Te Bao
School of Social Sciences
Nanyang Technological University
50 Nanyang Avenue
Singapore, 639798
Republic of Singapore
baote@ntu.edu.sg

Yun Dai
Department of Finance
Lingnan College
Sun Yat-sen University
135 Xingang Xi Road
Haizhu District
Guangzhou, 510275
P.R. China
daiy28@mail.sysu.edu.cn

Yongqin Wang
China Center of Economic Studies
Fudan University
220 Handan Road
Yangpu District
Shanghai, 200433
P.R. China
yongqinwang@fudan.edu.cn

Refunded NO_x-Emission Payments Scheme: A Viable Alternative to a Pigouvian Tax?

by

Arild Heimvik and Eirik S. Amundsen*

Received December 16, 2021; in revised form July 6, 2023;
accepted July 24, 2023

We consider a regulator seeking to achieve a specific target time path of NO_x emission reductions, using a refunded emission payments (REP) scheme. The paper's main objective is an efficiency comparison of two REP schemes with different refunding criteria. Both schemes may attain the targets and one of them can achieve the results from using an optimal single emission tax without any refunds. Distributional effects of applying the schemes on firms with different characteristics are also analyzed. If a single tax or a quota system cannot be used, an REP scheme based on emission reduction should be preferred.

Keywords: refunded emission payments, government policy

JEL classification code: H2, Q52, Q58

1 Introduction

Nitrogen oxides (NO_x), which consist of nitric oxide (NO) and nitrogen dioxide (NO_2), are potent gases that can cause considerable local damage. The main source of NO_x emission is the combustion of fossil fuels. These emissions lead to acid deposition and eutrophication, causing detrimental effects on both soil and water quality. NO_x (in the form of NO_2) can also have significant adverse impacts on human health. High concentrations contribute to local air pollutants' formation and lead to inflammation of the airways (European Environment Agency, 2018). Negative externalities like pollution can be internalized with corrective instruments,

* Arild Heimvik (corresponding author): Department of Economics, University of Bergen, Norway; Eirik S. Amundsen: Department of Economics, University of Bergen, Norway; Department of Food and Resource Economics, University of Copenhagen, Denmark. The paper benefitted from presentation at the Danish Environmental Economic Conference 2020 and comments and suggestions from Frode Meland, Lars Gårn Hansen, and Linda Nøstbakken. We are also indebted to two anonymous referees for valuable comments. The usual disclaimer applies.

Journal of Institutional and Theoretical Economics 180, 485–510 – ISSN 0932-4569
DOI: 10.1628/jite-2024-0002 –

such as Pigouvian taxes or emission permits (Baumol and Oates, 1988). In practice, political constraints can restrict the instrument choice of regulators.[1]

In a refunded emission payments (REP) scheme, a charge is put on the regulated firms' emissions, and the revenues are recycled back to the firms. Currently, there is limited experience with the use of REP schemes. Perhaps the most well-known example is Sweden, which introduced an REP scheme in 1992. A charge is levied per kg of NO_x emitted, and the collected funds are recycled back to the same firms – not as non-distortive lump sum transfers but rather in proportion to their output of useful energy (Sterner and Isaksson, 2006).[2] France operated with a similar system where the revenues were recycled back to firms as subsidies for abatement measures (Millock, Nauges, and Sterner, 2004). In 2008, Norway introduced a voluntary solution called the NO_x fund. The participating business organizations pay a charge upfront to the fund per kg of NO_x emitted. These revenues are then recycled back based on verified emission cuts (NOx-fondet, 2023).[3] The Norwegian system has annual and biannual emission targets announced several years ahead of fulfilment thus allowing for a gradual path of investments for abatement capacity starting with "the low hanging fruits" (i.e., for the period 2018–2025, the targets were announced in 2017).

In this paper, we investigate the case where a regulator seeks to achieve a specific target path of NO_x emission reductions. However, due to political constraints and other constraints,[4] we assume that a first-best Pigouvian tax is not available. Hence, the regulator tries to achieve cost-effective regulation with the use of an REP scheme.[5] Since the presence of constraints on regulation is not uncommon, it is an interesting analysis whether an REP scheme can function as a cost-effective instrument for reducing NO_x emission.

[1] See Amundsen, Hansen, and Whitta-Jacobsen (2022) for an analysis of indirect regulation of location-specific externalities from small-scale polluters.

[2] "Useful energy" is generally used as a benchmark concept for measuring output for industries as d under the Swedish REP scheme, whose primary goal is to affect the combustion technologies. Useful energy for power plants and district heating plants equals the energy sold. For other industries, useful energy is comprised of hot water, steam, or electricity produced in the boiler, used in heating or factory buildings or the production process (Sterner and Isaksson, 2006).

[3] Earlier versions of the Norwegian NO_X fund based the refund on the firms' expenses on NO_X-reducing investments and had two levels of refund payments depending on what kind of investment that was made. As of 2021, there is a single refund payment per kg verified NO_x reduction.

[4] The background for the introduction of the Norwegian NO_x fund was a fear that many smaller NO_X-emitting firms would not be able to survive the harsh single Pigouvian tax necessary to achieve the emission targets set by the Norwegian government (NHO, 2020). For a discussion of political and other regulatory constraints related to NO_X and other kinds of emissions inflicting costs on society, see Hagem, Hoel, and Sterner (2020).

[5] An alternative to using REP schemes on NO_x emission is to apply tradable emission permits issued free of charge according to some grandfathering rule. This would do the job of relieving the cost burden of the firms without being distortionary.

The main objective of the paper is an efficiency comparison of differing REP schemes of NO_x regulation, notably a comparison of the refunding criteria. The firms studied in the model produce energy that causes NO_x emission, and they can invest in new abatement technology. Hence, firms can reduce emission through production cuts or by investing in new abatement capacity. We examine two different designs of an REP scheme. First, we look at a version where firms pay a charge per unit of emission and receive refunds based on emission cuts. This scheme is based on the scheme currently in use in Norway. The second is an output-based scheme, where firms pay a charge per unit of emission and receive refunds in proportion to their output. This version is based on the system in place in Sweden and has a prominent place in the REP literature. We study the effects of the two REP schemes on output and investment incentives in new abatement technology.

We also take account of the fact that some economies not necessarily have only one single target to be achieved by a given date but have a plan with annual targets or milestones stretching over several years (e.g., Norway).[6] For this reason, we apply a dynamic model. To the best of our knowledge, this is the first paper to analyze the use of REP schemes with a dynamic model. Using a dynamic model allows us to analyze time-related issues such as time paths for economic instruments and paths for emission reductions. Further, our model assumes heterogeneous firms. We can then examine the effects of the two REP schemes on different firm types, given an exogenous emission target. Finally, since REP schemes are more recent additions as environmental policy, it is interesting to assess their potential as an alternative to a Pigouvian tax.[7]

A key assumption in the paper is that the regulator announces a target path of NO_x emission reductions that must be achieved using an REP scheme. It is not an assertion that the announced target path is the optimal solution for mitigating NO_x emission. We do not discuss the justification for this target path but merely assume that the regulator has announced it and that it is binding. The objective for the regulator in this paper is comparable to other established dynamic emission targets. For example, the emission target for CO_2 in the European Union (EU) is determined by the total emission cap, which is reduced by an annual percentage reduction. This increasingly stringent and binding cap raises the price of emission permits and reduces CO_2 emission to achieve the EU's long-term emission target. The Gothenburg Protocol establishes commitments for the reduction of NO_x emission, where final targets are set for the signatory countries. Furthermore, countries such as Norway have set commitments to ensure that the total NO_x emission over periods of two years does not exceed specified emission ceilings that decrease over time (NHO, 2020). In our model, we capture the cost of NO_x emission's negative

[6] Norway has had annual and biannual targets from 2008 onward. For the period 2018–2025, the target is to reduce NO_x emission by ca. 2.5% annually (NHO, 2020).

[7] Strictly speaking, the effective single tax necessary to achieve the target when there are no refunds, is not a Pigouvian tax since the politically determined targets announced by the regulator may not necessarily reflect the marginal social cost of emission. Therefore, we from now on use the term *effective single tax* instead of *Pigouvian tax*.

external effects by the shadow price of the emission constraint. The emission in our model is a function of the output of firms, as in Gersbach and Requate (2004), Sterner and Isaksson (2006), Fischer (2011), and Hagem, Hoel, and Sterner (2020).

Analyses of exogenous emission targets achievement are common in the environmental economics literature (Fell and Linn, 2013; Goulder and Parry, 2008; Wibulpolprasert, 2016). In the REP literature, the focus has been mainly on using REP schemes to reduce emission. Sterner and Isaksson (2006) discuss the theory of an output-based REP scheme and the Swedish experience from using this instrument. Gersbach and Requate (2004) showed that refunding based on market shares under perfect competition causes distortions since output levels become excessive. Under imperfect competition, however, a combination of an emission tax and an output subsidy could result in the first-best outcome. This result was supported by Fischer (2011), who also showed that with imperfect competition and endogenous refunds, significant market shares could result in reduced abatement incentives. Further, in an asymmetric Cournot duopoly, endogenous refunds could result in too high output and emission levels. Hagem, Hoel, and Sterner (2020) used a static model to compare two REP schemes under perfect competition, where refunds are given in proportion to output and as a share of expenditures for abatement equipment. They found that both schemes result in cost-ineffective abatement compared to a Pigouvian tax. Bontems (2019) combined both output- and expenditure-based refunding. He showed that this combination could remedy REP-scheme drawbacks where refunds are given for either output or abatement equipment.

Using econometric methods, other contributions have focused on the effect of REP schemes on the adoption of abatement technologies. Sterner and Turnheim (2009) studied the development of technical change for NO_x abatement from large stationary sources in Sweden. They found that an output-based REP scheme was important in reducing emission intensities. Furthermore, Bonilla et al. (2015) found that the REP scheme in Sweden had a positive effect on adopting post-combustion technologies to mitigate NO_x emission. On the other hand, Coria and Mohlin (2017) argued that it is not unambiguous whether an REP scheme provided better incentives than a standard emission tax for technological upgrades over time. The effect of the refund diminishes as the regulated sector becomes cleaner.

The paper that comes closest to ours is Hagem, Hoel, and Sterner (2020). However, our paper departs from this in that we take account of a time path of exogenous targets (or annual or biannual milestones) and, therefore, apply a dynamic model. Furthermore, our paper allows for increasing capacity costs instead of constant annuitized unit cost as in Hagem, Hoel, and Sterner (2020). Our paper also analyses a new specification of the Norwegian style REP scheme, where refunds are paid according to documented reduced NO_x emission from abatement technology and production adjustment instead of according to documented expenses on abatement technology as in Hagem, Hoel, and Sterner (2020). This version of the system has not, to our knowledge, been analyzed earlier in the literature.

Hence, our paper sets out to contribute to the REP literature in several ways. First, we examine whether an REP scheme can be a cost-effective instrument for reducing NO_x emission. Along with this we derive and discuss the necessary conditions for this to be the case in competitive markets. These results are compared with outcomes from using an output-based REP scheme. Second, our dynamic model allows us to investigate how a dynamic emission target can be achieved and how output, investments in new abatement technology, and economic instruments evolve over time. Third, we assume heterogeneous firms, and this allows us to explore how different firm types behave under different REP schemes for the same emission target. We are also able to compare the distributional effect for different firms using the two instruments, i.e., which firms are net receivers of money, and which are not.

The rest of the paper is organized as follows. In section 2, the theoretical model and its assumptions are introduced. Section 3 begins by deriving the solutions to the problem of the regulator. Next, we derive the results from the two REP schemes and discuss their implications. The analysis is extended in section 4, where we conduct comparative analyses to study firm behavior under the two REP schemes. Further, we look at the distributional effect for different firms in terms of net payments in section 5. A discussion concerning the merit of the two REP schemes is conducted in section 6. The paper is summarized, and concluding remarks are delivered in section 7.

2 The Model

In the model to follow[8] we focus on energy-producing firms that emit NO_x as part of their production process. The single firm is denoted by i, and there is a number of n firms ($i = 1,\ldots,n$). Considering an arbitrarily chosen firm, the production costs are integrated into a (gross) profit function (of output) $\pi_i(q_{it})$, where q_{it} denotes production at date t. Marginal profit is assumed decreasing in production, i.e., $\partial^2\pi_i(q_{it})/\partial q_{it}^2 = \pi''_{iqq}(q_{it}) < 0$. Further, firms are assumed heterogeneous and may differ with respect to, e.g., emission intensity. As already noted, a firm can reduce emission by either reducing production or by installing abatement technology, or both. Also, we assume competitive markets where the firms take output prices as given and, otherwise, ignore the actions of the other firms.

The capacity cost per unit abatement technology is denoted as $h_i(K_{it})$, where K_{it} is the capacity of abatement technology for firm i at date t. The function is increasing and convex in the capacity level of abatement technology, i.e.,

$$\frac{\partial h_i}{\partial K_{it}} = h_i'(K_{it}) > 0 \quad \text{and} \quad \frac{\partial^2 h_i}{\partial K_{it}^2} = h_i''(K_{it}) \geq 0.$$

Investment in new abatement technology at date t is denoted by k_t. The total cost function for abatement technology is multiplicative and expressed as $h_i(K_{it})k_{it}$. Since we solve explicitly for investment in new technology capacity, the multiplicative form of the cost function has convenient properties (see, e.g., Heimvik and

8 The nomenclature for the model is summarized in section A.1.

Amundsen, 2021). We use optimal control theory to highlight the accumulation of new abatement technology. The stock of technology acts as the state variable, and investments in new technology represent the control variable. The depreciation of the existing stock of abatement technology is denoted δ. The evolution of a firm's capacity of abatement technology is then expressed as $k_{it} - \delta K_{it}$.

The abatement technology may be integrated into the production process or be a so-called end-of-pipe technology. An end-of-pipe technology is an add-on measure used to comply with environmental regulations that reduces harmful substances arising as byproducts from production. Examples are scrubbers and catalytic converters (Frondel, Horbach, and Rennings, 2007). The emission function is denoted $m_{it} = m_i(q_{it}, K_{it})$ and has the following characteristics:

$$\frac{\partial m_{it}}{\partial q_{it}} = m'_{iq}(q_{it}, K_{it}) > 0, \quad \frac{\partial^2 m_{it}}{\partial q_{it}^2} = m''_{iqq}(q_{it}, K_{it}) \geq 0,$$
$$\frac{\partial m_{it}}{\partial K_{it}} = m'_{iK}(q_{it}, K_{it}) < 0, \quad \frac{\partial^2 m_{it}}{\partial K_{it}^2} = m''_{iKK}(q_{it}, K_{it}) \geq 0.$$

To allow for a more general abatement technology in the analytic models to follow, no specific assumptions are made for the sign of the cross-derivative: $\partial^2 m_{it}/(\partial q_{it} \partial K_{it}) = m''_{iqK}(q_{it}, K_{it})$, i.e., investments in new abatement technologies may affect the marginal emission from production. However, in the comparative analysis in the latter part of the article (section 4), we consider the case of a pure end-of-pipe technology, which amounts to assuming a separable emission function, i.e.,

$$\frac{\partial^2 m_{it}}{\partial q_{it} \partial K_{it}} = m''_{iqK}(q_{it}, K_{it}) = 0.$$

The objective of the regulator is to reduce NO_x emission in accordance with an announced target path $\overline{M}_t = \overline{M}_0 e^{-\omega t}$, where $\overline{M}_0$ and $\overline{M}_t$ is the target for date 0 and date t, respectively and ω is the rate of emission reductions the regulator seeks to achieve. Hence, the objective of the regulator may then be expressed as

$$M_t = \sum_{i=1}^{n} m_{it} \leq \overline{M}_t,$$

where M_t is the aggregate emission by the firms at date t. The target is to bring down emission by a date T, and keep it at this level for dates after T.

3 Model Applications and Analyses

3.1 Emission-Constrained Cost-Effective Optimum

In this section, we solve the optimization problem for the regulator. The objective is to maximize the total net present value of the difference between profit and the costs

of abatement technology capacity over the period from 0 to the terminating date T, given the regulator's binding target path of emission reductions. Henceforth, the solutions obtained in this section are referred to as solutions of the cost-effective optimum related to the emission targets announced by the regulator. Denoting the applied discount rate by r, the optimization problem reads

$$\max_{q_{it}, k_{it}, K_{it}} \int_0^T \sum_{i=1}^{n} [\pi_i(q_{it}) - h_i(K_{it}) k_{it}] e^{-rt}$$

subject to

$$\dot{K}_{it} = k_{it} - \delta K_{it},$$
$$\sum_{i=1}^{n} m_{it} \leq \overline{M}_t.$$

The first constraint denotes the equation of motion for the abatement technology capacity. The second constraint is the target of emission reductions announced by the regulator. By assumption, this constraint is binding. This results in a positive shadow price for the emission constraint. This is an expression of the costs of achieving the target path. Denoting the costate variable by λ_{it} and the shadow price by η_t, the corresponding present-value Hamiltonian reads

$$H_t = \sum_{i=1}^{n} [\pi_i(q_{it}) - h_i(K_{it}) k_{it}] e^{-rt} + \sum_{i=1}^{n} \lambda_{it}(k_{it} - \delta K_{it}) + \eta_t \left(\overline{M}_t - \sum_{i=1}^{n} m_{it} \right).$$

The optimization conditions are

(1) $$\frac{\partial H_t}{\partial q_{it}} = [\pi'_i(q_{it})] e^{-rt} - \eta_t m'_{iq}(q_{it}, K_{it}) = 0,$$

(2) $$\frac{\partial H_t}{\partial k_{it}} = -h_i(K_{it}) e^{-rt} + \lambda_{it} = 0,$$

(3) $$\frac{\partial H_t}{\partial K_{it}} = -h'_i(K_{it}) k_{it} e^{-rt} - \delta \lambda_{it} - \eta_t m'_{iK}(q_{it}, K_{it}) = -\dot{\lambda}_{it},$$

(4) $$\frac{\partial H_t}{\partial \lambda_{it}} = k_{it} - \delta K_{it} = \dot{K}_{it},$$

(5) $$\lambda_{iT} = 0, \quad \eta_{iT} \leq 0, \quad \overline{M}_T - \sum_{i=1}^{n} m_{iT} \geq 0, \quad \eta_{iT} \left(\overline{M}_T - \sum_{i=1}^{n} m_{iT} \right) = 0.$$

We take the time derivative of (2), insert into (3), and apply the equation of motion. Solving for the shadow price, we obtain the following expression:

(6) $$\eta_t = -\frac{(r+\delta) h_i(K_{it}) + h'_i(K_{it}) \delta K_{it}}{m'_{iK}(q_{it}, K_{it})} e^{-rt}.$$

Further, we insert (6) into (1) to obtain

$$\frac{\pi_i'(q_{it})}{m_{iq}'(q_{it},K_{it})} = -\frac{(r+\delta)h_i(K_{it})+h_i'(K_{it})\delta K_{it}}{m_{iK}'(q_{it},K_{it})} = \eta_t e^{rt}. \tag{7}$$

From (7), we have obtained conditions for the marginal cost of emission reductions related to production adjustment and to the use of abatement technology. For both mitigation measures, marginal costs of emission reductions divided by the marginal effect on emission reductions must be equal to the shadow price for the emission constraint. If emission is reduced by reducing output, the marginal cost of emission reductions is the foregone marginal profit divided by its marginal product on emission reduction. If a firm cuts emission through investment in abatement technology, the marginal cost is the extra abatement costs, divided by its marginal product on emission reduction. Further, as targets get harsher over time and the right-hand side of (7) increases, both production and investment in emission-reducing equipment will have to adapt so that the equality signs hold. Hence, as can easily be seen for the case of end-of-pipe technology, production will gradually decrease and the stock of emission-reducing equipment will gradually increase.

If the regulator introduces a single emission tax, i.e., τ_t, equal to $\eta_t e^{rt}$, per unit of emission, every single firm arrives at the same optimization conditions as those given in (1) to (5), and the cost-effective optimal solutions are obtained. This optimal single tax increases to achieve the determined target path. A key assumption in this paper, however, is that the optimal single emission tax is unavailable. In the next two sections, we analyze outcomes from using REP schemes where refunds are given for emission cuts and output.

3.2 *Refunds Based on Emission Reductions*

The REP scheme in this section is based on a version of the system currently in use in Norway. The scheme is a voluntary arrangement where participating firms pay a charge to a fund per kg of NO_x emitted. These revenues are then recycled back to the participating firms based on verified emission cuts. Further, there is a constraint on the refund, such that it cannot exceed 70% of the cost of the NO_x-reducing measure (NOx-fondet, 2023). If the firms meet their obligations through the NO_x fund, they are exempted from the government's alternative NO_x tax. If the firms are noncompliant however, they must pay the tax in proportion to their emission (Den norske Stat v/Klima- og miljødepartementet, 2017). In this paper, the goal is to examine whether an REP scheme of such a kind can be used as a cost-effective instrument to achieve a time path of emission targets.

In the model, firms pay a charge ρ_t per unit of NO_x emission upfront and receive a refund, $\beta\varphi_t$, per unit of actual emission cuts. The refund per unit is composed of a per-unit refund component, φ_t, and a constant refund constraint, β, with $0 < \beta < 1$. The refund constraint is meant to ensure that the refund does not exceed a given share of the cost of the NO_x-reducing measure undertaken by the firm. The effective refund rate is, thus, $\beta\varphi_t$. The emission cuts are defined as the difference from their

emission with no mitigation, $\hat{m}_i$. If there are no emission regulations, then from (1), the firm upholds production until marginal profit is zero. Then, there is no incentive to invest in abatement technology. Hence, the unregulated emission level is constant over time.[9] As the introduction of this REP scheme, to be denoted ERB, will affect the equilibrium market price at each date and therefore also the profit function, we denote the profit function as $\pi_i^{\mathrm{ERB}}(q_{it})$. The optimization problem for firm i reads

$$\max_{q_{it},k_{it},K_{it}} \int_0^T \left[q_{it}^{\mathrm{ERB}}(q_{it}) - h_i(K_{it})k_{it} - \rho_t m_{it} + \beta\varphi_t(\hat{m}_i - m_{it})\right]e^{-rt}$$

subject to

$$\dot{K}_{it} = k_{it} - \delta K_{it}.$$

Denoting the costate variable by ξ_{it}, the corresponding present-value Hamiltonian reads

$$H_t = \left[\pi_{it}^{\mathrm{ERB}}(q_{it}) - h_i(K_{it})k_{it} - \rho_t m_{it} + \beta\varphi_t(\hat{m}_i - m_{it})\right]e^{-rt} + \xi_{it}(k_{it} - \delta K_{it}).$$

The optimization conditions are

$$\frac{\partial H_t}{\partial q_{it}} = \left[\pi_{it}^{\mathrm{ERB}\prime}(q_{it}) - (\rho_t + \beta\varphi_t)m_{iq}'(q_{it},K_{it})\right]e^{-rt} = 0, \tag{8}$$

$$\frac{\partial H_t}{\partial k_{it}} = -h_i(K_{it})e^{-rt} + \xi_{it} = 0, \tag{9}$$

$$\frac{\partial H_t}{\partial K_{it}} = \left[-h_i'(K_{it})k_{it} - (\rho_t + \beta\varphi_t)m_{iK}'(q_{it},K_{it})\right]e^{-rt} - \delta\xi_{it} = -\dot{\xi}_{it}, \tag{10}$$

$$\frac{\partial H_t}{\partial \xi_{it}} = k_{it} - \delta K_{it} = \dot{K}_{it}, \quad \xi_{iT} = 0.$$

We take the time derivative of (9), insert into (10), and solve for the sum of the REP charge and the refund. The following expression is obtained:

$$\rho_t + \beta\varphi_t = -\frac{(r+\delta)h_i(K_{it}) + h_i'(K_{it})\delta K_{it}}{m_{iK}'(q_{it},K_{it})}. \tag{11}$$

By inserting (11) into (8), we obtain the following expression:

$$\frac{\pi_{it}^{\mathrm{ERB}\prime}(q_{it})}{m_{iq}'(q_{it},K_{it})} = -\frac{(r+\delta)h_i(K_{it}) + h_i'(K_{it})\delta K_{it}}{m_{iK}'(q_{it},K_{it})} = \rho_t + \beta\varphi_t. \tag{12}$$

[9] This result is conditional upon an emission function without drift over time. If drift is included, unregulated emission for the firm would be time-variant. This is not pursued further in this paper.

The optimality conditions in (12) equal those in (7). A combination of the emission charge and the refund rate can achieve a cost-effective solution if

$$\rho_t + \beta\varphi_t = \tau_t = \eta_t e^{rt}. \tag{13}$$

Hence, the optimal levels of emission and the production following from this REP scheme are cost-effectively optimal with respect to the targets announced by the regulator.[10] Further, as targets get harsher over time and the right-hand side of (12) increases, both production and investment in emission-reducing equipment will have to adapt so that the equality signs hold. Hence, as can easily be seen for the case of end-of-pipe technology, production will gradually decrease and the stock of emission-reducing equipment will gradually increase.

In the following, all optimal variables in this REP scheme are denoted by a *. An REP scheme is characterized by revenue neutrality and the budget constraint is binding for the sum of all the regulated firms, i.e.,

$$\rho_t^* \sum_{i=1}^{n} m_{it}^* = \beta\varphi_t^* \sum_{i=1}^{n} (\hat{m}_i - m_{it}^*). \tag{14}$$

Using (13) and (14) while recognizing that $\sum_{i=1}^{n} m_{it}^* = \overline{M}_t$, we obtain an analytical expression for the REP charge:

$$\rho_t^* = \tau_t^* \left(1 - \frac{\overline{M}_t}{\sum_{i=1}^{n} \hat{m}_i}\right). \tag{15}$$

The REP charge is a share of the cost-effective single emission tax, determined by the emission regulation's stringency. The cost-effective solutions can then be achieved using an REP scheme where the charge is at a lower level than the single emission tax. Hagem, Hoel, and Sterner (2020) showed that when refunds are given for abatement technology costs, the REP charge is also below the first-best emission tax. However, this scheme was not cost-effective as it was based on a similar earlier version of the Norwegian scheme where refunds were made according to expenses on abatement technology. In the current version of the Norwegian NO_x fund, the REP charge is also at a lower level than the NO_x tax set by the government. This was meant as an incentive to join the NO_x fund, in addition to the refund firms received (NOx-fondet, 2018).

To obtain the time path of the REP charge, we take the time derivative of (15):

$$\dot{\rho}_t^* = \dot{\tau}_t^* \left(1 - \frac{\overline{M}_t}{\sum_{i=1}^{n} \hat{m}_{it}}\right) - \tau_t \frac{\dot{\overline{M}}_t}{\sum_{i=1}^{n} \hat{m}_i}. \tag{16}$$

[10] Observe, however, that these two instruments are not equivalent when it comes to evaluating the social surplus from the regulation. In the REP scheme, revenue is returned to the firms while the revenue under the single emission tax is retained by the regulating authority and used otherwise. The effect of this may be very different.

The first term on the right side is positive since the optimal single emission tax increases to achieve the target path. The second term (including the minus sign) is positive since the target path decreases over time ($\sum_{i=1}^{n} \dot{m}_{it}^* < 0$). Hence, the REP charge must increase over time to achieve the target path.

We can derive an expression for the refund by inserting (15) into (13) and rearranging:

$$\beta\varphi_t^* = \tau_t^* \frac{\overline{M}_t}{\sum_{i=1}^{n} \hat{m}_i}. \tag{17}$$

The refund is also derived as a share of the cost-effective tax and is determined by emission regulation stringency. If β is reduced, then the refund component φ_t must increase to ensure that (17) holds. Hence, if refunds are given for all emission cuts, then the refund constraint does not determine the value of the refund.[11]

The time path for the refund is obtained by taking the time derivative of (17):

$$\beta\dot{\varphi}_t^* = \frac{\dot{\tau}_t \overline{M}_t + \tau_t^* \dot{\overline{M}}_t}{\sum_{i=1}^{n} \hat{m}_i}. \tag{18}$$

From (18), we can see that the time path of the refund is ambiguous. The denominator is positive, but the numerator is indeterminate. We know that to achieve the announced target path, the optimal single emission tax must increase. From (16), we have that the REP charge also increases over time to achieve the announced target path. Hence, there are two forces at play, the REP charge is increasing, and the total emission is decreasing. It is, therefore, indeterminate how the total budget develops as this is the product of REP charge and emission. However, if we take the time derivative of (13) and rearrange, we find that if the single emission tax increases at a higher rate than the REP charge, the refund increases as well.

With refund based on emission reductions, we find that there exists a combination of the emission charge and the refund rate, such that the REP scheme achieves the same outcomes as the optimal single emission tax. Further, the emission charge, ρ_t^*, must be lower than the single emission tax.

3.3 *Refunds Based on Output*

The REP scheme to be analyzed in this section is based on the one used in Sweden. A charge is levied per kg of NO_x emitted, and the funds are recycled back in proportion to the output of useful energy. The model in this section is based on the contributions of Gersbach and Requate (2004) and Sterner and Isaksson (2006). Both apply the same static model. We analyze an REP scheme where firms pay a charge μ_t per unit of emission and receive a refund proportional to their output at a rate equal to σ_t. We assume that output and emission can be aggregated such that

[11] For a discussion on when the regulator decides upon the share of revenues to be refunded, see Gersbach and Requate (2004).

$\sum_{i=1}^{n} q_{it} = Q_t$ and $\sum_{i=1}^{n} m_{it} = M_t$.[12] Further, the market share of firm i is defined as $s_{it} = q_{it}/Q_t$. Denoting this REP scheme by OB, the optimization problem for firm i then reads[13]

$$\max_{q_{it},k_{it},K_{it}} \int_0^T \left[\pi_{it}^{\text{OB}}(q_{it}) - h_i(K_{it})k_{it} - \mu_t m_{it} + \sigma_t q_{it} \frac{\sum_{i=1}^{n} m_{it}}{\sum_{i=1}^{n} q_{it}} \right] e^{-rt}$$

subject to

$$\dot{K}_{it} = k_{it} - \delta K_{it}.$$

Denoting the costate variable ε_{it}, the corresponding present-value Hamiltonian reads

$$H_t = \left[\pi_{it}^{\text{OB}}(q_{it}) - h_i(K_{it})k_{it} - \mu_t m_{it} + \sigma_t q_{it} \frac{\sum_{i=1}^{n} m_{it}}{\sum_{i=1}^{n} q_{it}} \right] e^{-rt} + \varepsilon_{it}(k_{it} - \delta K_{it}).$$

The optimization conditions are

$$\frac{\partial H_t}{\partial q_{it}} = \left[\pi_{it}^{\text{OB}\prime}(q_{it}) - \mu_t m'_{iq}(q_{it}, K_{it}) + \sigma_t \left(\frac{M_t}{Q_t} + s_{it} \left(m'_{iq}(q_{it}, K_{it}) - \frac{M_t}{Q_t} \right) \right) \right] e^{-rt} = 0, \tag{19}$$

$$\frac{\partial H_t}{\partial k_{it}} = -h_i(K_{it}) e^{-rt} + \varepsilon_{it} = 0, \tag{20}$$

$$\frac{\partial H_t}{\partial K_{it}} = [-h'_i(K_{it})k_{it} - (\mu_t - \sigma_t s_{it}) m'_{iK}(q_{it}, K_{it})] e^{-rt} - \delta \varepsilon_{it} = -\dot{\varepsilon}_{it}, \tag{21}$$

$$\frac{\partial H_t}{\partial \varepsilon_{it}} = k_{it} - \delta K_{it} = \dot{K}_{it}, \quad \varepsilon_{iT} = 0.$$

Denoting the optimal values of variables by ** while recognizing that $\sum_{i=1}^{n} m_{it}^{**} = \overline{M}_t$, the budget constraint can be expressed as

$$\mu_t^{**} \overline{M}_t = \sigma_t^{**} \sum_{i=1}^{n} q_{it}^{**} \frac{\overline{M}_t}{\sum_{i=1}^{n} q_{it}^{**}}. \tag{22}$$

From the budget constraint, the value of the REP charge must be equal to the refund component:

$$\mu_t^{**} = \sigma_t^{**}.$$

[12] This is the same assumption as made in Sterner and Isaksson (2006).

[13] Here, we follow Sterner and Isaksson (2006). In their paper, they argue that firm i knows that there are other firms with total emission (M_{-i}) and output (Q_{-i}). Although the firm may have expectations about these variables (based on previous years), they are treated as unknown constants in the firm's optimization.

To derive an expression for the REP charge, we first take the time derivative of (20) and set it equal to (21). Then, we substitute for the market share of firm i to obtain

$$\mu_t^{**} = -\frac{(r+\delta)h_i(K_{it}^{**}) + h'_i(K_{it}^{**})\delta K_{it}^{**}}{(1-s_{it})m'_{iK}(q_{it}^{**}, K_{it}^{**})}. \tag{23}$$

Inserting (23) into (19) provides us with the following expression:

$$\frac{\pi_i^{\text{OB}\prime}(q_{it}^{**})}{(1-s_{it})(m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**})} = -\frac{(r+\delta)h_i(K_{it}^{**}) + h'_i(K_{it}^{**})\delta K_{it}^{**}}{(1-s_{it})m'_{iK}(q_{it}^{**}, K_{it}^{**})} = \mu_t^{**}. \tag{24}$$

With output-based refunding, optimality conditions for both mitigation measures are influenced by the firm's output share. Further, the condition for production adjustments is affected by the difference between the emission intensity of firm i and the average emission intensity for all firms $(m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**})$. Further, as we saw in sections 3.1 and 3.2, when targets get harsher over time, resulting in an increase of the right-hand side of (24), both production and investment in emission-reducing equipment will have to adapt so that the equality signs hold. Hence, as can easily be seen for the case of end-of-pipe technology, production will gradually decrease and the stock of emission-reducing equipment will gradually increase.

3.4 *Comparison of Mechanisms*

As seen from (24), optimality conditions of the REP scheme based on output deviate from those of the REP scheme based on emission reductions which is given by (7) and (12) for the general case of $s_{it} > 0$ and $\partial^2 m_{it}/(\partial K_{it} \partial q_{it}) \neq 0$. However, in the case of $s_{it} \approx 0$, where there are many firms with insignificant output shares – thus corresponding to the assumption of perfect competition, condition (23) has an identical structure to that of (7) and (12). But this does not mean that the REP scheme based on output is cost-effective. As can be seen from (24) and (7), optimality conditions still differ. In section A.2, we prove formally that the REP scheme based on output is not cost-efficient when $s_{it} \approx 0$, irrespective of the sign of $\partial^2 m_{it}/(\partial K_{it} \partial q_{it})$.

As for the size of q_{it}^{**} compared to the size of q_{it}^{*}, this is in general indeterminate when $\partial^2 m_{it}/(\partial K_{it} \partial q_{it}) \neq 0$. However, if $(m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**}) < 0$, then from (24), the denominator in the first term is negative. Since $\mu_t^{**} > 0$, marginal profit must also be negative. The optimal production level for the firm then exceeds its unregulated production level, $\hat{q}_i$. Hence, $q_{it}^{**} > \hat{q}_i$. But it follows from (12) that $q_{it}^{*} < \hat{q}_i$. Hence, we must have $q_{it}^{**} > q_t^{*}$, for this case. However, no clear-cut result may be obtained with respect to the relationship between the production levels for firms where $(m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**}) > 0$, when $\partial^2 m_{it}/(\partial K_{it} \partial q_{it}) \neq 0$.

Clear-cut results, may, however, be obtained for perfect competition and separability of the emission function, i.e., where $s_{it} \approx 0$, and $\partial^2 m_{it}/(\partial K_{it} \partial q_{it}) = 0$ for all firms. As formally proven in section A.3, we have $q_{it}^{**} > q_{it}^{*}$ for all producers and further, that $\mu_t^{**} > \tau_t^{*}$. For the emission targets to be met, we must also have $K_{it}^{**} > K_t^{*}$. In this case investments in abatement technologies are independent of the level of production.

4 Firm Characteristics and Effects on Production, Abatement Technology, and Emission

In the following sections, we examine the behavior of different firms under the two REP schemes using comparative analysis. As the outcome of using an REP scheme based on emission reduction is identical to that of using a cost-effective single emission tax scheme, it suffices to compare with this. In the following we consider two firms under the same equilibrium solution (i.e., confronted with the same values of the policy instruments) and investigate cases where one of the two – otherwise identical firms – has a higher value of the parameter considered. We focus on the functions $m_i(q_{it}, K_{it})$ and $\pi_i(q_{it})$.

4.1 Difference of Emission per Unit of Output

To study firms with different emission per unit of output (i.e., the emission intensity), we define the following simplified emission function:

$$m_i(q_{it}, K_{it}) = \theta_i f(q_{it}) - \alpha g(K_{it}).$$

This function is separable and additive (i.e., the abatement technology may be considered an add-on, end-of-pipe technology). The symbol θ_i is a scalar that defines the emission intensity. The functions $f(\cdot)$ and $g(\cdot)$ are common to all firms. We assume $f'(\cdot) > 0$, $f''(\cdot) \geq 0$ and $g'(\cdot) > 0$, $g''(\cdot) \geq 0$. Also, except from the scalar of the emission function, we assume that all firms are otherwise identical (i.e., have the same profit- and cost functions) and have negligible market shares (i.e., $s_{it} \approx 0$). Equilibrium values for the REP scheme based on emission reductions and on output are, as earlier, indicated by * and **, respectively. In the following, we continue to consider negligible market shares of all firms in accordance with the assumption of perfect competition.

Consider first the REP scheme based on emission reductions and denote any two firms by i and j. Since these firms belong to the same equilibrium solution, they must face the same value of τ_t^{*}. From (12), we thus have

$$\frac{\pi_i^{\mathrm{ERB}\prime}(q_{it}^{*})}{\theta_i f'(q_{it}^{*})} = \frac{\pi_j^{\mathrm{ERB}\prime}(q_{jt}^{*})}{\theta_j f'(q_{jt}^{*})} = \tau_t^{*}. \tag{25}$$

The relationship between q_{it}^{*} and q_{jt}^{*} is implicitly determined by the functional form of (25), where τ_t is considered a constant (since both firms belong to the same equilibrium). Hence, taking the implicit derivative of (25) with respect to θ, we arrive at[14]

$$\frac{dq_t^{*}}{d\theta} = \frac{\tau f'(q_t^{*})}{\pi^{\mathrm{ERB}\prime\prime}(q_t^{*}) - \theta\tau f''(q_t^{*})} < 0.$$

Clearly, then, if $\theta_j > \theta_i$, we must have $q_{jt}^{*} < q_{it}^{*}$ since $\pi''(\cdot) < 0$ and $f''(\cdot) \geq 0$.

Likewise, for the REP scheme based on output, we must, from (24), have

$$\frac{\pi_i^{\mathrm{OB}\prime}(q_{it}^{**})}{\theta_i f'(q_{it}^{**}) - \overline{M}_t/Q_t^{**}} = \frac{\pi_j^{\mathrm{OB}\prime}(q_{jt}^{**})}{\theta_j f'(q_{jt}^{**}) - \overline{M}_t/Q_t^{**}} = \mu_t^{**}.$$

As expression (25) is also valid for this case – except that τ_t^{*} is replaced by μ_t^{**} – the same conclusion applies. Therefore, if $\theta_j > \theta_i$, we must have $q_{jt}^{**} < q_{it}^{**}$ since $\pi''(\cdot) < 0$ and $f''(\cdot) \geq 0$.

Hence, under both REP schemes, we find that a firm with a higher emission intensity will produce less than an otherwise identical firm with a lower emission intensity. As for abatement, this is the same for both firms under this case due to the separable and additive form of the emission function common to both firms. Otherwise, it is not possible to conclude which of the firms have the highest emission from production (i.e., the highest of $\theta_i f(q_{it}^{*})$ and $\theta_j f(q_{jt}^{*})$) since $\theta_j > \theta_i$ and $q_{jt}^{*} < q_{it}^{*}$. The answer to this question depends on the functional forms and parameters of the model. Generally, for the REP scheme based on emission reduction we find

$$\frac{dm_{it}^{*}}{d\theta} = f(q_{it}^{*}) + \theta f'(q_{it}^{*})\frac{dq_{it}^{*}}{d\theta}. \tag{26}$$

Inspection of signs shows that this expression is indeterminate. An identical expression is valid for the REP scheme based on output. Hence, from a policy point of view, while a firm with a higher emission intensity will produce less than an otherwise identical firm with a lower emission intensity, the firm's higher emissions per unit of output (θ) could entail that emissions are still higher.

4.2 Difference of Abatement Technology Costs

To analyze the difference in abatement technology, we define the function $\kappa_i h(K_{it})$, where κ_i is a scalar determining the level of the abatement cost function. For any two otherwise identical firms i and j, we must have

$$\frac{\kappa_i((r+\delta)h_i(K_{it}) + h_i'(K_{it})\delta K_{it})}{\alpha g'(K_{it})} = \frac{\kappa_j((r+\delta)h_j(K_{jt}^{*}) + h_j'(K_{jt}^{*})\delta K_{jt}^{*})}{\alpha g'(K_{jt}^{*})}. \tag{27}$$

[14] Strictly speaking, as the number of firms is countable (i.e., does not belong to a continuum), the comparison should be discrete in the sense of subtracting q_{it}^{*} from q_{jt}^{*} and then consider signs. However, for convenience, and to avoid long and awkward expressions, we study the sign of the derivate with respect to q_{it}^{*} instead since q_{it}^{*} is a monotone function of θ. The same approach is made for the other parameters under consideration.

Assuming the market shares are insignificant, condition (27) is the same for both REP schemes. Taking the implicit derivative of the optimality condition, (27) with respect to α, we arrive at

$$\frac{dK}{d\kappa} = \frac{(r+\delta)h(K)}{\tau\alpha g''(K) - (\kappa(r+\delta)+\delta)h'(K)) - h''(K)\delta K}.$$

An expression with the same structure is valid for the REP scheme based on output. Inspection of signs shows that the denominator is indeterminate, but strictly negative if $g''(K) = 0$. For this case, we can conclude that if $\kappa_j > \kappa_i$, then $K^*_{jt} < K^*_{it}$ and $K^{**}_{jt} < K^{**}_{it}$. Therefore, under both REP schemes, a firm with a higher abatement technology cost invests less in abatement technology than an otherwise identical firm with lower abatement cost. Emission is also larger, i.e., $m^*_{jt} > m^*_{it}$ and $m^{**}_{jt} > m^{**}_{it}$. For the case where $g''(K) > 0$, nothing in general can be said about the effects on capital stock and emission. As for emission from production, this is the same for both firms under this case due to the separable and additive form of the emission. Therefore, one may conclude that if $g''(K) = 0$, then a firm with a higher abatement cost has a higher emission than a firm with a lower abatement cost. Generally, for an REP scheme based on emission reductions, we have

(28) $$\frac{dm^*_{it}}{d\kappa} = -\tau^*_t \alpha \frac{dK^*_{it}}{d\kappa}.$$

An identical expression is valid for the REP scheme based on output. A policy point is therefore that the effects on production, abatement technology, and emission are the same irrespective of the REP scheme applied.

4.3 *Difference of Profit per Unit of Production*

Finally. to analyze the difference in the profit function, we define the function $\varsigma_i \Pi(q_{it})$, where ς_i is a scalar determining the level of the profit function. For any two otherwise identical firms i and j, under the REP scheme based on emission reductions, we must have

(29) $$\frac{\varsigma_i \pi'(q^*_{it})}{\theta f'(q^*_{it})} = \frac{\varsigma_j \pi'(q^*_{jt})}{\theta f'(q^*_{jt})} = \tau^*_t.$$

Likewise, under the REP scheme based on output for firms with insignificant market shares, we must have

$$\frac{\varsigma_i \pi'(q^{**}_{it})}{\theta f'(q^{**}_{it}) - \overline{M}_t / Q^{**}_t} = \frac{\varsigma_j \pi'(q^{**}_{jt})}{\theta f'(q^{**}_{jt}) - \overline{M}_t / Q^{**}_t} = \mu^{**}_t.$$

Taking the implicit derivative of condition (29) with respect to ς, we arrive at

(30) $$\frac{dq}{d\varsigma} = -\frac{\pi'(q)}{\varsigma\pi''(q) - \tau f''(q)} > 0.$$

An expression of the same structure is valid for the REP scheme based on output. Hence, for both schemes, it must be true that if $\varsigma_j > \varsigma_i$, then $q^*_{jt} > q^*_{it}$ and $q^{**}_{jt} > q^{**}_{it}$ since $\Pi''(\cdot) < 0$ and $f''(\cdot) \geq 0$. Generally, for the REP scheme based on emission reductions, we have

$$\frac{dm^*_{it}}{d\varsigma} = \theta f'(q^*_{it}) \frac{dq^*_{it}}{d\varsigma} > 0.$$

An identical expression is valid for the REP scheme based on output. Therefore, under both REP schemes, a firm with a higher profit per unit of production will have a higher production and a higher emission than an otherwise identical firm with a lower profit per unit of production. Also, if $f''(q^*_t) = 0$, and $\pi''(q^*_t)$ is a strictly negative constant, then the firms considered will have smaller effects on production and emission under the REP scheme based on output.

Summing up this section, it is interesting to note from a policy point of view, that there is really no difference between the two REP schemes when it comes to the directions of comparative effects of parameter changes of emission intensity, cost of abatement technology, and profit per unit of production. However, the size of the effects may differ between the two REP schemes.

5 *Comparison of Distributional Outcomes*

Under the REP schemes considered, total revenues from emission payments must equal total refunds to satisfy the property of revenue neutrality. However, for an individual firm, refunds can be lower, equal to, or higher than its emission payments. In this section, we examine the distributional outcomes for different firms under the two REP schemes. The net payment is the difference between emission payments and refunds. The expressions for the net payment of a firm in the emission-reduction and output-based REP scheme, R^*_{it} and R^{**}_{it} are shown in (31) and (32), respectively:

(31) $$R^*_t = \rho^*_t m^*_{it} - \beta_t \varphi^{**}_t (\hat{m}_i - m^*_{it}) = \tau^*_t m^*_{it} - \beta_t \varphi^*_t \hat{m}_{it},$$

(32) $$R^{**}_t = \mu^*_t m^{**}_{it} - \sigma^{**}_t q^{**}_{it} \frac{\overline{M}_t}{Q^{**}_t}.$$

At the outset, one may observe from (31) and (17) that a firm for which $m^*_i / \hat{m}_i > \overline{M}_t / \sum_{i=1}^{n} \hat{m}_i$ under the REP scheme based on emission reduction, is a net payer of money, while a firm for which the opposite is the case, is a net receiver. Hence, a firm that contributes a smaller relative reduction than the average emission reduction in the market, is a net payer. Further, from (32) and (22), we find that a firm for which $m^{**}_i / q^{**}_i > \overline{M}_t / Q^{**}$ under the output-based scheme, is a net payer, whereas a firm for which the opposite is the case, is a net receiver of money. Hence, a firm that contributes a smaller reduction per unit of production than the average reduction in the market, is a net payer.

Apart from these general results, further results may be obtained by considering the various characteristics of firms.

5.1 Difference of Emission per Unit of Output

To proceed with this analysis, we apply the same assumptions as in section 4.1. Inserting the emission function $m_i(q_{it}, K_{it}) = \theta_i f(q_{it}) - \alpha g(K_{it})$ into (31) and taking the implicit derivative with respect to θ, while recognizing that the unregulated production, $\hat{q}$, is the same for both firms (since the emission function does not enter into the decisions on this variable) and that $\widehat{K} = 0$, we arrive at

$$\frac{dR^*_{it}}{d\theta} = \tau_t \frac{dm^*_{it}}{d\theta} - \beta\varphi_{it}\frac{d\hat{m}_{it}}{d\theta} = \tau_t\left(f(q^*_{it}) + \theta f'(q^*_{it})\frac{dq^*_{it}}{d\theta}\right) - \beta\varphi_{it} f(\hat{q}).$$

As noted from (26), the sign of $dm^*_{it}/d\theta$ is indeterminate. However, if $dm^*_{it}/d\theta < 0$, i.e., firms with a higher emission intensity have lower emission after regulation, we can conclude that a firm with a high emission intensity pays less or receives more money[15] than an otherwise identical firm with a low emission intensity, is a net payer. This must be the case since the sum of all payments is equal to zero. On the other hand, if $dm^*_{it}/d\theta > 0$, then no firm conclusion can be drawn.

Likewise, for the REP scheme based on output, we find that

$$\frac{dR^{**}_{it}}{d\theta} = \mu_t \frac{dm^{**}_{it}}{d\theta} - \sigma_t \frac{\overline{M}_t}{Q^{**}_t}\frac{dq^{**}_{it}}{d\theta} = \mu_t\left[f(q^{**}_{it}) + \left(\theta f'(q^{**}_{it}) - \frac{\overline{M}_t}{Q^{**}_t}\right)\frac{dq^{**}_{it}}{d\theta}\right].$$

Again, as noted earlier, the sign of $dm^{**}_{it}/d\theta$ is indeterminate whereas $dq^{**}_{it}/d\theta < 0$. However, if $dm^{**}_{it}/d\theta < 0$, i.e., we can conclude that a firm with a high emission intensity pays more or receives less than an otherwise identical firm with a low emission intensity. Alternatively expressed, inspection of signs shows that a firm with a marginal emission less than the market-wide average emission pays more or receives less money. On the other hand, if $dm^{**}_{it}/d\theta > 0$, then no firm conclusion can be drawn.

5.2 Difference of Abatement Technology Costs

Applying the cost function $\kappa_i h(K_{it})$ from section 4.2 and taking the implicit derivative of (31) with respect to κ, while recognizing that q_{it} is the same for all firms and unaffected by the change due to separability of the emission function, we arrive at

$$\frac{dR^*_{it}}{d\kappa} = \tau_t \frac{dm^*_{it}}{d\kappa} = -\tau_t \alpha \frac{dK^*_{it}}{d\kappa}.$$

An identical expression is valid for the REP scheme based on output except that τ is replaced by μ. As noted from (28), the sign of $dm^*_{it}/d\kappa$ is indeterminate, but if $g''(K) = 0$, then $dK^*_{it}/d\kappa < 0$. Hence, for this case we can conclude for both REP schemes that a firm with high technology costs pays less or receives more money than an otherwise identical firm with low technology costs. Otherwise, if $g''(K) > 0$, no firm conclusions may be drawn.

[15] Observe that since the equilibrium will consist of many firms with many different characteristics, we cannot decide whether the firm considered is a net receiver or net payer.

5.3 Difference of Profit per Unit of Production

Applying the profit function $\varsigma_i \Pi(q_{it})$ from section 4.3 and taking the implicit derivative of (31) with respect to ς, while recognizing that K_{it} is the same for all firms and is unaffected by the change due to separability of the emission function, we arrive at

$$\frac{dR^{**}_{it}}{d\theta}\frac{dR^*_{it}}{d\varsigma} = \tau_t \frac{dm^*_{it}}{d\varsigma} = \tau_t \theta f'(q^*_{it})\frac{dq^*_{it}}{d\varsigma}.$$

As noted from (30), $dq^*_{it}/d\varsigma > 0$. Hence, one can conclude that for the REP scheme based on emission reductions, a firm with a high profit per unit of production pays more or receives less money than an otherwise identical firm with a low profit per unit of production.

As for the REP scheme based on output, we find

$$\frac{dR^{**}_{it}}{d\varsigma} = \mu_t \frac{dm^{**}_{it}}{d\varsigma} - \sigma_t \frac{\overline{M}_t}{Q^{**}_t}\frac{dq^{**}_{it}}{d\varsigma} = \mu_t\left(\theta f'(q^{**}_{it}) - \frac{\overline{M}_t}{Q^{**}_t}\right)\frac{dq^{**}_{it}}{d\varsigma}.$$

In general, this expression is indeterminate. However, if marginal emission from production is larger than the average emission, then a firm with high profit per unit of production pays more or receives less money than an otherwise identical firm with low profit per unit of production.

6 Choice of REP Scheme

The REP scheme based on output favors firms with a large share of production and a relatively small emission. For such firms one may imagine cases where they are net receivers under the REP scheme based on output, whereas they would be net payers under the REP scheme based on emission reduction. Such firms would seemingly prefer the output-based scheme over the scheme based on emission reduction. However, this is not the full story since gross profit from production and investment costs of abatement technology will be affected. As observed, all firms will produce more under the output-based scheme (assuming separability) and thereby have a higher gross revenue than under the other scheme. But, in order to comply with the emission targets, they will also incur larger investment costs than under the other scheme. Hence even though a firm will be better off in terms of net payments in the REP scheme based on output, it is not necessarily true that the same conclusion applies while considering the full opportunity cost.

In the model, we have considered the costs of achieving the time path of emission targets by maximizing the sum of net surpluses of the producers in the market and found that the REP scheme based on emission reduction has the highest value. In terms of forgone net surpluses (the opportunity costs), the REP scheme based on emission reduction has, therefore, also the lowest cost of achieving the targets and is, thus, cost-effective. As no market failures are assumed (except for the emission externality taken care of by the REP schemes) and, further, that perfect competition is assumed, the least cost found is also the least social cost of achieving the targets. Hence, collectively, the firms will be better off under the REP scheme based on emission reduction, even though some firms may be worse off. Clearly also, if all firms are identical, then all firms will be better off under the REP scheme based on emission reduction.

From the point of view of society, the REP scheme based on emission reduction is clearly preferable when it comes to minimizing the cost of achieving the given target time path. If the target path is reflecting the true social cost of emission, i.e., where the marginal social cost at all times is equal to the marginal abatement cost (in terms of production reduction and abatement technology invested), then the REP scheme is also socially optimal (and cost-effective). However, as the analysis is based on the assumption that a regulator fixes the target path that may not necessarily be socially optimal (as presumably often is the case for many pollutants), one cannot conclude that the REP scheme based on emission reduction always gives a first-best (i.e., socially optimal) result with respect to internalizing the externality. For instance, if the target path is too harshly set, it may well turn out that the production level under the REP scheme based on output is closer to the socially optimal production level than that of the REP scheme based on emission reduction, but then again, the stock of abatement capital under the REP scheme based on output would be too high compared to what is socially optimal.

All in all, it is difficult from an economic point of view to see why an REP scheme for emission reduction should be set to benefit the level of production as in the Swedish REP scheme. For this to be optimal, there must exist some kind of positive externality of production, which – in the proper meaning of the concept of externality – there normally is not. And if it were, then a separate subsidy isolated from the emission reduction is called for. Presumably, the setup of the Swedish system has been influenced by the lobbying of the larger producers that emit NO_x in Sweden (Sterner and Isaksson, 2006). Anyway, it seems like a firm conclusion that an REP scheme based on emission reduction should be preferred to an REP scheme based on output.

7 Summary and Concluding Remarks

In an REP scheme, a charge is levied per kg of NO_x emission from the regulated firms, and the collected revenues are refunded back to the same group of firms. In this paper, we analyze the use of two refund alternatives as instruments to regulate

negative externality. In the first version, firms receive refunds for their total emission cuts. In the second, refunds are given in the proportion to the firms' output.

Using a theoretical model, we show that a properly designed REP scheme can achieve the results from using an optimal cost-effective single emission tax without any refunds. For this to be the case three features must be in place. First, both production adjustments and installment of abatement technology must be eligible for refunds. Second, the necessary REP charge is then lower than the cost-effective single tax. Finally, the REP charge and the refund rate must be derived appropriately such that the sum of these two equals the optimal emission tax in each period.

In this paper an optimal REP charge and refund are derived analytically, and their time paths are determined to achieve the politically set target path of emission reductions. This REP scheme is based on refunds for NO_x reductions. An output-based REP scheme can also be used to achieve the specified target path, but, as shown, it provides nonoptimal incentives for both mitigation measures. To achieve the targets with this scheme, the REP charge must be higher than the cost-effective single tax, for the case of end-of-pipe technologies. Further, with refunds given in proportion to output, investments in new abatement technology must exceed the optimal cost-effective levels as well.

Generally, from this part of the analysis, a firm conclusion may be drawn. The REP scheme based on emission reductions is superior to that of an REP scheme based on output when it comes to achieving a given target time path of emission reductions at least costs to society. However, as the analysis is based on the assumption that a regulator fixes the target path that may not necessarily be socially optimal (as may be typical for many pollutants), one cannot conclude that the REP scheme based on emission reduction always gives a first-best (i.e., socially optimal) result with respect to internalizing the externality.

The firms studied in our model are heterogeneous. This allows us to study how different firms behave under the two REP schemes, given the binding target path of emission reductions. We use comparative analysis to alter firm characteristics and compare otherwise identical firms using a separable emission function relevant for end-of-pipe abatement technologies. It turns out that even though the REP scheme based on emission reductions is the better of the two, the comparative analysis effects of the two schemes are quite similar. For both REP schemes it is shown that a firm with higher emission intensity reduces production more than an otherwise identical firm with lower emission in abatement functions while the direction of the effect on emission is indeterminate.

Further, our results show that an otherwise identical firm with lower technology costs invests more in abatement technology under both schemes and emits less provided that the abatement function is linear. In the general case no clear-cut results apply. However, for the case of varying profitability, firm conclusions may be drawn. Hence, for both REP schemes, a firm with a higher profit per unit of production produces more and has a larger emission than an otherwise identical firm with a lower profit per unit of production.

The results are somewhat more varied between the two REP schemes when it comes to distributional outcomes of individual firms with respect to net payments. Generally, we find that a firm that contributes a smaller relative emission reduction than the average emission reduction in the market, is a net payer under the REP scheme based on emission reduction. Further, we find that a firm that contributes a smaller reduction per unit of production than the average reduction in the market, is a net payer under the REP scheme based on output.

For differences of emission intensity under the REP scheme based on emission reduction nothing general can be said as to which firms are net payers and which are net receivers. For the REP scheme based on output, however, it may generally be concluded that a firm with a marginal emission less than the market-wide average emission per unit of output pays more or receives less money than an otherwise identical firm with a lower emission intensity.

Further, our results show that an otherwise identical firm with lower technology costs pays less or receives more money under both schemes provided that the abatement function is linear. If this is not linear, then no firm conclusion may be drawn. However, for the case of varying profitability, firm conclusions may be drawn. Hence, one can conclude that for the REP scheme based on emission reductions, a firm with a high profit per unit of production pays more or receives less money than an otherwise identical firm with a low profit per unit of production. For the REP scheme based on output, more qualifications are needed. Hence, a firm with a marginal emission intensity larger than the average intensity pays more or receives less money than an otherwise identical firm with low profit per unit of production. The opposite conclusion applies if the marginal emission intensity is smaller than the average intensity.

Apart from the ability of the studied REP schemes to achieve a specific target path of emission reductions, there are also additional arguments in favor of introducing this type of instrument.[16] In an experimental setting, Kallbekken, Kroll, and Cherry (2011) found that recycling tax revenues can increase public support for environmental taxation. The use of a refunding mechanism could also make it easier to introduce charges to obtain efficient emission reductions (Johnson, 2007; Sterner and Isaksson, 2006). Further, refunds can address concerns about regulated firms' competitiveness (Sterner and Isaksson, 2006). Finally, the use of refunding could reduce the problem of emission leakage (Bernard, Fischer, and Fox, 2007; Fischer and Fox, 2012; Fischer, Greaker, and Rosendahl, 2017). While the use of an REP scheme could be appealing for a regulator, there are also drawbacks from using this instrument rather than an emission tax or auctioned emission permits. Such instruments generate revenue that could be used to reduce distortionary taxes (Goulder et al., 1999). Furthermore, REP schemes violate the pure polluter-pays principle since emitting firms receive refunds even though they are responsible for emissions.

[16] For a discussion about political economy and lobbying concerning REP schemes, see Fredriksson and Sterner (2005) and Aidt (2010).

Appendix

A.1 The Nomenclature Used in the Model

Table
Description of Variables

Symbol	Description
m_{it}	emission level of NO_x at date t for firm i
$\hat{m}_i$	emission level for firm i, with no regulations
m_{it}^*	optimal emission level at date t for firm i, given the binding target path, under emission reductions-based refunding
m_{it}^{**}	optimal emission level at date t for firm i, given the binding target path, under output-based refunding
$\overline{M}_t$	total target level of NO_x emission at date t
Ω	rate of emission reductions desired by the regulator
q_{it}	production at date t for firm i
$\hat{q}_i$	production for firm i, with no regulations
q_{it}^*	optimal production at date t for firm i, given the binding target path, under emission-reductions-based refunding
q_{it}^{**}	optimal production at date t for firm i, given the binding target path, under output-based refunding
K_{it}^*	optimal capacity of abatement technology at date t for firm i, given the binding target path, under emission-reductions-based refunding
K_{it}^{**}	optimal capacity of abatement technology at date t for firm i, given the binding target path, under output-based refunding
k_{it}	investment in new capacity of abatement technology for firm i at date t, with $k_{it} \geq 0$
$\pi_i(q_{it})$	profit for firm i at date t
R_{it}^*	net payment for firm i at date t under emission-reductions-based refunding
R_{it}^{**}	net payment for firm i at date t under output-based refunding
τ_t	optimal single emission tax at date t
ρ_t^*	REP charge in the emission reductions REP scheme at date t
φ_t^*	support rate in the emission reductions REP scheme at date t
β	refund constraint the emission reductions REP scheme, with $0 < \beta < 1$
$\beta\varphi_t^*$	refund in the emission reductions REP scheme
μ_t^{**}	REP charge in the output REP scheme at date t
σ_t^{**}	refund in the output REP scheme at date t
m_{iq}'	marginal effect of production on emission for firm i
m_{iK}'	marginal effect of abatement technology on emission for firm i
r	market discount rate
δ	depreciation rate of abatement technology capacity
T	termination date of problem considered

A.2 *Proof that the REP Scheme Based on Output (where $s_i \approx 0\ \forall i$ and $\partial^2 m_{it}/(\partial K_{it}\partial q_{it}) \neq 0$) Is Not Cost-Effective*

We know that the REP scheme based on emission reductions is cost-effective, with optimal values equal to q_{it}^*, K_{it}^*, τ_t^*, and $\sum_{i=0}^{n} m_i(q_{it}^*, K_{it}^*) = \overline{M}_t$. Assume, contrary to the claim, that the REP scheme based on output is cost-effective, such that $q_{it}^{**} = q_{it}^*$, $K_{it}^{**} = K_{it}^*$. In that case, we must, from (12) and (24), have

$$\mu_t^{**} = \frac{\pi_i^{\mathrm{OB}\prime}(q_{it}^{**})}{m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**}} > \frac{\pi_i^{\mathrm{ERB}\prime}(q_{it}^*)}{m'_{iq}(q_{it}^*, K_{it}^*)} = \tau_t^*.$$

But then, from (12) and (19), it follows that

$$\mu_t^{**} = -\frac{(r+\delta)h_i(K_{it}^{**}) + h_i'(K_{it}^{**})\delta K_{it}^{**}}{m'_{iK}(q_{it}^{**}, K_{it}^{**})}$$
$$> -\frac{(r+\delta)h_i(K_{it}^*) + h_i'(K_{it}^*)\delta K_{it}^*}{m'_{iK}(q_{it}^*, K_{it}^*)} = \tau_t^*,$$

which is a clear contradiction if $q_{it}^{**} = q_{it}^*$, $K_{it}^{**} = K_{it}^*$. Hence, we must conclude that the REP scheme based on output cannot be cost-effective. *Q.E.D.*

A.3 *Proof that $q_{it}^{**} > q_{it}^*$, $K_{it}^{**} > K_{it}^*$, and $\mu_t^{**} > \tau_t^*$ if $m''_{iqK}(q_{it}, K_{it}) = 0$*

Assume, contrary to the claim that $q_{it}^{**} \leq q_{it}^*$ and that $\mu_t^{**} \leq \tau_t^*$. From (12) and (24), we must then have

$$\text{(A1)} \qquad \mu_t^{**} = \frac{\pi_i^{\mathrm{OB}\prime}(q_{it}^{**})}{m'_{iq}(q_{it}^{**}, K_{it}^{**}) - \overline{M}_t/Q_t^{**}} \leq \frac{\pi_i^{\mathrm{ERB}\prime}(q_{it}^*)}{m'_{iq}(q_{it}^*, K_{it}^*)} = \tau_t^*.$$

As, by assumption, $\pi_i''(q_{it}) < 0$, $m''_{iqq}(q_{it}, K_{it}) \geq 0$, and $m''_{iqK}(q_{it}, K_{it}) = 0$, (A1) implies a clear contradiction.

Next, assume $q_{it}^{**} \leq q_{it}^*$ and that $\mu_t^{**} > \tau_t^*$. From (12) and (24), we must then have

$$\text{(A2)} \qquad \mu_t^{**} = -\frac{(r+\delta)h_i(K_{it}^{**}) + h_i'(K_{it}^{**})\delta K_{it}^{**}}{m'_{iK}(q_{it}^{**}, K_{it}^{**})}$$
$$> -\frac{(r+\delta)h_i(K_{it}^*) + h_i'(K_{it}^*)\delta K_{it}^*}{m'_{iK}(q_{it}^*, K_{it}^*)} = \tau_t^*.$$

Taking the implicit differential with respect to μ_t^{**} and τ_t^* of (A2), we find that $dK_{it}^{**}/d\mu_{it}^{**} > 0$ and $dK_{it}^*/d\tau_{it}^* > 0$. Hence, it follows that $K_{it}^{**} > K_{it}^*$. Since this must be true for all firms, we have $\sum_{i=0}^{n} m_i(q_{it}^{**}, K_{it}^{**}) < \overline{M}_t$, which cannot be optimal. Hence, for all relationships between μ_t^{**} and τ_t^*, we must conclude that the assumption that $q_{it}^{**} \leq q_{it}^*$ for all i, leads to contradictions. Hence, we have $q_{it}^{**} > q_{it}^*$.

Next, observe that the result that $q_{it}^{**} > q_{it}^*$ for all i leads to the conclusion that $K_{it}^{**} > K_{it}^*$ as we otherwise would have $\sum_{i=0}^{n} m_i(q_{it}^{**}, K_{it}^{**}) > \overline{M}_t$, which contradicts the prerequisite that the REP scheme based on output complies with the emission constraints. As $K_{it}^{**} > K_{it}^*$, it also follows from (A2) that $\mu_t^{**} > \tau_t^*$. *Q.E.D.*

References

Aidt, Toke S. (2010), "Green Taxes: Refunding Rules and Lobbying," *Journal of Environmental Economics and Management*, 60(1), 31–43.

Amundsen, Eirik S., Lars Gårn Hansen, and Hans Jørgen Whitta-Jacobsen (2022), "Regulation of Location-Specific Externalities from Small-Scale Polluters," *Environmental & Resource Economics*, 82(3), 507–528.

Baumol, William J., and Wallace E. Oates (1988), *The Theory of Environmental Policy*, 2nd edn., Cambridge University Press, Cambridge.

Bernard, Alain L., Carolyn Fischer, and Alan K. Fox (2007), "Is there a Rationale for Output-Based Rebating of Environmental Levies?" *Resource and Energy Economics*, 29(2), 83–101.

Bonilla, Jorge, Jessica Coria, Kristina Mohlin, and Thomas Sterner (2015), "Refunded Emission Payments and Diffusion of NO_x Abatement Technologies in Sweden," *Ecological Economics*, 116, 132–145.

Bontems, Philippe (2019), "Refunding Emissions Taxes: The Case for a Three-Part Policy," *The B.E. Journal of Economic Analysis & Policy*, 19(2), DOI: 10.1515/bejeap-2018-0123.

Coria, Jessica, and Kristina Mohlin (2017), "On Refunding of Emission Taxes and Technology Diffusion," *Strategic Behavior and the Environment*, 6(3), 205–248.

Den norske Stat v/Klima- og miljødepartementet (2017), "Miljøavtale om reduksjon av NO_x-utslipp for perioden 2018–2025 ('NO_x-avtalen 2018–2025')," https://www.nho.no/siteassets/nox-fondet/miljoavtalen/nox-avtalen-2018-2025-signert.pdf, accessed October 17, 2023.

European Environment Agency (2018), "Nitrogen Oxides (NOx) Emissions," https://www.eea.europa.eu/data-and-maps/indicators/eea-32-nitrogen-oxides-nox-emissions-1, accessed October 17, 2023.

Fell, Harrison, and Joshua Linn (2013), "Renewable Electricity Policies, Heterogeneity, and Cost Effectiveness," *Journal of Environmental Economics and Management*, 66(3), 688–707.

Fischer, Carolyn (2011), "Market Power and Output-Based Refunding of Environmental Policy Revenues," *Resource and Energy Economics*, 33(1), 212–230.

— and Alan K. Fox (2012), "Comparing Policies to Combat Emissions Leakage: Border Carbon Adjustments versus Rebates," *Journal of Environmental Economics and Management*, 64(2), 199–216.

—, Mads Greaker, and Knut Einar Rosendahl (2017), "Robust Technology Policy against Emission Leakage: The Case of Upstream Subsidies," *Journal of Environmental Economics and Management*, 84, 44–61.

Fredriksson, Per G., and Thomas Sterner (2005), "The Political Economy of Refunded Emissions Payment Programs," *Economics Letters*, 87(1), 113–119.

Frondel, Manuel, Jens Horbach, and Klaus Rennings (2007), "End-of-Pipe or Cleaner Production? An Empirical Comparison of Environmental Innovation Decisions across OECD Countries," *Business Strategy and the Environment*, 16(8), 571–584.

Gersbach, Hans, and Till Requate (2004), "Emission Taxes and Optimal Refunding Schemes," *Journal of Public Economics*, 88(3-4), 713–725.

Goulder, Lawrence H., and Ian W. H. Parry (2008), "Instrument Choice in Environmental Policy," *Review of Environmental Economics and Policy*, 2(2), 152–174.

—, —, Roberton C. Williams III, and Dallas Burtraw (1999), "The Cost-Effectiveness of Alternative Instruments for Environmental Protection in a Second-Best Setting," *Journal of Public Economics*, 72(3), 329–360.

Hagem, Cathrine, Michael Hoel, and Thomas Sterner (2020), "Refunding Emission Payments: Output-Based versus Expenditure-Based Refunding," *Environmental & Resource Economics*, 77(3), 641–667.

Heimvik, Arild, and Eirik S. Amundsen (2021), "Prices vs. Percentages: Use of Tradable Green Certificates as an Instrument of Greenhouse Gas Mitigation," *Energy Economics*, 99, DOI: 10.1016/j.eneco.2021.105316.
Johnson, Kenneth C. (2007), "Refunded Emission Taxes: A Resolution to the Cap-versus-Tax Dilemma for Greenhouse Gas Regulation," *Energy Policy*, 35(5), 3115–3118.
Kallbekken, Steffen, Stephan Kroll, and Todd L. Cherry (2011), "Do you Not Like Pigou, or do you Not Understand him? Tax Aversion and Revenue Recycling in the Lab," *Journal of Environmental Economics and Management*, 62(1), 53–64.
Millock, Katrin, Céline Nauges, and Thomas Sterner (2004), "Environmental Taxes: A Comparison of French and Swedish Experience from Taxes on Industrial Air Pollution," *CESifo DICE Report*, 2(1), 30–34.
NHO (2020), "Story about the NOx Fund," https://www.noxfondet.no/en/articles/about-the-nox-fond/, accessed October 23, 2023.
NOx-fondet (2018), "Årsrapport 2017: Rapport om oppfyllelse av reduksjonsforpliktelsen for 2017 i henhold til Miljøavtalen om NOx 2011–2017," https://www.nho.no/siteassets/nox-fondet/rapporter/2018/nox-fondet-arsrapport-2017_oppslag.pdf, accessed October 17, 2023.
— (2023), "GUIDELINE: Application for financial support under the NOx Agreement 2018–2025," version 5.1, https://www.noxfondet.no/siteassets/dokumenter/vedlegg/2023/nox-veiledning-engelsk-versjon-oktober-2023.pdf, accessed October 23, 2023.
Sterner, Thomas, and Lena Höglund Isaksson (2006), "Refunded Emission Payments Theory, Distribution of Costs, and Swedish Experience of NO_x Abatement," *Ecological Economics*, 57(1), 93–106.
— and Bruno Turnheim (2009), "Innovation and Diffusion of Environmental Technology: Industrial NOx Abatement in Sweden under Refunded Emission Payments," *Ecological Economics*, 68(12), 2996–3006.
Wibulpolprasert, Wichsinee (2016), "Optimal Environmental Policies and Renewable Energy Investment: Evidence from the Texas Electricity Market," *Climate Change Economics*, 7(4), DOI: 10.1142/S201000781650010X.

Arild Heimvik
Department of Economics
University of Bergen
Box 7802
5020 Bergen
Norway
arild.heimvik@econ.uib.no

Eirik S. Amundsen
Department of Economics
University of Bergen
Box 7802
5020 Bergen
Norway
eirik.amundsen@econ.uib.no

Entry-Deterrent Licensing Revisited

by

Shuai Niu*

Received March 3, 2023; in revised form August 14, 2023;
accepted November 16, 2023

Consider a market with an incumbent and a potential entrant. The potential entrant owns an innovation so that it can produce more efficiently than the incumbent if it enters the market by starting a new firm. The potential entrant can license the innovation to the incumbent. The gain from licensing is higher if the potential entrant does not enter the market. It is possible that the potential entrant enters without licensing and does not enter with licensing. Licensing may reduce social welfare if it induces the potential entrant not to enter the market.

Keywords: process innovation, licensing, entry, welfare

JEL classification code: L13, L24, L41

1 Introduction

Innovations used within an industry often originate outside the boundaries of established manufacturing firms (Hienerth and Lettl, 2011).[1] Sometimes external innovators enter the market by creating new firms to drive technological change in an industry. For example, in the computer industry in the early 1980s, Sun Microsystems exploited its technological leadership by entering the workstation market and competing with incumbents such as Digital, IBM, and Apollo Systems (Gans and Stern, 2003). Sometimes external innovators do not create new firms but license their innovations to incumbents. For example, in 2000 the technology company Growth Networks formed an alliance with the incumbent Cisco Systems to provide

* Shandong University, Jinan, P.R. China. I would like to thank Gerd Muehlheusser and the two anonymous referees for helpful and insightful comments, which I believe have helped me substantially improve the paper. I am grateful also to the financial support from The National Natural Science Foundation of China (grant 72192842). I am responsible for all the remaining errors.

[1] According to Reinganum (1983), technological change in an industry can occur outside the industry because the incumbent may invest less than the challenger in a given research project if the innovation is sufficiently revolutionary and the inventive process is stochastic.