ANALYSIS

Refunded emission payments theory, distribution of costs, and Swedish experience of NO\textsubscript{X} abatement\footnote{An early version of this paper was RFF Discussion Paper 00-29.}

Thomas Sterner*,1, Lena Höglund Isaksson1,2

Department of Economics, University of Göteborg, Box 640, 40530 Gothenburg, Sweden
Resources for the Future, Washington, USA
International Institute for Applied Systems Analysis (IIASA), Austria

Received 7 August 2003; received in revised form 2 March 2005; accepted 7 March 2005
Available online 11 May 2005

Abstract

In this paper, we discuss the effect of refunding environmental charges. Taxes are often resisted by polluters because they imply both abatement and tax costs. When charges are refunded, the incentives for abatement are essentially the same as for a tax as long as there are many producers. The incidence and distribution of costs is however different. Lower net tax payments reduce resistance from the polluters and make refunded emissions payments politically easier to implement at a sufficiently high charge level to have significant abatement effects. We describe and examine the refunded emissions payment scheme as a policy instrument and compare it with taxes and permits with regard to allocative properties, distribution of costs, property rights, and, consequently, the politics of implementation. As an empirical example, the Swedish charge on nitrogen oxides is analyzed.

Keywords: Refunded emission payments; Tax-subsidies; Two-part instruments; NO\textsubscript{X}; Environmental policy; Incentives; Output-based allocation

JEL classification: Q28; Q25; H23

1. Introduction

This paper analyzes the properties of a refunded emission payment (REP) scheme in which polluters pay a charge on pollution and the revenues are returned to the same collective of polluters as refunds in proportion to output. This instrument is considered to have been successful when applied to the abatement of nitrogen oxide (NO\textsubscript{X}) from large stationary sources in Sweden. The Swedish nitrogen charge is refunded in proportion to the physical measure of...
energy produced. This is one of the rather few charges that is directly refunded in the OECD although there are other cases where the fees are used in funds for purposes that the polluters may benefit from and influence. This applies to both road funds and water charges in many countries. Many developing and formerly planned countries such as Colombia and China also have quite extensive systems of environmental charges that are, to a large extent, refunded in different ways (Wheeler et al., 2000).³

Sweden has long had a very aggressive policy on the precursors to acid rain. Most of Scandinavia has old geological structures with little calcium and thus little buffering capacity. Sweden is one of the countries that has been most affected by acid rain leading to considerable effects on lake and forest ecosystems. This is one of the reasons why there is a very determined policy on sulphur and nitrogen emissions. When it comes to sulphur oxide there is a tax of over 4000 USD/ton which has led to dramatic reductions and very low emission levels by international standards (Hammar and Löfgren, 2001). For NOₓ, the ambitions were equally high but emission reductions were found to be much more difficult. In 1992, the Swedish government therefore decided to create an economic incentive to reduce NOₓ emissions. A charge was set at 40 Swedish kronor (SEK) per kilogram of NOₓ emitted. This corresponds to 6000 USD/ton,⁴ which can be compared to permit prices that are usually in the hundreds of dollars (although occasionally higher) in the US programs for NOₓ permits. The few other countries in Europe that have NOₓ fees, like France, Italy and Galicia in Spain, all have fee levels of about 150 USD/ton or less (Convention on Long-Range Transboundary Pollution, 1999, see also Cansier and Krumm, 1997).

Several reasons underlie Sweden’s decision to adopt this particular instrument design for NOₓ (SEPA, 1997). It was clear that the flexibility of a market-based instrument was needed since abatement costs were thought to vary considerably between firms, and advantages to scale and rapid development in abatement technology were expected. Permit trade could of course have offered the necessary flexibility but a price-type instrument was wanted partly because of habit. Fees and taxes are common while permits are novel and would thus have imposed considerable planning costs and (at least perceived) uncertainty on the ministry for the environment and the Swedish Environmental Protection Agency (SEPA).

There was also a feeling that in this type of situation, the vital issue is to get a high charge accepted and really implemented. There were many practical problems (related to high monitoring costs, exemptions, exit and entry, small polluters, trade effects, etc. that will be discussed later) as well as considerable resistance from polluters. In these circumstances, refunding is an attractive option since it strongly reduces the total cost to the affected polluters and hence the resources they will expend on lobbying to stop or lower the charge.

In Section 2, we review the relevant literature on two-part tariffs and relate the REP to other price instruments and to the corresponding quantity instruments. In Section 3, we present an overview of how the REP instrument works and compare it to other two-part tariffs and permit schemes. The rest of the paper focuses mainly on the Swedish experience of NOₓ abatement. Section 4 analyses the choice of technology and the incentives for abatement. This section also looks at the distribution of costs and the political economy of this experience. Section 5 discusses the importance of the output effect foregone by not using a tax. Section 6 finally discusses some policy implications and conclusions.

2. Two-part instruments

Environmental taxes are in simple textbook models a fully adequate policy instrument, but there are cases when they are not fully satisfactory. In these cases, it may be better to use more complex instruments which we will refer to as “two-part instruments”. These typi-
cally combine a tax or charge with some exemptions, allowances, refunds or subsidies. For instance, when there are asymmetries in information or power and influence, it is well known that simple taxes may not be appropriate, see Sterner (2002). Polluting firms may be able to resist taxes since they can form powerful lobbies and the authorities may not be able to face their threats of relocating or going out of business. Felder and Schleiniger (2002) show how effective market-based instruments, such as taxes, are politically hard to implement because of opposition by those who stand to lose environmental rents. Schneider and Volkert (1999) use a Public Choice approach to explain the difficulty in implementing such incentive-oriented policies. In many cases, governments are trying to entice firms to make heavy investments in abatement equipment. However, Gersbach (2002) shows that the inability of the regulator to precommit to a sufficiently tough regime of, for instance, taxes creates incentives for firms not to invest in the abatement equipment. This implies that the firms hold up the regulator and Gersbach shows that this can be solved by a self-financing tax-subsidy mechanism (with a subsidy on clean production) shows that this can be solved by a self-financing tax-subsidy mechanism (with a subsidy on clean production paid by the taxes on the “dirty” production). In practice, there is a tendency for policymakers therefore to abandon charges for other instruments like licensing or permits. We know, however, from Weitzman (1974) that price-type instruments, such as charges, are sometimes preferable to quantity-type instruments. In such cases, REPs or other two-part instruments, combinations of tax and refund or allowance, may be useful since they cost less for the polluters than a tax and thus cause less political opposition.

An early example of a two-tier price mechanism is found in Porter (1974). Each period, polluters pay an “entry” tax on baseline emissions but receive a refund for the pollution abated. Similar ideas can be found in Kohn (1991). Disclosure is the main idea behind the mechanisms put forth by Porter and Kohn. Reporting of both anticipated and actual emissions makes deception more difficult. The firm still pays for actual emissions—but the tax has a clever disclosure mechanism. Yohe and MacAvoy’s suggested simultaneous tax on emissions with a refund for proven abatement is similar.

Two-part instruments are also often motivated by monitoring concerns. Examples include deposit—refund systems (Bohm and Russell, 1985) or “presumptive taxes” (Eskeland and Devarajan, 1996). The latter are taxes on inputs or outputs that are complements to pollution. To avoid taxing those who abate there must be a possibility for polluters to claim refunds by proving abatement or clean technology. This shifts the burden of proof to the polluters and reduces the information and administrative costs for the regulators. The polluters who have installed cleaner technology will monitor and report, and those who have not are presumed to be polluting. Another interesting example is provided by Fullerton and Kanneman (1995), who show that a combination of tax and subsidy is needed for waste management when monitoring is impossible.

As shown by Pezzey (1992, 2003) and Pezzey and Park (1998), schemes that combine taxes and allowances or entitlements are a natural analogue to permit schemes in which some permits are allocated free of charge. The symmetry between price and quantity instruments is often overlooked. Different instruments correspond in a Coasian manner to different distributions of the cost burdens and to different perceptions of the property rights to effluents. If society has these rights, the appropriate price instrument is a tax and the corresponding quantity instrument is an auctioned permit. If the firm has the rights, the price instrument is a ‘pure’ subsidy where the baseline is the emissions the firm would have had in the absence of regulation. The quantity analogue is, however, not the usual scheme in which plants receive permits to cover a fraction of their past emissions, but the allocation of a sufficient number of free permits to cover the same baseline—combined with the repurchase by the government of permits not used due to abatement. Other instruments correspond to intermediate concepts of ownership (see Table 1).

Grandfathering means that the (free) permits are allocated in proportion to some historical variable. Usually they are allocated as a percentage of past emissions during a base year. This corresponds to an intermediate concept of property rights and the corresponding price instrument is the tax-with-allowances or TWA. Such a scheme is analyzed in Pezzey (op. cit.) and Farrow (1995). Polluters are assigned entitlements or property rights (much like grandfathered permits) to a “baseline” emission level $\bar{e}_j$ set by the regulator. They pay a charge per unit of emissions above the baseline and receive a payment per emis-
sion “saved” whenever emissions are below the baseline. The firms thus pay a net fee $\mathcal{F}(e_j - \bar{e}_j)$. This is equal to a tax $t$ and a fixed payment $\mathcal{F}$. Notice that, though similar in some respects, it is not identical to the REP scheme exactly because the payment is fixed. As shown by Pezzey and Farrow (op. cit.), the TWA scheme has the same short- and long-run efficiency properties as a tax. This is a price-type instrument but still based on rights. New firms have no property rights and must pay for their emissions.

Emission entitlements can also be allocated in proportion to other criteria, including output (as in the US phase-out of lead program in which it was the lead content of gasoline that was controlled, see Fischer, 2001 on the theory of output allocation and (Hahn and Hester, 1989 or Holley and Anderson, 1989) on the lead program). This is akin to regulating emission rates rather than emissions. It is also related to so-called benchmarking programs. The analogue among price mechanisms is the output-based refunding of charges discussed in this paper since the refund (which is analogous to the allocation of rights) is in proportion to output.

Emission payments have traditionally been taxes, but if polluters hold rights, or if they simply have the power to resist a tax, then Table 1 shows there is still a menu of price-type alternatives to tradable emission permit (TEP) schemes. Just as permits can be partly grandfathered, emission payments can be partially refunded. Many other hybrid instruments exist, such as grandfathered permits with a charge. By reducing net payments, the political objections of the polluters are minimized in the same way as with the distribution of free permits. There is, however, some difference between the addition of a fixed entitlement or allowance as in the TWA and the refunding of charges as in the REP. The fixed entitlement is the right for existing firms to a baseline level of emissions. It has the advantage of being based on rights and of having no effect on exit or entry. It does, however, involve the state in obligations to indefinite payments to firms even after they cease to exist but this need not be a difficulty: The entitlement is an asset which, like land, can be sold by the firm if it ceases production. The REP does not have this feature since the refund is paid in relation to output.

Table 1

<table>
<thead>
<tr>
<th>Rights to the environment and selection of policy instruments</th>
<th>Type of instrument</th>
<th>Holder of ownership rights to the environment</th>
<th>Quantity-type instrument</th>
<th>Price-type instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Society</td>
<td>Auctioned Permits output allocated to cover some share of pollution emitted</td>
<td>Polluter</td>
<td>Tax</td>
<td></td>
</tr>
<tr>
<td>Intermediate (State grants rights in proportion to output)</td>
<td>Grandfathered permits to cover some share of pollution emitted</td>
<td>Polluter</td>
<td>Tax-subsidy TWA</td>
<td></td>
</tr>
<tr>
<td>Intermediate (Firms have some “prior appropriation” rights)</td>
<td>Free permits with buyback from state to correspond with abatement</td>
<td>Polluter</td>
<td>Pure subsidy</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Rights to the environment and selection of policy instruments

5 An example is CFCs in the US where permits were taxed to limit the windfall to the permit recipients (Tietenberg, 1995).
and this might in some cases be seen as an advantage. Since the firms control output, the REP may create perverse incentives for entry (to cash in on the refund) or exit from (to avoid the charge) the system. This risk can however be controlled by the environmental authorities. REPs are also different from TWAs when it comes to the distribution of costs. The latter favour established firms while REPs may benefit new firms (which often tend to be “cleaner”) since rights can be gained simply by producing output.

Since there are many related price-type mechanisms based on combining a tax with an allowance, payment or refund, we might categorize them by the following criteria:

1. The payment might either be the total resource rent \( tE \) or a share thereof. It might be a fixed entitlement or a refund. Gersbach and Requate (2004) analyze partial refunding and show that in some situations an optimal degree of refunding can be defined.
2. The allocation might be based on firm emissions, firm output, or some other variable.
3. The allocation might be based on either current or historical values of that variable.

Potentially, a large number of different instruments can be created, but some combinations may be less natural, less practical or less likely than others. Allocating rights or refunds on the basis of current emissions is clearly nonsense. One could imagine refunds of current fees based on historical figures for firm output (or allocating emission rights in relation to historical output), but we will not consider these possibilities here. In the particular institutional context of a refunded emission payment, it is probably more natural to use current output. One reason for this is that firms that cease to exist need not be paid. To keep matters manageable, we will stick to the three instruments we already have discussed: tax, TWA and REP. Each of these has an exactly analogous quantity-type instrument: auctioned permits, grandfathered permits and output allocated permits.

Traditionally, subsidies have been resisted strongly by economists for at least three reasons. The economic reason is that subsidies (if related to production) lower the average cost of production and encourage both demand and excessive entry (Baumol and Oates, 1988). The administrative reason is that it is costly to make emission reduction payments indefinitely to firms that have exited the industry. Finally, payments may be resisted on the political ground that society, and not firms, should hold ultimate rights to the environment. The last argument appears to disregard the fact that some firms may have ‘prior appropriation’ rights and in practice do wield considerable influence in defending their rights. The other two arguments hinge on the exact definition of the rights to an entitlement in relation to entry and exit. They do not apply to a TWA program where only existing firms have rights and there is no perverse entry effect simply since firms cannot gain permits by entry.

Experience from real policy making shows that there is typically considerable interest in the use of charges collected and that distributional issues may determine the political feasibility of a policy instrument. Traditionally, economists have not focused strongly on these issues but this changed with the extensive debate on double dividends. As shown in Goulder et al. (1999), an instrument such as a tax operates through four different mechanisms:

1. The abatement effect.
2. The output substitution effect.
3. Tax-interaction effect.
4. Revenue-recycling effect.

The first of these is the effect that is the main concern of the decision maker when it comes to \( \text{NO}_x \) emissions. It includes all efforts to reduce emissions of \( \text{NO}_x \) per unit of energy produced. The second is, in our context, the fall in demand caused by the rise in price of energy (or other final product) due not only to abatement costs but also to tax payments. In principle, this is part of an efficient response in a first-best setting. In other contexts such as limited competition.

\[ \text{This effect can be further split into ‘input substitution’ and ‘pure abatement’ effects but this distinction is not relevant to our model.} \]

\[ \text{In the Swedish \( \text{NO}_x \) scheme, the right to entry is determined by SEPA. Zero-emission energy producers, such as wind or nuclear power plants, are not eligible, because the purpose of the scheme is to promote abatement in large combustion furnaces.} \]

\[ \text{There is also the possibility of using charges collected to finance abatement investments. This is actually quite common but not included here since it requires an extra modeling step in which firms apply for these funds.} \]
or, for instance, in a small open economy, the case is
not so clear-cut. If other countries do not pursue
similar policies, politicians may actually want to cre-
ate incentives for cleaner technology, but avoid, as far
as possible, to put extra fiscal burdens on the firms
that could lead to their relocation. The situation is
similar if a policy only targets one group of firms
(large firms in the case of the Swedish NO\textsubscript{x} charge).
Furthermore, the political economy is complex and
the output effect is not an effect politicians usually
emphasize. They do not want to be seen to abate NO\textsubscript{x}
by reducing output of electricity or electricity-using
industries. Quite the opposite—they typically go
great lengths to make sure that there are no detrimen-
tal effects on output, employment and competitiveness
of national industry. Effects 3–4 do not exist in a first-
best world but are the result of pre-existing distor-
tionary taxes to finance public consumption. The tax-
interaction effect is, as the name suggests, the (neg-
tive) effects related to the interaction between distor-
tionary taxation and the change in the consumption
basket due to the environmental policy instrument.
The revenue-recycling effect is the efficiency gain
from using the additional emissions tax revenue to
finance cuts in the distortionary labour tax.

Different instruments will display these effects
more or less fully. Taxes as well as cap and trade
permits (auctioned or free) will have full effects 1–2.
Only taxes and auctioned permits will have effect 4.
Virtually all market-based instruments (including
REPs) and even optimal performance-based standards
will have effect 1, while mandatory technology, for
instance, only has a partial abatement effect. It is an
empirical issue to judge how important the various
effects are in each particular case.

3. The REP scheme: design and comparison to
similar instruments

Assume we have \( n \) profit-maximizing firms \( (i=1, \ldots, j, \ldots, n) \), where \( j \) is an arbitrarily chosen firm
among the \( n \) firms) with production costs \( c_j(e_j, q_j) \) to
produce output \( q_j \) with emissions \( e_j \). We assume \( c_j \) is
sufficiently well behaved to give an interior solution.
An REP scheme imposes a charge of \( t \) per unit of
emission and refunds revenues to the firms propor-
tionally to output. The firms now face total supply
costs of operation \( C_j=c_j(e_j,q_j)+t[e_j−\sum_i e_i(q_i/\sum q_i)] \).
We assume that both emissions and output can be
aggregated to suitable totals and that firms act com-
petitively and take the market prices of output (\( p \)) as
well as the actions of all other agents as given. We
then maximize the profit function (Eq. (1)),\(^9\) giving
first-order conditions (2) and (3):

\[
\Pi_j = pq_j - c_j(e_j, q_j) - te_j + t \sum_i q_i \sum_t e_i
\]

\[ - c'_je_j = t(1-s_j) \]  \hspace{1cm} (2)

\[ c'_jq_j - t(E/Q)(1-s_j) = p_{\text{REP}}, \] \hspace{1cm} (3)

where total emissions \( E \) is \( \sum q_i \) total output \( Q \) is
\( \sum q_i \) and firm \( j \)’s output share \( (s_j) \) is \( q_j/Q \). For a firm
in an REP scheme, the net payment \( (N_j) \) consists of
the tax and refund terms of Eq. (1). They can be
written \( N_j=te_j(1-\Phi_j) \), where \( \Phi_j=s_j/(e_j/E) \) is rela-
tive environmental effectiveness. When \( \Phi_j>1 \), firm \( j \)
receives a net refund and is a winner in the scheme
and if \( \Phi_j<1 \) firm \( j \) is a loser. For an average firm,
\( \Phi_j=1 \) and net payment is zero.

It appears from the profit function that agent \( j \)
knows all the outputs and emissions of all agents but
this is not assumed. Agent \( j \) simply knows that there
are other agents with total emissions and output
which we can call \( E_{-j} \) and \( Q_{-j} \). Although agent \( j \)
may have expectations on these variables (based on
previous years), he treats them as unknown con-
stants in his optimization. All we assume is that
each firm is small and independent in the usual
sense of a competitive market. To compare this
calculation with a tax, (or auctioned permits) we
drop the refund term in Eq. (1). The corresponding
first-order conditions are

\[ - c'_je_j = t \]  \hspace{1cm} (4)

\[ c'_jq_j = p_{\text{tax}}. \] \hspace{1cm} (5)

A comparison between Eqs. (2) and (4) shows there
is considerable similarity between abatement incen-
tives under tax and REP schemes. The difference in

\(^9\) This model is identical to one of those analyzed in Gersbach and
incentive with respect to abatement due to refunding (which would be the same for output allocation of permits) is a factor \((1 - s_j)\) where \(s_j\) is the firm’s market share. With many small firms, \(s_j\) approaches 0, and the marginal cost of abatement \((-c'_j)\) approaches \(t\) for all firms. In this case, a refunded charge levied on all firms implies practically the same abatement level as a tax. With large output shares, the factor \((1 - s_j)\) would be more significant however. Also, we see from Eqs. (3) and (5) that the output price will not be the same. The refund in the REP implies a lower marginal cost compared to a tax of the same magnitude (see further Section 5).

In a TEP scheme with grandfathered permits, each firm is given an exogenous number of permits \((\bar{e}_j)\) that is normally related to historic emissions. The instrument would, however, operate in the same manner if the rights were allocated in proportion to some other variable as long as it is not related to anything the firm can influence (for auctioned permits \(\bar{e}_j = 0\)). Assuming the appropriate number of permits are issued and traded competitively, marginal abatement costs will be equalized at the price \(t\). The profit function then becomes

\[
\Pi_j = pq_j - c_j(e_j, q_j) - t(e_j - \bar{e}_j)
\]

Because \(\bar{e}_j\) is a constant, it is easy to see that the first-order conditions will correspond to the case of a tax (compare Eqs. (4) and (5) and see Pezzey, 1992 or Xepapadeas, 1997).

A TEP could similarly be made analogous to the REP and the profit function in (1) by allocating permits in proportion to current output. There are subtle institutional and practical details that tend to make the permits systems slightly different in practice from their price-type counterparts. The choice of allocation based on current output implies that neither firms nor regulator knows total emissions or output beforehand. In an REP, this is not a problem since the taxes are collected first and refunded afterwards so current values can be used for allocation. In the case of a rights-based system, it is normally desirable to know the number of rights “allocated” at the beginning of the year. Another possible arrangement is however a system of tradable performance standards based on emission coefficients as in Fischer (2001). This was the system used in the US for the phase-out of lead from gasoline. Each firm is given a number of permits equal to a fixed (target or “benchmark”) emission coefficient multiplied by its current production, that is, \(q_j\bar{r}\), where \(\bar{r}\) is the target emission coefficient. Note that current output affects the number of permits obtained and total emissions. The corresponding profit function is given in Eq. (7), which is analogous to Eq. (1):

\[
\Pi_j = pq_j - c_j(e_j, q_j) - P_T(e_j - q_j\bar{r})
\]

The first-order conditions are again equivalent to (2) and (3) with the price of permits \(P_T\) instead of \(t\), and the net payment that polluters make or receive is \(N_j = P_T(e_j - q_j\bar{r}) = P_T e_j (1 - \Phi_j)\). As in the REP scheme, the determining factor for being a net buyer or seller of permits is the firm’s relative environmental effectiveness \((\Phi_j)\).

REPs have recently been analyzed by Gersbach and Requate (2004) in the context of imperfect competition and pre-investment in cleaner technology. They show that while refunding of emission taxes is harmful under perfect competition, REPs can be first-best instruments under imperfect competition as long as the marginal damage from pollution exceeds the output distortion caused by imperfect competition. They also show that when there is pre-investment in cleaner technology with short-term abatement opportunities, a first-best REP scheme can be constructed where only part of the refunding is in proportion to market shares and part in proportion to shares in abatement investment.

4. The Swedish experience of NO\(_x\) fees and their effect on abatement levels

As mentioned in the introduction, the Swedish EPA wanted to reduce NO\(_x\) emissions through a high fee. A crucial problem was that a tax on all point sources of airborne NO\(_x\) emissions was infeasible because monitoring was too expensive for small units. NO\(_x\) emissions are not like sulphur dioxide (SO\(_2\)) emissions, which can easily be calculated based on the sulphur content of the fuel minus any abatement (measured from sulphur captured in ash or sludge). For SO\(_2\) emissions, a sulphur tax based on fuel content (with refunds for abatement) is an appropriate instrument (compare Yohe and MacAvoy,
NO$_x$ emissions are only partly due to the nitrogen content of the fuel. They are mainly due to an unintended chemical reaction in the combustion chamber between nitrogen and oxygen from the air. The extent and speed of this reaction happens to be highly nonlinear in temperature and other combustion parameters. This has two implications that are important in this context.

(1) Firstly, it implies that there is a large scope for NO$_x$ reduction through various technical measures which we may refer to as ‘abatement’ (although some of them may also have other effects). They include changing the shape, temperature, oxygen and moisture content of the combustion chamber. Fuel switching will automatically have a large effect on NO$_x$ as will exhaust gas re-circulation (EGR). Furthermore, there are various other specific abatement strategies such as adding reduction agents (ammonia) or passing exhausts through catalytic converters. The range in emission coefficients (emissions in relation to output) among the plants in the Swedish system has consistently been at least 20:1, indicating that abatement possibilities are large in practice too. In fact, one of the functions of an environmental charge that appears to have been particularly relevant in this case is of course to stimulate innovation in abatement technology (see, for example, Milliman and Prince, 1989).

(2) Secondly, physical measurements and monitoring are absolutely necessary. Experience from the Swedish program has shown that it is not just the installation of certain pieces of equipment but the fine-tuning of their operation that leads to NO$_x$ reduction (Högglund Isaksson, in press). For NO$_x$, actual monitoring is required both for outside inspectors and, indeed, for plant managers to know emission levels and thus which measures are successful. This is due to the complex nature of the chemical processes involved. Basing fee payments on actual measurements (which are in the Swedish case typically continuous or hourly) ensures that the payments are fair and that they do imply a real incentive for abatement.

If small units were to be exempted because of the high monitoring cost, then a high tax, levied only on large units, would be perverse, encouraging the operation of small, and usually less efficient units rather than big ones. Presumptive taxes (assuming fixed emission coefficients for all plants without monitoring equipment) would have been an alternative but presumptive taxes would not imply any incentive for abatement. At the very high level of the charge considered, such taxes would also have engendered considerable resistance (even in a tax-tolerant country such as Sweden). The argument would have been made that they put Swedish firms at a competitive disadvantage to foreign producers (particularly if the tax were based on estimated coefficients and had no incentive effect).

If the decision makers only would have had regular taxes at their disposal, they would probably have had to settle for a tax level that would have been just a small fraction of the level chosen—which would have implied very small incentives for abatement. A high tax would have engendered too much resistance. The politically feasible solution, which still permits a high charge level, was to impose the fee only on the large combustion furnaces but refund it to the same group of firms. This minimizes the effects related to the distribution of costs, as well as trade-sensitivity and thereby political resistance and lobbying. In fact, it even splits the industry lobby since the cleaner-than-average firms actually benefit instead of losing (see Fredriksson and Sterner, 2005).

The fee initially applied to all boilers, stationary combustion engines, and gas turbines producing at least 50 GWh of energy per year. Roughly 200 plants were affected including not only the energy sector but also pulp and paper mills, food, metal and other manufacturing as well as waste incineration plants. It should be noticed that a necessary requirement for the REP is that both emissions and output can be aggregated in some suitable manner. In 1996 and 1997, the limits for eligibility were lowered to 40 and 25 GWh/year, respectively, bringing in about 170 new units into the scheme (SEPA, 1993–1998). The SEPA manages the scheme at a small administrative cost amounting to 0.2–0.3% of revenues (SEPA, 1997). The entire, remaining, revenue of about 600 million SEK ($90 million) per year is refunded to the same sources that paid the charge but in proportion to output of useful energy. “Useful energy” produced has been accepted as a relevant and neutral yardstick for measuring output from this heterogeneous group of industries since the main goal is to affect combustion technologies. For power plants
and district heating plants it is equal to the energy sold. For other industries, the energy is defined as steam, hot water or electricity produced in the boiler and used in production processes or heating of factory buildings.\textsuperscript{10}

The basis for refunding is an important general issue. For NO\textsubscript{x} in Sweden, a physical basis was chosen but when thinking of this instrument in general, there may be other contexts where monetary measures of output would be more natural. Obviously this would be the case when no common physical measure can be found. However monetary measures such as total sales value, gross or net profits also are problematic since they may be manipulated by internal pricing and by the legal structure within which a firm operates. Let us compare two firms, both of which have equally big furnaces. One produces steam for district heating while the other produces steam that is used in the production of paper. Clearly the sales value of the paper produced in the latter will be several orders of magnitude higher than the simple district heat produced in the former. Refunding in proportion to sales value would have given virtually all the refunds to the paper producer. In this particular case the physical emission coefficients (g NO\textsubscript{x}/Wh energy) were the main policy target and thus energy produced was a natural basis for refunding. Other situations would have to be judged on a case-by-case basis.

As already mentioned, the abatement incentives will basically be the same as for a tax as long as market shares are small. There are currently 365 units in the scheme. They produced a total of 51 TWh in 2000 and the largest unit produced 1.1 TWh, giving it the biggest output share of 2.2%. If this plant did their arithmetic in detail, this would mean that taking the refund into account their incentive to abate would only be 0.978 \times 6000 USD/ton NO\textsubscript{x}, which is 5870 instead of 6000 $/ton. Clearly this is not a major factor (in this case) given the fairly large uncertainties of other types that surround decision making in this type of context (and also given the fact that taxes in other countries are at a level of $100 or less).

The empirical results are considered fairly satisfactory. It is not easy to determine the cost efficiency of a program such as this one empirically. Such a study would ideally compare the Swedish scheme with that of another country that uses a tax or other policy, but no countries have comparable policies. France has a tax, but its level is less than 1% of the Swedish level. Good data is needed, but there is little monitoring of plants outside this scheme. French plants, for instance, may use sporadic measurements or pre-abatement estimates based on engineering information on expected abatement effects from using various abatement techniques. Such “rule of thumb” data is questionable, in particular since experience from the Swedish plants shows large variations in abatement due to fine-tuning, as demonstrated by Höglund Isaksson (in press) using detailed post-abatement data in a study of 114 plants subject to the Swedish NO\textsubscript{x} charge. Millock and Sterner (2004) compare the French and Swedish schemes and find that the Swedish one has had much greater effect in terms of abatement. This study also provides a comparison with US data (from Burtraw and Evans, 2004) and shows that the emission factors in Sweden are very low by US standards. Even if the comparison is focused exclusively on coal fired plants (which have the highest values and are common in the US but unusual in Sweden), the Swedish plants still have emission coefficients around one third of the US Standard for new sources. Most US plants are far above this standard. There is also a detailed engineering comparison that allows for a direct comparison. The U.S. Environmental Protection Agency (1998) has compared NO\textsubscript{x} emissions from selected plants with similar technology (coal-fired boiler plants with selective catalytic reduction [SCR]) in different countries, including a plant from Sweden. The Swedish plant had about half the emission rates of the plants in

\begin{footnotesize}
\textsuperscript{10} Treating steam and electricity equally does not take into account the inherent difference in exergy or quality between these two forms of energy. Heat is less valuable in the sense that large losses are inevitable when converting heat into electricity (but not when making heat out of electricity). One could thus argue that counting heat and electricity equally implies an implicit subsidy to those plants with a large share of heat production. The plants producing steam for district heating benefit and these plants are often publicly owned. However, the way useful energy is defined implies that heat used in industrial production of paper, iron or other goods also counts. The only category of production that would be clearly disfavored would be condensation production of only electricity (with no use of waste heat). There is, however, hardly any such (fossil) electricity production in Sweden, where there is a fairly large demand from district heating systems.
\end{footnotesize}
the other countries. The report ascribes this discrepancy largely to the REP scheme:

“The Swedish retrofitted unit, in contrast, demonstrates that NOx levels well below the Swedish standard (and also below the German or United States standards) are achievable. ... The Swedish regulatory system, incorporating an economic incentive, clearly motivates [the Swedish plant] to achieve minimal NOx rates rather than just comply with the applicable emission standard". [p. 37].

The Swedish data also gives some further interesting insights. Between 1992 and 2000, mean emission rates were reduced by 40% for all the plants in the system (see Fig. 1). This reduction occurred in spite of the inclusion of many new smaller plants in 1996 and 1997. Data from these newcomers reinforce the impression that emission reductions are mainly due to the REP scheme. Their emission rates on entry into the scheme are similar to those of the original plants on entry in 1992, but they immediately start to fall in response to the charge. The reduction for the median of the original 190 plants in the system is over 50%. In an analysis of 162 abatement measures undertaken by combustion plants targeted by the NO\textsubscript{x} charge, 76 measures were claimed by the plants not to have been undertaken without the introduction of the NO\textsubscript{x} charge (Höglund Isaksson, in press). Decisive reasons for the other measures were compliance with parallel quantitative standards both for NO\textsubscript{x} and other pollutants as well as cost-effectiveness unrelated to NO\textsubscript{x} reduction. This result further strengthens the conclusion that the charge has been a principal cause for the extensive emission reductions.

The reduction is substantial because, technically, NO\textsubscript{x} abatement is considered relatively difficult. Detailed figures show that not only the average but the most and least polluting plants per unit of output have improved by roughly the same margins (SEPA, 1993–1998).

The role of the fee compared to local or national regulations to which the plants are subject has also been investigated by SEPA (1997). They found that reductions of 14,570 tons of NO\textsubscript{x} were due to investments in equipment and an additional 4000 tons were due to fine-tuning the combustion process to minimize the NO\textsubscript{x} creation during combustion. Without the fee, and with only quantitative plant limits, they estimate that emission reductions would have been less than a third of the observed. Virtually all plants were operating well below their legal limits, and the REP scheme is thus the main cause of the observed reductions.

Detailed analyses show that fuel choice (and in some cases fuel switching) is fairly important. Particularly, this applied to the early stages of the program. In 1992, the overall average emission coefficients were around 0.25 kg NO\textsubscript{x}/MWh energy output while the corresponding figures for coal and oil fired plants were close to 0.5 kg NO\textsubscript{x}/MWh and for waste incineration the figures were 0.75 kg NO\textsubscript{x}/MWh. Progress has however been considerably more rapid for the latter fuels and the gap has been reduced considerably. In 1998–1999 the averages for oil, bio-energy and waste incineration plants were all between 0.2 and 0.3 kg NO\textsubscript{x}/MWh (while coal and gas were below 0.2 kg NO\textsubscript{x}/MWh).

The REP scheme collected just over 500 million SEK in the year 2000 (75 million USD) from the companies. The net payments of the different sectors are described in annual data from SEPA (1993–1998). Due to the refund, the median firm only paid 4% of the tax as a net fee. 46% of the firms had net returns instead of paying a net fee. Only 7% of the firms in this scheme paid 50% or more of what they would

11 From 1980 to 1996, Sweden reduced total SO\textsubscript{2} emissions by 81%, but total NO\textsubscript{x} emissions (including transport) by only 22% (SEPA 1993–1998). One of the reasons for the small reduction is that NO\textsubscript{x} emissions are only partly due to nitrogen emissions from fuel as explained earlier.

Fig. 1. Mean NO\textsubscript{x} emission rates for plants targeted by a NO\textsubscript{x} charge in Sweden, 1992–2000. Note that ‘average’ refers to the average of all plants (a number that increases over the years) while ‘median 190’ refers to the median of the same 190 plants that have been included from 1992 and are still in the scheme 2000.
have paid without a refund and the sum of all the net payments was equal to 94 million SEK or less than 20% of the gross charges. It should be clear that this kind of refunding will reduce the resistance by industry to the policy so that, for instance, lobbying against the charge (or for a lower fee level) will be much weaker to the policy so that, for instance, lobbying against the kind of refunding will reduce the resistance by industry 20% of the gross charges. It should be clear that this payments was equal to 94 million SEK or less than have paid without a refund and the sum of all the net abatement costs in the energy sector are above the than other sectors. Indeed, in some cases, marginal sector has also invested more heavily in abatement overall profit from the scheme, because the energy production or supply curve (\( p^{\text{tax}} \)) is given as \( c'_j q \). The marginal cost of production under a tax will be higher than in the unregulated case both because of the tax payment itself and because the tax leads to abatement, implying a higher market price and a smaller level of production. With REPs, the marginal cost curve \( (p^{\text{rep}}) \) is given by \( c'_j q - t (1 - s_j) E / Q \). The marginal costs of production are higher than without any regulation (because of abatement costs) but lower than under a tax because of the rebate, thus the supply curve will be lower than the corresponding curve for the tax case. Since \((1 - s_j) \) is close to 1 when many firms are targeted, the supply curve will be lower by an amount very close to \( rt \), where \( r \) is the average rate of emissions \( (E / Q) \). This amount can be seen as the tax cost for the average firm of an extra unit of production. If this amount is large relative to the marginal production costs \((c'_j q \) in Eq. (3)) then the output effect is potentially important.

With the general function used, this effect is not easy to see since the marginal costs of production depend on the level of abatement. If we, however, assume additive separability between production and abatement costs, it becomes clearer. Our general cost function is then replaced by a pure production cost \( c(q) \) that only depends on output and a pure abatement cost \( c(e,q) \) that depends on both emissions and output. If we assume \( c_j(e_j,q_j) = c_{y,j}(q_j) + c_{a,j}(e_j,q_j) \), then the first-order conditions become \( p^{\text{Unreg}} = c'_{y,j} \) for the unregulated \( p^{\text{Unreg}} = \partial c_{y,j} / \partial q_j \) for the tax case, and \( p^{\text{REP}} = \partial c_{y,j} / \partial q_j + \partial c_{a,j} / \partial q_j - t (1 - s_j) E / Q \) for the REP case. We thus have \( p^{\text{Unreg}} < p^{\text{REP}} < p^{\text{tax}} \), that is, the supply price under REP will be intermediate between the unregulated supply price and the price under a Pigovian tax. This is natural since the price rises due to abatement costs but not due to the tax payment itself.

Empirically, the output effect may be more or less important: In the case of \( \text{NO}_x \) in Sweden it is small since the cost share of the fee is low, the demand elasticity of the product is small, and technical abatement possibilities large. In our particular case the average refund (which of course is equal to the average fee for the average firm) was about 1.4 USD/MWh, corresponding to between 5% and 10% of the

5. Output effects

As mentioned in Section 2, policy instruments may operate through a number of different mechanisms. In addition to abatement, a tax has, for instance, an output effect that may or may not be a crucial mechanism. Under an environmental tax, the marginal cost of production or supply curve \( (p^{\text{tax}}) \) is given as \( c'_j q \). The marginal cost of production under a tax will be higher than in the unregulated case both because of the tax payment itself and because the tax leads to abatement, implying a higher market price and a smaller level of production. With REPs, the marginal cost curve \( (p^{\text{rep}}) \) is given by \( c'_j q - t (1 - s_j) E / Q \). The marginal costs of production are higher than without any regulation (because of abatement costs) but lower than under a tax because of the rebate, thus the supply curve will be lower than the corresponding curve for the tax case. Since \((1 - s_j) \) is close to 1 when many firms are targeted, the supply curve will be lower by an amount very close to \( rt \), where \( r \) is the average rate of emissions \( (E / Q) \). This amount can be seen as the tax cost for the average firm of an extra unit of production. If this amount is large relative to the marginal production costs \((c'_j q \) in Eq. (3)) then the output effect is potentially important.

With the general function used, this effect is not easy to see since the marginal costs of production depend on the level of abatement. If we, however, assume additive separability between production and abatement costs, it becomes clearer. Our general cost function is then replaced by a pure production cost \( c(q) \) that only depends on output and a pure abatement cost \( c(e,q) \) that depends on both emissions and output. If we assume \( c_j(e_j,q_j) = c_{y,j}(q_j) + c_{a,j}(e_j,q_j) \), then the first-order conditions become \( p^{\text{Unreg}} = c'_{y,j} \) for the unregulated \( p^{\text{Unreg}} = \partial c_{y,j} / \partial q_j \) for the tax case, and \( p^{\text{REP}} = \partial c_{y,j} / \partial q_j + \partial c_{a,j} / \partial q_j - t (1 - s_j) E / Q \) for the REP case. We thus have \( p^{\text{Unreg}} < p^{\text{REP}} < p^{\text{tax}} \), that is, the supply price under REP will be intermediate between the unregulated supply price and the price under a Pigovian tax. This is natural since the price rises due to abatement costs but not due to the tax payment itself.

Empirically, the output effect may be more or less important: In the case of \( \text{NO}_x \) in Sweden it is small since the cost share of the fee is low, the demand elasticity of the product is small, and technical abatement possibilities large. In our particular case the average refund (which of course is equal to the average fee for the average firm) was about 1.4 USD/MWh, corresponding to between 5% and 10% of the
cost of energy production (SEPA, 1993–1998). The cost share in final production will for the non-energy firms be even smaller. If we assume a price elasticity\(^{12}\) of \(-0.2\) to \(-0.3\), this would give a hypothetical reduction of about 1\% in output. This means that a tax of the same high level as the REP charge level of 6000 $/ton would, in addition to the 40\% reduction in emission coefficients, have achieved a further reduction of one percentage point due to lower product demand. A tax of equal magnitude is however not a reasonable comparison because the power of the firms is too great. A tax would have been successfully resisted and if a pure tax had been used at all, it would most likely have been at a level of just a few hundred $/ton. Such a tax would have given insignificant output effects and much smaller abatement effects than the high fee made possible by refunding.

The output effect of a tax may not always be desired. This is the case if the instrument is only to be applied to a subgroup of polluters (in this case, the larger firms since costs of monitoring are too high to include the small ones). A tax would also have been fought with the argument that goods (including electricity) are internationally traded and that production could move to neighbouring countries that have no NO\(_x\) policy. It would be alleged that if the plants move to Germany or Denmark, Sweden would lose the jobs but get the NO\(_x\) as transboundary pollution.

6. Policy implications and conclusions

Although there is a form of ‘subsidy’ implicit in the refunded emission payment (REP), it is merely a repayment of a fee which may thus be set much higher than would have been possible for a tax. The crucial difference between a traditional subsidy and an REP or a tax-with-allowance (TWA) scheme is that the industry is not “subsidized” as a whole. Only in comparison to a pure tax (of equal magnitude, which may well be politically impossible) could one say that there is a subsidy. However, there is no ‘subsidy’ in relation to a policy of mandatory emission standards. Compared to such standards, the REP has two advantages: the flexibility of a market-based instrument in that each firm can choose the extent and timing of its abatement investments and the advantage of encouraging technical progress in abatement even after a target rate of emissions is achieved.

Naturally, the REP scheme has its limitations. For some environmental issues such as climate change, the potential output substitution and revenue-recycling effects forgone may be important. As with all price-type instruments, setting the optimal payment level is a tricky task. However the refunding does reduce the distributional and political tension concerning the charge level. The basis of refunding in an REP scheme requires a common output, which can be hard to define. Total revenues that include capital gains, consulting or financial services, etc. would often be unsuitable and maybe open to manipulation. In the particular case of NO\(_x\), energy output provided a reasonable common yardstick but in many other cases it would be hard to find a common measure all firms would agree to. Even in this case the use of energy rather than sold output might be considered as a distortion although we think it is small since the goal of the decision maker appears to have been to affect combustion technology. For small numbers of firms or oligopolies with large output shares, the abatement incentive would be weakened by the refund. The REP scheme does not follow the “polluters pay principle” with respect to the cost of unabated pollution (but neither do standards and other regulations).

In comparison with the TWA, a nice feature of the REPs is that they are by definition budget neutral and there is no risk the state commits itself to making indefinite payments for abatement.

Although pure REP schemes are rare, the Swedish NO\(_x\) charge described here is rather unique, other schemes have included elements of refunding to the polluters. Water and sewage charges sometimes have some of the characteristics of schemes described here, as do some energy efficiency policies (Stavins, 2001). In transitional or developing countries the use of revenues from environmental charges is a central topic. They are often earmarked for specific purposes related to the firms charged. Sometimes this is criticized due to concerns about the efficient use of public funds but this earmarking reflects political concerns that need to be taken into account. We believe that the REP scheme may be particularly attractive in devel-

\(^{12}\) The price elasticity for electricity might be a little higher, but many of the plants concerned produce steam for district heating where the price elasticity is almost zero.
oping countries where regulators typically lack the
power to directly confront all the polluters at once,
and the polluters’ threats of bankruptcy carry more
political weight than in richer countries. At the same
time, some sort of economic incentive for the adop-
tion of cleaner technology is urgently needed in these
countries. Therefore, it is important to use an instru-
ment that will not meet with too much resistance from
the business community. In these countries restricted
competition may also be more of a concern, which, as
shown by Gersbach and Requate (2004), is a further
argument for refunding.

A refunded emission payment offers a new option
to policymakers who want to choose an environmen-
tal policy instrument that will be particularly effective
in encouraging abatement to reduce pollution. The
REP scheme is the price-type counterpart to the out-
put-based allocation of tradable permits. It can be
compared to the TWA scheme that is a counterpart
to grandfathered permits. An empirical example is the
Swedish charge on NO\textsubscript{x}, which is the main factor
behind the 40% reduction in emission rates in 7
years for large stationary combustion plants. A tax
of this size would not have been politically feasible.
Nor would it have been perceived as fair if only
targeted at a subset of large polluters.

The REP scheme is a promising alternative for
many situations when taxes are neither feasible nor
desirable, for example, in a small open economy,
when the distributional consequences of taxes are
unacceptable or the opposition to taxes is too strong.
It offers an interesting alternative to permits, parti-
cularly when the regulator wants a price-type instru-
ment but does not want to place the full cost burden
on the polluters. The REP scheme also can target
subsets of polluters and thereby limit perverse incen-
tives to entry and exit. For a small open economy in
which trade-sensitivity is an issue, refunding makes
political sense.

Symbols used

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Π\textsubscript{j}</td>
<td>Firm j’s profit</td>
</tr>
<tr>
<td>(\Phi_j)</td>
<td>Firm j’s environmental effectiveness</td>
</tr>
<tr>
<td>(c_{j\Phi})</td>
<td>Marginal costs with respect to emissions</td>
</tr>
<tr>
<td>(c_{jQ})</td>
<td>Marginal costs with respect to production</td>
</tr>
<tr>
<td>(C_j)</td>
<td>Total supply costs, including costs of production and taxes, refunds, etc.</td>
</tr>
<tr>
<td>(c_j(e_j,q_j))</td>
<td>Production costs for firm j to produce output (q_j) with emissions (e_j)</td>
</tr>
<tr>
<td>(c_{y,j})</td>
<td>Firm j’s production cost when production and abatement costs are separable</td>
</tr>
<tr>
<td>(E)</td>
<td>Total emissions ((\sum e_j))</td>
</tr>
<tr>
<td>(E_{-j})</td>
<td>Total emissions by all other agents except (j)</td>
</tr>
<tr>
<td>(e_{j})</td>
<td>Firm j’s emissions of single pollutant</td>
</tr>
<tr>
<td>(\bar{e}_j)</td>
<td>Allocation of emissions permits to firm j in a permit or tax-subsidy scheme</td>
</tr>
<tr>
<td>(N_j)</td>
<td>Firm j’s net payment or refund</td>
</tr>
<tr>
<td>(\bar{P})</td>
<td>Market price of output</td>
</tr>
<tr>
<td>(\bar{p}\textsuperscript{tax}, \bar{p}\textsuperscript{rep})</td>
<td>Supply curve/marginal cost/market price of output for firms regulated by a tax or refunded emission payment scheme, respectively</td>
</tr>
<tr>
<td>(P_T)</td>
<td>Price of emissions permits</td>
</tr>
<tr>
<td>(Q)</td>
<td>Total output ((\sum q_j))</td>
</tr>
<tr>
<td>(Q_{-j})</td>
<td>Total output by all other agents except (j)</td>
</tr>
<tr>
<td>(q_j)</td>
<td>Firm j’s output of a single product ((i=1,\ldots, j,\ldots, n) is firm index)</td>
</tr>
<tr>
<td>(R)</td>
<td>Average rate of emissions ((E/Q))</td>
</tr>
<tr>
<td>(\bar{r})</td>
<td>Target rate of emissions</td>
</tr>
<tr>
<td>(s_j)</td>
<td>Firm j’s market share</td>
</tr>
<tr>
<td>(t)</td>
<td>Payment or tax level per unit of emissions</td>
</tr>
</tbody>
</table>

Acknowledgements

Thanks to Dallas Burtraw, Scott Farrow, Carolyn
Fischer, Olof Johansson-Stenman, Alan Krupnick,
Mike Toman and anonymous referees for useful com-
ments. Financial support from Naturvårdsverket,
Göteborg Energi AB, Sida, and Jan Wallanders och
Tom Hedelius’ stiftelse is gratefully acknowledged.

References

Burtraw, D., Evans, D., 2004. NO\textsubscript{x} emissions in the United States: a potpourri of policies. In: Harrington, W., Morgenstern, R.,
Sterner, T. (Eds.), Choosing Environmental Policy, Comparing Instruments and Outcomes in the United States and Europe.
RFF Press.


