (1998), 'Tradable Carbon Allowance Auctions: How and Why to Auction', Airlie Carbon Trading Papers, Center for Clean Air Policy, Washington DC, www.ccap.org
[72] Hargrave, T., Kerr, S., Helme, N. and Denne, T. (2000), 'Treaty Compliance as Background for an Effective Trading Program', in S. Kerr (ed.) *Global Emissions Trading. Key Issues for Industrialized Countries*, Edward Elgar, pp. 43–83.
[73] It is probably less as a multiplier of the expected costs of compliance at the time the programme was designed.

requirements. Inflated figures could continue to be used until an adequate monitoring system was established.

5.1.7. Banking and borrowing

Banking and borrowing offer considerable advantages to a cap and trade system by providing flexibility in emission reductions over time in addition to flexibility over space.

Banking of allowances enables firms that hold more allowances than they need in any year to retain the excess allowances for use or sale in later years. Borrowing allows firms to use an allowance from a future year to cover emissions in the current year. Banking and borrowing reduce price volatility in the allowance market and limit the costs of over- or under-compliance, but at the possible expense of emissions being above the annual cap in some years.

Unlimited banking has been included in the US SO_2 trading system. Coupled with the stringent penalties for non-compliance, this has led to banking of considerable quantities of allowances (see Figures 5 and 6). Firms have managed their allowance holdings to limit the risks of non-compliance, and banking has limited the costs of over-compliance. In a system involving a two-step cap (with step reductions in 1995 and 2000), banking has led to a relatively smooth, downward-sloping emissions path.

Banking and borrowing enable firms to smooth the price curve of allowances[74].
In the absence of banking and borrowing, allowance prices are likely to be much more volatile; at the end of a compliance period, they may rise to very high levels (if there are shortages in supply), or fall to low levels (if there is excess supply). The system of tradable Packaging Recovery Notes (PRNs) currently used in the UK to ensure compliance with recycling and recovery obligations has no banking and borrowing, and is subject to end-of-year price spikes.

Banking improves compliance with the cap and trade system. Because it ensures that allowances have a value at the end of the compliance period, it limits the costs of over-compliance: excess allowances can be retained for use in the following year. In the absence of banking, firms will weigh up the risks of non-compliance with the costs of holding excess allowances. They will seek to maintain allowance holdings at the end of the compliance period very close to their expected emission levels. Depending on the

[74] Under unlimited banking and borrowing, and perfect information, firms might be expected to act in such a way that the price of allowances increased over time at a rate equal to the discount rate. If less than this rate, firms would have an incentive to borrow allowances from the future, as prices next year discounted to the present day would be less than prices this year. Borrowing has the effect of reducing future availability (supply) of allowances, thereby increasing expected future prices, and of increasing current supply, thereby reducing current prices. Borrowing would occur until discounted prices were equal. Likewise, with banking, if prices were expected to rise faster than the discount rate, firms would reduce emissions more now, and transfer allowances to future time periods. Again, this would occur until discounted prices were equal in all time periods.

stringency of compliance penalties, some may hold too few allowances. In contrast, if banking is allowed, the cost of holding excess allowances is the difference between their price this year and their discounted value next year.

Borrowing allows emissions to exceed the cap early in the programme and, while providing many of the cost-reduction advantages of banking, can reduce public confidence. The efficiency advantages of borrowing can be built into a cap and trade system using a declining cap. This simulates the behaviour of firms when borrowing is possible and cost-effective—ie, more reductions are made later and fewer now. The efficient use of allowances and pricing outcome over time can still be achieved, provided that there is access to unlimited banking and that the rate of decline of the cap is as steep as the maximum that the market would derive.

5.1.8. Registry

Trading schemes require a registry to manage the allowance holdings of the participants, and to compare allowance holdings with emissions.

The first requirement of such a system is that every allowance has a unique number and that, at any time, a record is kept of ownership. The number might include information regarding its year of issue and its validity, but this information might be kept centrally, against which the number can be compared.

5.1.9. Mechanism for exchange

Most markets start through a system of bilateral or over-the-counter exchanges. When there is sufficient volume of exchanges, an exchange may develop which enables transfers at a central point. This is something that the market will establish and does not require government intervention.

Brokers charge a small fee for matching buyers and sellers. Although initially higher, US brokers typically charge 50 cents per allowance transferred for SO_2 allowances worth $100–$200 per ton.

5.2. Does a Cap Make Sense?

The cap and trade system requires a cap on total use of a substance. For individual firms this does not function as a physical restriction. Where trading is allowed, the instrument works as a price mechanism. If there is a liquid market, firms will be able to obtain volumes of the substance, albeit at a price that the market determines is required to restrict activity levels so that the cap is not exceeded.

Economic theory would suggest that the charge-based approach is the best way to address an externality problem. A charge set equal to marginal damage ensures an efficient level of pollution.

A cap and trade approach has proven to be a useful mechanism when a government has set a target as a domestic policy tool and wishes to meet the target at least cost. We do not address the merits or otherwise of using targets here, but assume that they are introduced with some assessment of the acceptability of the cost of achievement, and that the benefits of doing so exceed the costs. Using a trading mechanism to meet targets provides industry with greater certainty or ability to predict cost impacts than if the government were to use a charge-based approach, with ongoing adjustments (trial and error). Firms are likely to be better able to predict markets than they can predict the government. Use of targets has increased over time in environmental policy as governments adopted "management by objectives" and other strategic management tools similar to those adopted by businesses.

There are also very practical reasons for using a trading mechanism rather than a charge. Specifically it avoids the complications surrounding revenue raising and government concerns over introducing additional taxes. This has certainly been one reason why the US adopted a trading approach to tackling acid rain, and has since extended the successful programme to other pollutants.

Concerns over permit price are reduced through initial analysis of expected permit prices. And where there is genuine price uncertainty, hybrid permit-charge systems can be adopted, such as the UK's Renewables Obligation Certification system. Here a buy-out system allows firms to pay a charge rather than purchase a certificate, setting a limit on certificate prices (or in fact the cost impact of certificates). The revenue problems are avoided through redistributing the revenues within the system—they are returned to the certificate holders.

The practical effectiveness of cap and trade systems, coupled with ongoing and increasing government concerns over additional taxes, has meant that trading systems appear to be the subject of analysis and implementation more than charge systems in international environmental policy.

5.3. Introducing a Cap and Trade System

This section discusses the key practical issues involved in establishing a cap and trade system for hazardous substances, eg pesticides.

5.3.1. The Cap

The key component of a cap and trade system is the cap. Currently limits to the use, import, manufacture or even total allowable effect of hazardous substances have not been introduced. Such an approach would be required to make a cap and trade system practicable.

There are various forms in which caps have been used. These include absolute caps (eg on emissions or production) and relative caps (eg on the rate of use of a substance).[75] A physical cap does not, in practice (and assuming a liquid market for allowances), function as a physical constraint; rather the price of allowances rises to limit demand.

5.3.2. Trading Design

An important lesson is one articulated by the US EPA as the philosophy that guided the development of its various programmes—that the government should not do anything that the private sector can do.[76] There can be a tendency to over-design trading systems and for the government to seek to set rules for elements of the system (eg, the mechanisms for exchange) that will readily be established by firms participating in the allowance market.

5.3.3. Allocation

Gifts of allowances based on historical activity data has been the most widely used approach to allocation and provides a simple means for establishing a trading system. Auctions have been used in limited circumstances to assist new entrants and to provide price transparency in the establishment of new trading systems. Recently there has been a shift towards use of rate-based allocation, particularly to deal with potential problems faced by new entrants. Set against this, theory suggests that auctions are the best option for distribution of all allowances because of the potential for economic efficiency improvements over grandfathering.

Approaches to allocation can take up extraordinary amounts of time in developing a trading system because of the importance to individual firms of the wealth transfers that result. The approach adopted is probably less important to the eventual functioning of the system. This means that these decisions are often made on very pragmatic (and often political) grounds in order to ease the introduction.

---

[75] Rate-based approaches can be problematic because they can subsidise production—firms that have a rate of use of a substance below the targeted level will generate credits for every unit of output.

[76] B. McLean, personal communication.

5.3.4. Banking and borrowing

Although tradable permit systems are usually seen as means for trading between firms within a time period, much of the efficiency gain in practical example has been achieved through trading over time, ie shifting the time period in which environmental improvements are made.

Banking has been included in all the US trading systems apart from RECLAIM as a means for limiting costs and reducing price volatility. Banking has the additional benefit of increasing the likelihood of compliance by reducing the costs of over-compliance.

5.3.5. Compliance Penalties

Effective compliance penalties are vital to defining the value of allowances. Usually these are established in the design phase to be several times as high as the expected permit price. The UK Renewables Obligation has introduced a variant in which the penalty is a legitimate buy-out of the trading scheme and used to set a limit on the price of traded certificates (or in fact a limit on the financial impact of the certificates).

6. Conclusions

6.1. The Case for Economic Instruments

In examining the potential for use of economic instruments, the argument is:

- the current regulatory system ensures identification, evaluation and management of risks, so that risk levels are reduced;
- in doing so, it imposes costs on introducers of hazardous substances that increase the costs of entry;
- this is particularly significant and potentially problematic because it applies differently to new substances from substances that are currently used—it is likely to limit the introduction of some beneficial products;
- in a competitive market, the current costs of authorisation may not be being passed on in product prices, because they have little impact on variable costs;
- thus, once products are authorised, there is no additional incentive to limit use and there is residual risk associated with hazardous substances.

In this context, economic instruments can provide incentives for risk reduction and environmental improvement by increasing the costs of every unit of substance used. This can increase economic efficiency because, for the residual emissions, economic instruments can ensure that prices reflect marginal costs, including environmental damage costs. This can be an efficiency improvement, regardless of the impacts of the existing regulatory system.

6.2. Lessons from Practice

There are considerable practical difficulties in implementing a charge equal to marginal damage. In particular, there is no current data set that would allow either the quantification of relative damage in a suitable metric (eg monetary terms), nor are there data relating to the probability of a risk event. And levying charges on different products simply on the basis of intrinsic hazardousness of the substance does not necessarily bear any relationship to the relative probability of environmental damage. Developing such a data set would be useful for evaluating and improving current regulatory approaches, independent of any move towards use of economic instruments.

In other countries, charges have been levied in a more arbitrary way simply to cause an environmental improvement. Such an approach could be used. In the absence of any information about likely costs, a low level charge would be likely to have some positive environmental impact and yield an efficiency improvement by making private costs closer to real resource (social) costs. However, because an environmental charge changes the relative prices of individual hazardous substances (in addition to the price of hazardous substances relative to other goods), the changes in prices may not reflect the differences in effects. A uniform charge (eg per kg) which does not take account of the way in which a substance is used, may lead to negative environmental impacts because of product substitution amongst hazardous substances.

Environmental charges raise revenues and might be used as an alternative revenue source for ERMA. They could be designed to meet criteria for good revenue raising mechanisms—they could both correct a market failure by charging an existing externality (environmental damage from hazardous substances) and could remove an existing distortionary revenue raising mechanism (a charge on first introducers of new substances that can discourage market entry of beneficial products). They would need to apply to all uses of specific substances, not just new introductions.

Tradable permits, in the form of a cap and trade system, would need to be introduced alongside an aggregate limit on total production or imports of specific substances. Defining a cap and assessing the potential for use of targets in hazardous substances policy is an important next step in pursuing this policy option. Tradable permits can enable the achievement of such an aggregate limit at least cost.

Experience in other countries includes examples of both instruments applied to hazardous substances. It includes charges on pesticides, and tradable permits for VOCs and lead. The choice between the instruments depends largely on the objectives. If targeting quantifiable physical improvements, tradable permits can provide greater certainty of outcome, but with some cost uncertainty. In contrast, charges provide greater price and cost certainty, while having uncertainty over environmental outcome.

6.3. Possible Next Steps

Introducing economic instruments would provide a positive addition to the existing regulatory system. They could yield environmental improvements consistent with the objectives of the HSNO Act, and environmental charges could raise revenue for ERMA in a more economically efficient way than existing mechanisms, particularly to cover the costs of risk assessment for new substance introductions. To design an instrument that achieves these improvements in an economically efficient way requires analysis to quantify current damage effects of hazardous substances, and the relationship to quantities of individual substances introduced to the economy. To understand expected outcomes, it also requires an assessment of the expected effects of the instruments on prices, demand and supply of specific substances.

Introducing a simple (eg per kg of pesticide) low-level charge is an option that could yield efficiency improvements as a revenue raising tool and would be likely to have positive but limited incentive effects. There is a slight risk that relative prices of substances are changed in a way that does not relate to risk levels.