Carbon leakage: a mechanism design approach

Lassi Ahlvik and Matti Liski*

October 5, 2017

Abstract

Polluting firms often press policy makers to offer compensations or roll back regulations by threatening to relocate. Yet, only firms know if they can actually carry out the threat. Using a mechanism design approach, we show how exclusions and the intensity of regulation can be used to save on public funds and allocate regulatory assignments efficiently. Some leakage is always optimal, and the intensity of regulation increases when leakage is introduced; the regulation can even become stronger than in the first best. We provide an illustrative quantification of the optimal carbon leakage policy for the EU emissions trading program.

Keywords: Climate change; emissions trading; carbon leakage; private information; mechanism design.

JEL codes: D82; L51; Q54; Q58

*Ahlvik (lassi.ahlvik@nhh.no): Economics department of the Norwegian School of Economics. Liski (matti.liski@aalto.fi): Economics department of Aalto University. We thank Chang-Koo Chi, Baran Doda, Bård Harstad, Daniel Hauser, Johannes Hörner, Jussi Lintunen, Juan-Pablo Montero, Matias Pietola, Julia Salmi, Morten Sæthre, Po Yin Wong, Andrey Zhukov and participants at presentations in Athens (EAERE), Bergen, Helsinki, Paris, Oslo, Toulouse (11th TSE conference on climate and energy), and Zurich for valuable comments and discussion. Matti is grateful for funding from the Academy of Finland program New Energy, and from the Aalto University project SAGA.

1 Introduction

It is tricky to design local regulations on global externalities. Firms may relocate and merely shift the externality-causing activities to another location if the regulation is too costly and inadequately compensated. Yet, the true ease of moving is only known by the firm, and any compensation scheme comes at a risk of paying unnecessarily high transfers to some of the firms. Regulated industries fiercely lobby for compensations, emphasizing the cost of regulation and easiness to relocate production to other countries. If the aim is to keep all the firms, these claims need to be respected although they are difficult to verify. But doing so will create potentially large private rents to the industries.

The issue has become topical in the EU Emission Trading Scheme (ETS), where industrial emitters considered to be susceptible to "carbon leakage" have received both generous allocations of free emission allowances and direct monetary compensations.[1] In concrete terms, the electricity producers received windfall profits estimated to reach one billion euros between 2005 and 2013; the aviation sector, which has been partly included in the EU ETS since 2012, was estimated to have received two billion euros in windfalls.[2] The total windfall for the european industries from compensation schemes planned to prevent industry relocation has been estimated to be as high as 25 billion euros, with companies in cement, petrochemical and steel industries gaining the most. One reason suggested for this is "*industrial-scale lobbying from the vested interests which has led to vast windfalls being handed to many companies in the energy-intensive sector*".[3]

This set-up raises an important question for the policy design that we address in this

[1] In total 43 % of the allowances are given away for free during 2013-2020. Sectors deemed to be exposed to a significant risk of carbon leakage receive 100 % of their estimated allowance need for free, whereas the free allocation to non-leakage sector is gradually reduced to 30 % by year 2020. In addition, the most energy-intensive sectors can be given monetary compensation through national state aid schemes (EC, 2017).

[2] Harrabin, Roger. 2006. "£1 bn windfall from carbon trade." *BBC*, May 1. http://news.bbc.co.uk/2/hi/science/nature/4961320.stm. Nelsen, Arthur. 2012. "Airlines could net £1.6bn windfall from EU carbon trading scheme, report says." *The Guardian* 11 January. https://www.theguardian.com/environment/2012/jan/11/airlines-windfall-eu-carbon-trading

[3] Krukowska, Ewa. 2016. "EU industry got $27 billion carbon plan windfall, study says." *Bloomberg*, March 15. http://www.bloomberg.com/news/articles/2016-03-14/eu-industry-got-27-billion-cap-and-trade-windfall-study-says. Carrington, Damian. 2013. "Carbon fat cats are killing the emissions trading mouse", *The Guardian*, February 14. https://www.theguardian.com/environment/damian-carrington-blog/2013/feb/14/carbon-emissions- carbon-tax

paper: how should such local regulations on a global problem be designed to reach environmental targets and to limit "leakage" without creating excessive private rents? In particular, *how much* industries under leakage risk should be compensated, and *how* they should be compensated — with lump-sum transfers or by compromising on the level of regulation? The main challenge, that seems unavoidable, is that only firms know, firstly, the true costs of moving to another location and, secondly, the cost of compliance if staying in the regime. Taking a mechanism design approach, we study the optimal design of environmental policies when firms have such two-dimensional private information.

Two main results arise from our analysis. First, we find that although carbon leakage is socially wasteful, the solution to the information problem calls for some positive leakage as an optimal outcome. It is central to the mechanism that there is a tool for disciplining threats. This can be achieved by making firms responsible for emphasizing a low cost of relocation — the mechanism holds firms accountable for the claims in the sense that they can be excluded from the regime. The optimal level of compensation strikes a balance between firms' marginal propensity to relocate, net damages due to leakage, and the social cost of wasting public funds to overcompensate industries. Second, we show that the industries exposed to the risk of leakage should not be given a lower level of regulation but leakage instead calls for a *stricter* regulation, accompanied by lump-sum transfers. Without leakage, only the firms with low abatement costs receive information rents, and a lower emissions price is a way to reduce those rents as, for instance, in Lewis (1996). With leakage, the low-cost firms are no longer "cornered" and may relocate without contributing to emissions reductions, leading to an increase in the optimal emissions price attracting firms with low costs to stay. The other side of the same coin is that the higher emissions price taxes the information rents of firms that merely buy the right the pollute but are not close to moving. Together, these effects always increase the level of regulation; the optimal emissions price may even turn out to be higher than the Pigouvian first-best price.

The distortions in our mechanism stem from the regulator's urge to compensate firms differently depending on their abatement costs on one hand, and the inability to condition compensation on the unobservable characteristics of the firm on the other hand. Figure 1 illustrates how the emissions price can be used to target firms based on their unobservable costs. The regulator faces constant marginal damages (MD) from emissions, and the marginal cost (MC) curve represents marginal abatement costs, aggregated over a mass of small firms. High-cost firms pay the emissions price and face the cost denoted by area A, whereas low-

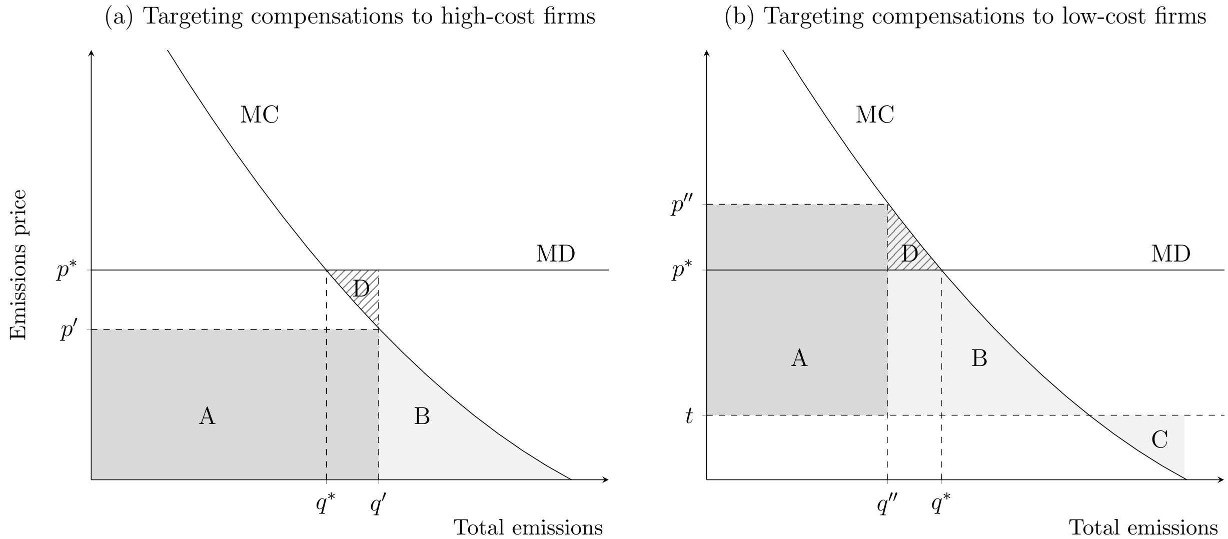

Figure 1: Illustration of the model and illustrations of second-best policies when the problem is leakage from (a) high-cost firms or (b) low-cost firms

cost firms choose to cut emissions facing costs in area B. Now, it is possible to target compensations to high-cost firms by choosing an emissions price that is lower than the Pigouvian price $p' < p^*$ as shown in Figure 1a; this reduces the cost for firms in area A without affecting the inframarginal firms in area B. However, deviating from the Pigouvian level comes at a price as it creates a deadweight loss (area D). On the contrary, compensations can be targeted to low-cost firms by choosing a higher emissions price $p'' > p^*$, accompanied by a respective lump-sum compensation t, as shown in Figure 1b. This would reduce the cost for firms in area B without affecting the costs of inframarginal firms in area A. This policy creates firms that *gain* from regulation (area C), and in addition, a deadweight loss (area D).

It is not *a priori* clear if more compensation should be paid to high- or low-cost firms. This will be endogenous in our model and depend on empirically testable characteristics of the industry. We begin by analyzing the case with no correlation between abatement and relocation costs in Section 3, which captures a general trade-off between two countervailing effects. On one hand, low-cost firms are the most valuable to keep in the regime, as they create the largest "climate surplus", avoided global externality minus local abatement costs, when cutting emissions. In fact, sufficiently high-cost firms do not cut emissions even if they stay and hence their potential relocation does not cause carbon leakage as such. This *climate* *surplus effect* makes the regulator inclined to target compensation to low-cost firms. On the other hand, the total cost burden of the regulation is the highest for high-cost firms, while low-cost firms may even gain from regulation (as indicated by area C in Figure 1b). This *abatement cost effect* makes the regulator want to target compensation to high-cost firms. We show that the abatement cost effect dominates when industries are relatively clean and the outside option is accurately known, leading to a downward distortion in regulatory stringency (Figure 1a). Conversely, the climate surplus effect dominates when industries are emission-intensive and there is a lot of uncertainty about firms' relocation costs, leading to an upward distortion in the level of regulation (Figure 1b).

For some industries, it may well be that abatement cost and relocation costs are correlated, although is not *a priori* clear if such a correlation exists and if the sign should be positive or negative.[4] In the analysis of perfect correlation (Section 4.1), we find that when there is a negative correlation between relocation and abatement costs, the high-cost firms are first to relocate. To compensate these high-cost firms, the regulator compromises on the level of regulation, following the logic presented in Figure 1a. With a strong positive correlation, low abatement costs are associated with the cheapest relocation cost and only the low-cost firms move. To target compensation to these low-cost firms, the regulator sets a higher level of regulation, in line with Figure 1b. Last, with a weak positive correlation, the medium-cost firms leave the regime and the net effect on the emissions price turns out to be ambiguous.

A natural extension is to consider a game between two uninformed local policy makers setting their environmental policies non-cooperatively (Section 4.2). The firm's relocation decision is now altered since the destination country may also have an environmental policy in place. We find that the information problem brings about a new strategic effect in this leakage game, absent in the strategic environmental policy literature. Intuitively, the firm's relocation becomes environmentally less harmful when it moves to a regime with regulations in place, which in turn reduces the need to compensate firms for their "climate surplus". One would expect that if firms have fewer options to escape regulations, the optimal emissions prices would go up. Yet, we find that the new strategic interaction distorts

[4] Ederington *et al.* (2005) consider U.S. manufacturing and trade data and find that the least "footloose" firms are the ones with the largest pollution abatement costs, suggesting a positive correlation between moving and abatement costs. In contrast, in a note written together with Ralf Martin we look at the firm level EU data from Martin *et al.* (2014) and find evidence for a negative correlation (Ahlvik *et al.*, 2017).

emissions prices downwards, below the first best, in a non-cooperative Nash-equilibrium between two symmetric regions.

In Section 5, we provide an illustrative quantification of the optimal carbon leakage policy for key sectors in the EU emissions trading program, based on the firm-level data from Martin *et al.* (2014). The data allow us to draw representative relocation risk distributions for five key sectors that together form 62 per cent of the total industry emissions covered by the EU ETS. With representative values for the social cost of carbon emissions and the social cost of public funds, we can quantify the optimal carbon leakage, distortions in the emissions price, and the fraction of the sectoral cost that is optimally covered from public funds. Two main results arise from our quantification: the optimal policies vary heavily across sectors and the size of the distortions are non-negligible. The optimal carbon leakage is $3-24$ per cent and the carbon prices are increased upwards by $21-32$ per cent compared to the benchmark without leakage. Total windfalls pocketed by the five key sectors under the optimal policy are 7.6 billion euros.

1.1 Related literature

Our study contributes to the literature on environmental regulation under privately informed polluters. It is well known that the mere existence of private information does not need to lead to distortions from the first best: absent budgetary concerns, the regulator can reach the socially efficient outcome by a Vickrey-Clarke-Groves -type mechanisms as in Dasgupta *et al.* (1980); see also Montero (2008). If firms cannot choose to opt out from regulation, the first-best can also be implemented through a Bayesian mechanism and any budgetary objectives can be satisfied by lump-sum transfers, following d'Aspremont and Gérard-Varet (1979, 1983). Technically, our model differs from these papers as public funds are valuable and firms' privately known opportunity to move introduces a participation constraint that prevents arbitrarily large lump-sum taxes on firms. It follows that the optimal second-best policy is distorted away from the first-best.

Some studies consider second-best environmental policies, where participation constraint is either based on zero-profit condition (Spulber, 1988; Kim and Chang, 1993) or voluntary participation of firms (Lewis, 1996; Montero, 2000) or countries (Helm and Wirl, 2014; Martimort and Sand-Zantman, 2015).[5] Our study builds on this literature and extend the

[5] Zero-profit condition ensures that all firms prefer to stay in the market. Voluntary participation means

earlier analyses in three ways. First, carbon leakage is fundamentally different from, say, bankruptcy as relocating firms will continue to produce global negative externalities. Second, unlike previous studies, we allow the use of exclusions as a screening device. Leakage plays a similar role than excluding consumers from using public goods (Hellwig, 2003; Norman, 2004) or preventing natural monopolies from serving the market (Baron and Myerson, 1982). As a third difference, the cited studies assume that the regulator knows the firms' outside option and can therefore offer just enough compensation to make the high-cost firms indifferent between exiting and staying. With leakage, however, the regulator cannot guarantee the indifference as the outside option is not observable. To incorporate this effect, we solve a self-selection model with random participation following Rochet and Stole (2002), where firms possess two-dimensional private information.[6] It follows that not just must the regulator leave positive information rents to some high-cost firms, but also some low-cost firms relocate.

There is a vast literature on how policies should be designed to take into account a threat of carbon leakage.[7] Given that firms' private information on the leakage propensity is indisputably an essential feature of the problem, it is rather surprising that it has received little attention. Greaker (2003) is among the first to introduce a model with one-dimensional asymmetric information about firms' ease of moving in the context of strategic environmental policy, but he leaves the formal analysis of a truth-revealing mechanism for future research. Harstad and Eskeland (2010) consider a signalling model where firms have private information about their need for permits, but there is no leakage in their model, an essen-

---

that firms should make no loss relative to the counterfactual situation. These two conditions are very similar in respect that, in the case of optimal policy, they only bind for high-cost firms who are left with no information rents. It follows that, to limit information rents, the regulator always distorts the level of the regulation downwards. This is in contrast to the present study, where the level of regulation is often distorted upwards.

[6]The upward distortion, one of the main results of this paper, is not a typical feature of previous models with random participation, such as Rochet and Stole (2002); however it arises naturally in our setting with the global externality problem.

[7]Carbon leakage, where the benefits of uncoordinated climate policies are offset by increased emissions in unregulated regimes, may refer to at least three distinguishable phenomena: (i) leakage through changes in fossil fuel prices (Harstad, 2012; Harstad and Liski, 2013; Böhringer *et al.*, 2014c), (ii) increased production in unregulated regimes via changes in output prices (Markusen, 1975; Hoel, 1996; Fischer and Fox, 2012; Meunier *et al.*, 2014; Böhringer *et al.*, 2014b) or (iii) firms' physically moving their production in countries with laxer environmental regulation (Markusen *et al.*, 1993; Motta and Thisse, 1994; Hoel, 1997; Ulph and Valentini, 1997; Schmidt and Heitzig, 2014; Martin *et al.*, 2014). This study focuses on the third category.

tial feature of the policy design problem. The study that has motivated us most is Martin *et al.* (2014), who construct firm-level measures for the relocation risk and then derive what they call the fundamental economic logic of industry compensation: "*[...] efficiency requires that payments be distributed across firms so as to equalize marginal relocation probabilities, weighted by the damage caused by relocation*". With private information, the optimal mechanism no longer follows this logic: the relocation propensities will be differentiated depending on observable actions or characteristics of the firms. Last, Meunier *et al.* (2016) consider how output-based allocations can be optimally used to manage leakage and volatility, with focus on uncertainty rather than on asymmetric information.

2 The set-up

Consider an industry with a unit mass of polluting firms, each firm characterized by a cost of reducing one unit of emissions $c \in [\underline{c}, \overline{c}]$ ($\underline{c} \geq 0$) and a cost of relocation $\theta \in \mathbb{R}$.[8] Both c and θ are private information to the firm. Regulator knows the distribution of firms' types. The distribution of abatement costs follows continuous density function $f(c)$, with $F(c)$ denoting the respective cumulative distribution. The conditional density and the cumulative distribution for relocation costs are $g(\theta|c)$ and $G(\theta|c)$, respectively. The mass of firms with cost c and migration cost less than θ is thus $G(\theta|c)f(c)$. We make the usual assumptions that the distributions are "regular":

Assumption 1. *(Monotone hazard rate assumption). For all c, θ,*

$$\frac{f(c)}{1 - F(c)} \text{ and } \frac{g(\theta|c)}{1 - G(\theta|c)}$$

are non-decreasing in c and θ, respectively.

[8] It is natural to think that $c \geq 0$, but technically there is nothing preventing firms with negative marginal abatement costs – for instance if firms postpone privately profitable abatement measures when expecting higher price for them in the future. The assumption of full support for θ guarantees that some firms are immobile and have very large relocation costs so that there will always be a mass of firms with abatement cost c staying in the regime. The cost of regulation may not be the only reason to relocate production, so there may be other privately known reasons affecting the outside option. These are all captured by θ. Notably, some firms may move even with no regulation in place; this is captured by these firms having a negative θ. The assumption of full support makes the forthcoming analysis technically simpler as we avoid points of non-differentiability at upper and lower bounds.

We begin the analysis without correlation between θ and c in Section 3, but to introduce the full model at one go, we use the general notation $G(\theta|c)$.

What we define as "firm" can be interpreted more broadly as a unit of production such that abatement costs are independently distributed across production units that can be relocated individually. A real-world company can therefore consist of several of these "firms". Leakage may be caused by either firms actively relocating existing production units to other regimes, or "investment leakage" when firms expand into another location for new production. Our model incorporates both kinds of leakage.

Firms' abatement cost c can be seen as a technical parameter related to the abatement process, and it is customary to treat it as private information (Dasgupta *et al.*, 1980; Montero, 2008); evidently, this is why market-based instruments are generally preferred over command-and-control measures. The relocation cost should be interpreted as the net effect of the physical cost of moving and the expected decrease in profits due to choosing another, supposedly less preferable location. The firm may not know the relocation cost accurately, but it is reasonable to assume that the firm has more possibilities to estimate this parameter than the regulator, who may even be uninformed of *where* the firm could potentially move to.

2.1 Firm's problem

An individual firm faces two decisions. First, it either stays in the regime and complies with the regulation, or relocates production to another location with no regulation in place. Second, conditional on staying, the firm chooses how much to cut emissions. Formally, $x = [0, 1]$ is the amount that a firm cuts its counterfactual emissions that are normalized to unity. A firm receiving monetary transfer t faces the following net cost of regulation:[9]

$$C(t, c, x) = cx - t. \tag{1}$$

The regulator has two ways to compensate the firm and reduce the net cost of regulation; it can either offer a higher transfer t, or alternatively lower the required emission reduction x. There is an important distinction between these two ways of compensation, however: while the former is independent of firms' abatement cost, the latter depends on it such that a decrease in x is more valuable for firms with high c. If the offered compensation is

[9]The case of general convex abatement cost function is analyzed in Section 4.3

insufficient, the firm may find it worthwhile to avoid the regulation and face the relocation cost. The condition for this to happen is:

$$C(t, c, x) > \theta. \tag{2}$$

It is critical that the transfer is conditional on firm's staying in the regime. Moreover, it is assumed that it is not possible to relocate a part of the production – the production units can only move in entirety or not at all. After a unit moves, the regulator loses the possibility to regulate it or collect lump-sum transfers.

The problem is essentially one of two-dimensional private information. To carry on with the analysis, we follow the techniques developed by Rochet and Stole (2002) and focus on non-stochastic mechanisms of the form $\{T(\hat{c}), X(\hat{c})\}$: we consider a direct revelation mechanism where firms announce their abatement cost $\hat{c}$, taking the decision to stay as given. For any given mechanism proposed by the regulator we then calculate the net cost of regulation: $cX(\hat{c}) - T(\hat{c})$, and solve for the mass of firms that have high enough relocation costs to stay in the regime. In other words, we model firms' relocation decision as an indirect mechanism where no report on relocation cost $\hat{\theta}$ is made. This is without loss of generality, given the assumption that firms cannot move partially, and the restriction to a deterministic mechanism.[10] In incentive-compatible mechanisms firms find it optimal to truthfully announce their abatement costs c:

$$c = \arg\min_{\hat{c}} \left\{ cX(\hat{c}) - T(\hat{c}) \right\} \text{ for all } c. \tag{3}$$

Let $C(c)$ be the indirect net cost function of the firm c who truthfully reports his type. A firm stays in the regime if its net cost, when truthful, is lower than the relocation cost: $C(c) \leq \theta$. The mass of firms of type c for which this holds, as a function of $C(c)$ and c, is:

[10] In a stochastic mechanism, firms would report both their abatement and relocation costs, after which the regulator would arrange a lottery and randomly exclude a share of firms, using the probability of exclusions to screen for both c and θ. Acknowledging that such a stochastic mechanism can be more efficient than the deterministic one, we follow the literature of random participation (Rochet and Stole, 2002; Lehmann *et al.*, 2014) and focus on the deterministic mechanism that have a simple market interpretation. A stochastic mechanism could be implemented, for example, by a scheme of random audits with the probability of being audited depending on the reports. However, such a mechanism may suffer from the known problems related to verifying randomization, see Laffont and Martimort (2002).

$$\phi(C(c),c) = \Big(1 - G(C(c)|c)\Big) f(c). \tag{4}$$

The mass of staying firms $\phi(C(c),c)$ is decreasing in $C(c)$ reaching zero as the net cost approaches to infinity. We use $\phi'(C(c),c)$ as shorthand for $d\phi(C(c),c)/dC(c)$. We also define the inverse hazard rate (over C):

$$\eta(C(c)) = \frac{\phi(C(c),c)}{\phi'(C(c),c)}. \tag{5}$$

2.2 Regulator's problem

The social planner designs a regulation scheme for the industry by choosing the level of compensation ($T(c)$) and the level of regulation ($X(c)$) for firms of type c who stay in the regime, so that the incentive compatibility condition is satisfied and the social welfare is maximized:

$$\max_{X(c),T(c)} \int_{\underline{c}}^{\overline{c}} \Big(\gamma - cX(c) - \lambda T(c)\Big)\phi(C(c),c) - D\Big(1 - \phi(C(c),c)X(c)\Big) dc \tag{6}$$

subject to equations (3) and (4) holding. Planner's objective function (6) consists of four components. First, the government gets some benefit γ when a firm stays in the regime; this parameter captures the firm's contribution to social surplus through profits, employment impacts, or any possible direct political benefits from having the firm.[11] Second, the planner considers the cost of abatement, which for firm of type c, reducing $X(c)$ units of emissions, is simply $cX(c)$. Third, the revenue collected by the mechanism is socially valuable, for instance, because they can be used to decrease distortionary taxes elsewhere. The social cost of compensating a firm is $\lambda T(c)$ where parameter λ denotes the social cost of public funds. Last, the government considers damages due to pollution, with marginal damages assumed to be linear, D.[12] The location of global emissions is irrelevant and therefore

[11] The value of γ is industry-specific, observable and does not depend on firm's type; it could be excluded from the analysis with minor substantial changes. However, it allows describing pollution intensive industries as those with low value-added per unit of emissions (low γ), and therefore variations in γ will help in analyzing how the optimal mechanism varies with observable characteristics across industries. Also, γ will be a key parameter in our numerical calibration in Section 5.

[12] For climate change, a constant marginal damage is not a poor approximation. First, emissions from a single jurisdiction are likely to have a small impact on the slope of the global damage curve. Second, the

pollution damages occur either when a share of firms $1-\phi(C(c),c)$ relocates, or when firms stay but do not cut emissions $\phi(C(c),c)(1-X(c))$. The last term of equation (6) is the sum of these two components.

It is immediately clear that in our model the firms' relocation decisions are not equally harmful; the regulator is more concerned about leakage by firms who could cut emissions with low abatement cost. To see this, we define the *net losses from relocation* as the decrease in social welfare when a firm of type c leaves:

$$\Delta(X(c),T(c),c) = -\Big(\gamma + (D-c)X(c) - \lambda T(c)\Big) \tag{7}$$

Let us denote this term $\Delta(c)$ in short. When firms do not cut emissions ($X(c)=0$) their relocation does not cause emission leakage as such, for they would pollute the same amount in either regime. In that case the regulator only dislikes relocation because of the fixed component γ and because these firms contribute to the public funds via a negative $T(c)$, for instance when they pay the emissions tax or buy emissions permits in an auction. Emission leakage only occurs when a firm that cuts emissions ($X(c)>0$) moves, and then climate surplus $(D-c)X(c)$ will be lost. The low-cost emission reductions create the largest climate surplus and, other things equal, the regulator suffers the most when the firms with lowest abatement costs move. The regulator is keen to keep the low-cost firms in the regime but cannot directly pay compensation to those firms unless abatement costs are observable. This will be the key driver behind the distortions in our leakage mechanism.[13]

3 The leakage mechanism

There are two main aspects of the policy we are interested in. The first is the *level of regulation*, that is, the total amount of required pollution reduction in sectors under the

global expected damage changes only slowly over time, and therefore it will be a matter of decades before the social cost can be seen to change as a function of emissions; the marginal damage can change in the near term only if severe climate impacts become material (see Gerlagh and Liski 2016). Third, constant damages can well approximate the predictions of the comprehensive climate-economy models (Golosov *et al.*, 2014; van den Bijgaart *et al.*, 2016). We will extend the analysis to the more general case where the damages from emissions may be convex in Section 4.3.

[13]This is a difference to the model developed by Rochet and Stole (2002) where, conditional on trade taking place at a given price, the monopolist does not care about customers' privately known valuation parameter. Here, the regulator directly cares about participating firms' type-dependent surplus as seen in equation (7).

leakage risk. The second aspect is the *level of compensation*: how much of the private cost of regulation should be covered from the public funds? Intuitively, we can think that the regulator sells two "tickets", one for the right to stay in the regime and another for the right to pollute if the firm chooses to stay. The tickets offered may vary across firms having different observable characteristics. The right-to-pollute ticket implements the level of the regulation and the right-to-stay gives the level of compensation (which can also be negative). We first consider a situation where these tickets can be conditioned on abatement cost c. Also, throughout section 3 we focus on the case without correlation between c and θ.

3.1 Benchmark: c public, θ private information

We begin by considering a benchmark where the regulator observes abatement cost c of each firm, and relocation cost θ is the only parameter that is private information to the firms. We can think that the regulator must offer a single contract for pools of observationally identical (c dimension) but in fact heterogeneous firms (θ dimension), and define industry-specific contracts this way. Although perhaps unrealistic, this case is useful for the more comprehensive mechanism where both c and θ are private information. Technically, the regulator maximizes social welfare (6) but without incentive compatibility constraint (3).

Proposition 1. *When c is public information and θ is firms' private information, the optimal policy $(X_B(c), T_B(c))$ takes the following form:*

$$X_B(c) = \begin{cases} 1 & \textit{if } c \leq c_B^* = \frac{D}{1+\lambda} \\ 0 & \textit{otherwise} \end{cases} \qquad \text{(intensive margin)}$$

$$\Delta(c)\phi'(C_B(c), c) - \lambda\phi(C_B(c), c) = 0 \qquad \text{(extensive margin)}$$

where $C_B(c) = cX_B(c) - T_B(c)$.

The first part of Proposition 1 is the (modified) Pigouvian rule for pricing externalities: firms cut emissions when their marginal abatement cost is lower than marginal damage from pollution, when expressed in terms of public funds $D/(1+\lambda)$.[14] Where revenues are not valuable to the regulator ($\lambda = 0$) the rule collapses to the famous Pigouvian principle,

[14]This is similar to Laffont and Tirole (1996). But the distortions in their paper are different than those in our analysis since they consider government monopoly pricing of permits using a single instrument (the price), rather than the optimal mechanism.

setting the marginal abatement cost threshold equal to the marginal damages from pollution, $c_B^* = D$. Where $\lambda > 0$, firms that cut emissions pay lower taxes, implying a loss of tax revenues that are valued as such. Private resources are spent on reductions to avoid emissions taxes and therefore they are costly in terms of public funds, diminishing the effective weight put on marginal damages. This first part of the Proposition gives the intensive margin of the regulation, that is, it tells us how much is required from the firms that stay in the regime.

The second part of Proposition 1 gives the extensive margin of the regulation in the sense that it determines the types to be excluded from the program – the optimal leakage. The condition represents a well-known trade-off between efficiency (leakage) and rent-extraction (overcompensation). Consider a marginal increase in compensation for type c: (i) it reduces surplus losses $\Delta(c)$ from marginal firms leaving the regime with mass $\phi'(C_B(c), c)$ and (ii) it increases the overcompensation paid to all the remaining firms of type c (with mass $\phi(C_B(c), c)$). Some carbon leakage is part of the optimal second-best policy, although it is clearly inefficient given that under perfect information, any excluded firm could simply be exempted from regulation – these exempted firms would pollute the unregulated amount but the regulator would not have to suffer the loss of γ. Firms' private information about θ therefore leads to inefficiencies in the extensive margin in this benchmark case with known c. The benchmark compensation that implements these distortions is characterized as:

Corollary 1. *When c is observable:*
(i) For $c > \frac{D}{1+\lambda}$, the optimal compensation schedule is flat,

$$T_B'(c) = 0.$$

(ii) For $c = \frac{D}{1+\lambda}$, the optimal compensation jumps discontinuously

$$\lim_{c \to D/(1+\lambda)^-} T_B(c) - \lim_{c \to D/(1+\lambda)^+} T_B(c) = c.$$

(iii) For $c < \frac{D}{1+\lambda}$, the optimal compensation satisfies

$$T_B'(c) = \begin{cases} > 0 & \textit{iff } \lambda\eta'(C_B(c)) > 1 \\ < 0 & \textit{iff } \lambda\eta'(C_B(c)) < 1 \end{cases}$$

Part (i) of the Corollary tells us that all firms that do not cut emissions are paid the same compensation regardless of their abatement cost. For firms that do not cut emissions, their abatement cost becomes irrelevant both in terms of social welfare and their private cost of regulation. We call this the *base compensation level*. To understand part (ii), consider a marginal firm with costs $c = D/(1+\lambda)$ so that the regulator is indifferent between this firm cutting emissions and polluting. Because of this indifference, the firm is guaranteed the same cost of regulation $C(c)$ irrespective of the choice of $X(c)$ and it follows that the transfer has to jump by $D/(1+\lambda)$, to exactly cover the cost of action $X(c) = 1$. Part (iii) of the Corollary 1 shows how the optimal benchmark compensation varies with abatement cost for firms that reduce emissions. Should more or less be paid to firms with the lowest abatement cost? There are two countervailing effects. On one hand, as shown in equation (7), the regulator wants the low-cost firms to stay and reduce emissions, because they create the largest climate surplus: this surplus grows with slope 1 when c marginally decreases. We call this the *climate surplus effect*. On the other hand, the regulation is cheaper for the low-cost firms as they can cut emissions with low costs. As the private net cost of regulation decreases linearly in c when a firm cuts emissions, the low-cost firms need a smaller compensation to stay; see equation (1). This reduced transfer in turn changes the social surplus by $\lambda\eta'(C(c))$. This is what we call the *abatement cost effect*. Whether the low-cost firms are paid the most then ultimately depends on which of these two effects dominates, as shown by the condition for the slope of the schedule in part (iii) of Corollary 1.

3.2 Complete mechanism: c and θ private information

When both abatement and moving costs are firms' private information, the transfers can no longer be made conditional on the observable characteristics (c dimension). They can only be made conditional on firms' observable actions.

Proposition 2. *When both c and θ are firms' private information, the optimal policy $(X(c), T(c))$ takes the following form:*

$$X(c) = \begin{cases} 1 & \text{if } c \leq c^* = \frac{D}{1+\lambda} + \frac{\mu(c^*)}{(1+\lambda)\phi(C(c^*), c^*)} \\ 0 & \text{otherwise} \end{cases} \qquad \text{(intensive margin)}$$

where

$$\mu(c^*) = \int_{\underline{c}}^{c^*} \Big[\Delta(c)\phi'(C(c), c) - \lambda\phi(C(c), c)\Big] dc.$$

Compensation to firms that do not cut emissions ($X(c) = 0, c > c^*$) is $T^* = T(c)$ and given *by*

$$\Delta(c^*)\phi'(-T^*,c^*) - \lambda\phi(-T^*,c^*) = -\frac{f(c^*)}{1-F(c^*)}\mu(c^*), \quad \text{(extensive margin)}$$

while firms that cut ($c \leq c^*$) receive $T(c) = T^* + c^*$.

As shown in the Appendix, the policy problem can be formally solved by choosing two prices: a flat baseline compensation level T^*, and a top-up for those who cut $T(c) = T^* + c^*$. In contrast to Proposition 1, there are now two types of distortions as both the intensive and extensive margins will differ from the benchmark with observable abatement costs.

To understand the source of the distortions, consider a small increase in the intensive margin c^*. This change makes firms around c^* to cut emissions but also, due to the incentive compatibility condition, leads to higher transfers to all the firms with c lower than c^*. The net welfare effect of this increase, integrated over all the types that cut emissions, is captured by the distortion term $\mu(c^*)$. Next, notice that keeping the intensive margin c^* fixed, any base transfer T^* paid to high-cost firms that do not cut emissions must also be paid to the low-cost firms – the welfare effect of this, it turns out, is captured by the same distortion term $\mu(c^*)$. For a positive $\mu(c^*)$ the regulator values the extra compensation paid to low-cost firms and consequently chooses a stricter level of regulation (higher c^*) and a higher base level of compensation (higher $T(c)$ for $c > c^*$) than in the benchmark. For a negative $\mu(c^*)$ the regulator dislikes the extra compensation to low-cost firms, and sets a lower level of regulation (lower c^*) and a lower base compensation level (lower $T(c)$ for $c > c^*$).

Sectors with no leakage. To illustrate, let us look at the special case where firms are immobile and will stay in the regime regardless the net cost of regulation they face.[15] Now the only relevant piece of private information is the firm's abatement cost. This is an extreme case of Proposition (2) where $\phi'(C(c),c) = 0$ and $\phi(C(c),c) = f(c)$, so that the distortion term becomes:

$$\mu(c^*) = \int_{\underline{c}}^{c^*} -\lambda f(\tilde{c})d\tilde{c} < 0 \tag{8}$$

It can immediately be seen that with no leakage, the level of regulation is thus always distorted downwards. Solving the integral in (8) and plugging it in the optimal policy we get:

[15] We can interpret the situation as one where the firms' outside option becomes publicly known and thus the firms can be made indifferent between staying and leaving.

$$c^* = \frac{D}{1+\lambda} - \frac{\lambda}{1+\lambda}\frac{F(c^*)}{f(c^*)}. \tag{9}$$

This result is familiar from earlier studies on optimal second-best environmental policies, see for instance Lewis (1996): only the low-cost firms receive information rents, and a lower emissions price offers a way to reduce those rents at the cost of efficiency (see also Baron and Myerson 1982). By contrast, when firms can relocate and the relocation cost is their private information, a mass of socially valuable firms with low abatement costs will also have low moving costs and will move. Therefore the regulator is now more constrained to extract their rents and will optimally increase the emissions price to encourage the low-cost firms' participation. In addition, and in contrast to the model without leakage, a mass of firms with high abatement cost will also receive information rents – those which also have high relocation costs. The higher price serves to extract a fraction of these rents.

Corollary 2. *All else equal, industries with carbon leakage are always assigned a stricter level of regulation (higher emissions price) than industries without leakage.*

This follows from comparing the distortion term in Proposition 2 with $\Delta(c)\phi'(C(c)) > 0$ and where some firms relocate $\phi(C(c), c) < f(c)$ to that presented in equation (8).

Comparison to benchmark with observable c. Whether the optimal second-best policy entails upward ($\mu(c^*) > 0$) or downward ($\mu(c^*) < 0$) distortions is endogenous in the model and depends on the characteristics of the industry. To be more precise, it will depend on the interplay of the previously introduced climate surplus and abatement cost effects: on one hand, the low-cost firms are the most valuable to keep in the regime for their emission reductions but, on the other hand, these firms have low cost of compliance and are therefore less likely to relocate. To better grasp this result, we make a link between the case of unobservable c and the benchmark case with known c.

Proposition 3. *The sign of distortions, term $\mu(c^*)$, is positive (negative) if climate surplus effect (abatement cost effect) dominates:*

$$\lambda\eta'(C_B(c)) < (>) 1 \textit{ for all } c < c^*.$$

The proposition tells us whether the level of regulation is higher or lower than the Pigouvian

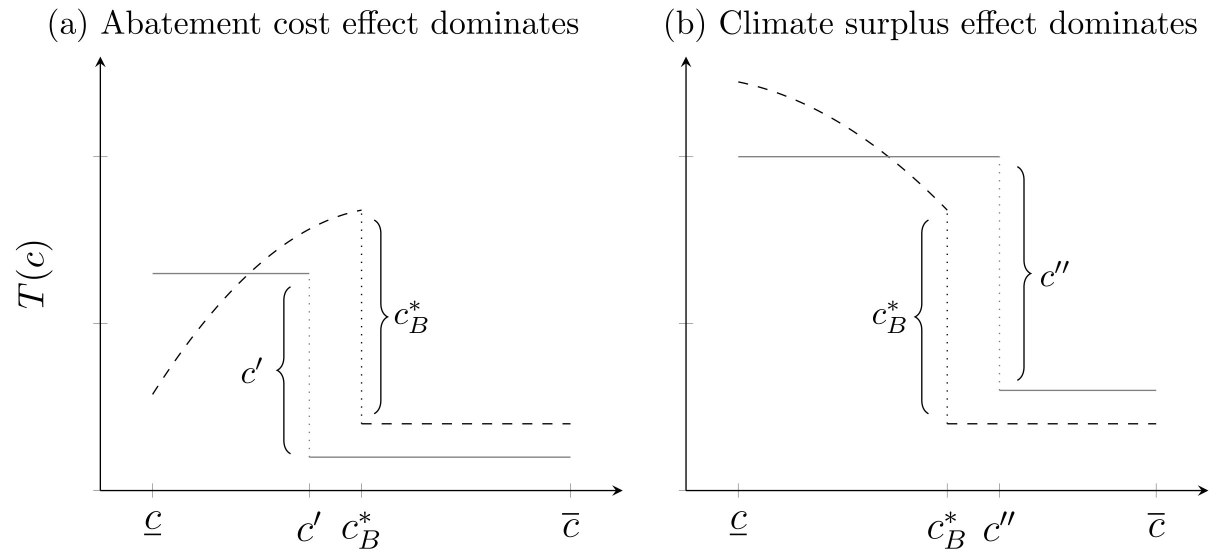

Figure 2: Optimal compensation in the benchmark (dashed line) and in the mechanism (solid line) for the two cases: (a) abatement cost effect dominates, (b) climate surplus effect dominates.

benchmark, $c_B^* = D/(1+\lambda)$.[16] The two cases are illustrated in Figure 2. The benchmark compensation, characterized in Corollary 1 and depicted as dashed line in the Figure, can be conditioned on firm's abatement cost as c is observable. But when c is not observable, the regulator is restricted to offer the same contract for all firms with the same action – the contract offered is chosen to minimize the deviations from the benchmark, weighted by the mass of type c firms. If the abatement cost effect dominates ($\lambda\eta'(C_B(c)) > 1$) as in Figure 2a, the benchmark compensation is increasing in $c < c^*$ (see Corollary 1) and the regulator wants to limit compensations to the low-cost firms in the benchmark. The optimal contract seeks to achieve this same by distorting the intensive margin, that is, the emissions price downwards ($\mu(c^*) < 0$). In contrast, if the climate surplus effect dominates ($\lambda\eta'(C_B(c)) < 1$), the benchmark compensation is decreasing in $c < c^*$ and the regulator would like to target compensation to the low-cost firms and tax the high-cost firms. In the optimal mechanism this is achieved by distorting the intensive margin upwards ($\mu(c^*) > 0$).[17] Intuitively, by

[16] Clearly, the condition in Proposition 3 is sufficient but not necessary since the benchmark schedule can be nonmonotonic. We provide further comparative statics results in the next Section.

[17] The appearance of upward distortion is not standard in models of random participation, but it stems

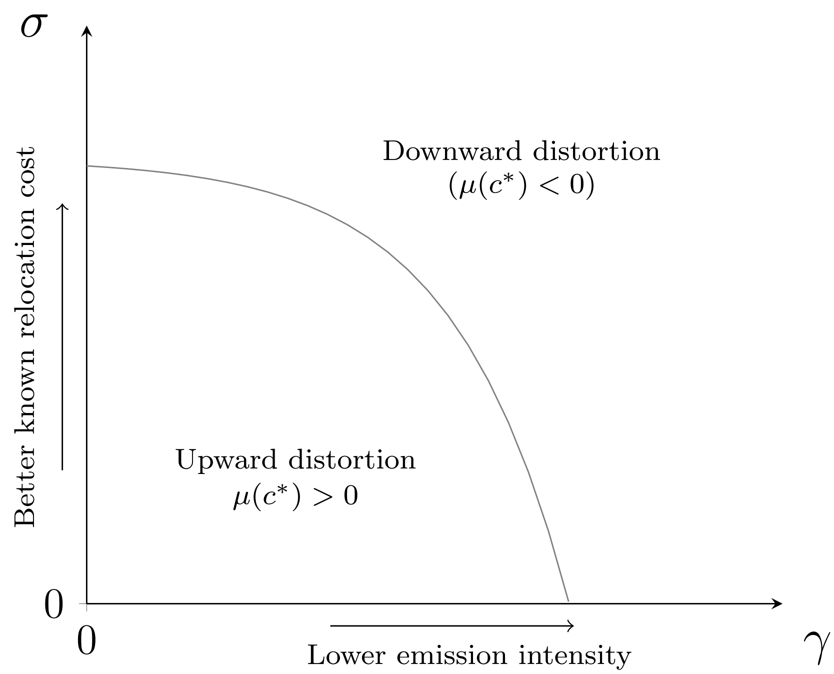

Figure 3: Sign of the distortion as a function of pollution intensity and relocation cost

setting a high price for emissions, the regulator taxes those firms that will not move in any case (high θ) and for whom abatement is not an option (high c). The combination of a high baseline T^* together with a high price on emissions thus implements a tax-subsidy scheme where firms self-select to be taxed or subsidized, depending on their abatement costs.

Comparative statics. For a more complete grip on the factors that determine how different industries should be treated, we carry out a sensitivity analysis with respect to two key industry-specific characteristics. We make an additional assumption for the distribution describing the outside options, namely that the hazard rate $g(\theta|c)/(1 - G(\theta|c))$ in Assumption 1 is not just increasing, but also convex in θ for all c. A corresponding assumption has been exploited before (Rochet and Stole, 2002), and it holds for many commonly used distributions such as the normal distribution. First, we look at how the distribution of outside

from our climate surplus effect, or more exactly, from the fact that firms' privately known type c appears directly in regulator's welfare function (6). This is not true in Rochet and Stole (2002) and, in their setting, assuming a non-decreasing inverse hazard rate (or $\eta'(\cdot) \geq 0$) is sufficient to eliminate any upward distortions. In our model the limiting case for upward distortions is $\eta'(\cdot) \geq 1$ (see Proposition 3) – where the right hand side (equal to 1) is exactly the climate-surplus effect.

options shapes the optimal mechanism. To this end, we introduce another parameter $\sigma > 0$, which captures the degree of uncertainty about the industry's relocation costs. Taking θ as a zero-mean risk, we can consider a mean-preserving spread around a positive mean $\hat{\theta}$ controlled by the precision parameter σ, so that the condition for firm's leaving becomes:

$$C(c) > \frac{\theta}{\sigma} + \hat{\theta}.$$

Hence $\sigma \to \infty$ captures the case where the outside option is perfectly known; smaller values of σ denote more uncertainty about the outside option. Second, we carry out sensitivity analysis with respect to γ. For very pollution-intensive industries, the value-added per unit of pollution, γ, becomes small and the compensation is mainly motivated by damages caused by emissions. For industries that pollute a little, γ is relatively large.

Proposition 4. *All else equal, for industries with lower pollution intensity (higher γ) or more uncertain relocation cost (higher σ):*

(i) *When c is an observable characteristics, the compensation of costs is greater, $\partial T_B(c)/\partial\gamma > 0$ for all c, and the sign of $\partial T_B(c)/\partial\sigma$ is ambiguous.*

(ii) *When c is private information and $\eta''(x) < 0$, the distortion term, $\mu(c^*)$, is strictly lower, $\partial\mu(c^*)/\partial\gamma < 0$ and $\partial\mu(c^*)/\partial\sigma < 0$.*

Broadly speaking, there are two reasons in the model why the regulator dislikes firms' relocation: the loss of fixed-value component given by γ and the loss of climate surplus $(D-c)X(c)$ that depends negatively on the abatement costs. For pollution intensive industries the latter effect becomes relatively very valuable, which calls for compensations to low-cost firms by distorting emissions price upwards. Conversely, for firms that have high value per pollution unit (high γ), the latter effect becomes small and the incentive to distort emissions price upwards disappears (see Figure 3).

When the outside option becomes accurately known (large σ), we approach the model of Lewis (1996) where the firms with high abatement costs can be made indifferent between staying and moving, leaving them with no information rents. Choosing a lower emissions price is a means to extract information rents from the low-cost firms. For very inaccurately known outside options (small σ), on the other hand, the mass of "escaping" low-abatement cost firms increases, so they must be compensated by lump-sums and tighter regulation, that is, the value of their contribution to reductions must be increased.

4 Extensions

4.1 Correlated private information

So far we have assumed no correlation between the relocation and abatement costs; it is not *a priori* clear which way the correlation should go in a given industry. For instance, in some industries the firms with the lowest location-specificity of productivity may also have the cheapest abatement options. Then, firms with low c can be the most prone to move, that is, they are "footloose and pollution-free".[18] In some other industries, the ease of relocation may increase together with abatement costs.[19] The correlation is an empirical question, and potentially important since, it turns out, it links closely to the pattern of compensations and distortions.[20] The general reasons for the distortions outlined in our main analysis do not change with correlation, but the analysis with correlation helps to identify new circumstances where downward and upward distortions arise.

To introduce correlation to the model in a tractable way, we follow the literature on type-dependent participation constraints (Lewis and Sappington, 1989; Maggi and Rodriguez-Clare, 1995; Jullien, 2000) and analyze a perfect correlation between θ and c:

$$\theta = b + kc \tag{10}$$

Here b is a scaling factor and k is publicly known parameter depicting the sign of the correlation such that for $k > 0$ the lowest cost firms face the lowest relocation cost and $k < 0$ is the opposite case. Firms still possess private information about both their abatement and relocation costs, but full correlation renders this information one-dimensional: truthful reporting of the firm's abatement cost immediately reveals also the outside option to the regulator. For ease of exposition, let us adopt the convention that the firm's type is c. Using

[18]This is the title of Ederington *et al.* (2005) who find, using U.S. manufacturing and trade data, that the firms with largest pollution abatement costs also tend to be the least geographically mobile.

[19]Levinson and Taylor (2008) find that pollution abatement costs, despite being small fraction of value-added, have an economically significant impact on U.S. trade volume with Canada and Mexico. While not direct evidence, the result is consistent with a negative correlation between mobility and abatement costs.

[20]We looked at the question of correlation empirically using the data from Martin *et al.* (2014) in a note written together with Ralf Martin. We find support for a negative correlation between the propensity to move and a measure of abatement costs, contradicting the positive correlation found in Ederington *et al.* (2005).

equation (10), the condition for leaving, equation (2), can be rewritten as:

$$cX(c) - T(c) > b + kc. \tag{11}$$

We can distinguish three situations, which turn out to be substantially different. First, if the cost of moving and abatement have a negative association $k < 0$, high-type firms (high c) have the lowest cost of moving. In that case, the firm always wants to overstate its type c as this way it can exaggerate both the ease of moving and cost of compliance. In this sense, the incentives to lie in both dimensions are *concurring*. Second, with strong positive correlation $k > 1$, the two types of costs are positively associated. The low-type firms still have the lowest cost of cutting emissions but they can also easily move: the firm cannot simultaneously overstate compliance costs and the ease of moving. However, since the correlation is strong, the latter effect dominates and despite these *conflicting incentives*, the firm always has an incentive to understate its type c. Third, with weak positive correlation $0 < k < 1$, the firm faces *countervailing incentives*: it wants to understate its type to emphasize the willingness to leave if it is not required to cut emissions ($X(c) = 0$), and overstate its type to emphasize the high compliance costs when required to cut ($X(c) = 1$).

When the regulator cares sufficiently about the public funds (λ high enough), it is always optimal to exclude some firms from the regime to save on compensations. Similarly, if externality cost D is sufficiently large, it will be optimal to require cuts at least from some firms. These two conditions ensure that the mechanism implements an interior outcome as follows:[21]

Proposition 5. *The (interior) optimal mechanism under perfect correlation.*
(i) Concurring incentives, $k < 0$***:*** *firms* $c \leq c^*$ *cut* ($X(c) = 1$) *where*

$$c^* = \frac{D}{1+\lambda} - \frac{\lambda}{1+\lambda}\frac{F(c^*)}{f(c^*)},$$

and firms with $c > c' = -(T^* + b)/k$ *leave where*

$$\Delta(c')f(c') = k\lambda F(c'),$$

[21] This is a technical difference to the basic model: with full correlation, the finite support of c determines the support for θ by equation (10) which can no longer be infinite, in contrast to what we assumed in Section 2. To facilitate comparison with the results of the basic model, we only focus on the interior solution where some firms cut emissions and some are excluded from the regime.

and $c' > c^*$. All staying firms ($c \leq c'$) receive T^*, but $T^* + c^*$ if they also cut ($c \leq c^*$).
(ii) Conflicting incentives, $k > 1$***:*** *firms $c \leq c^*$ cut,*

$$c^* = \frac{D}{1+\lambda} + \frac{\lambda}{1+\lambda}\frac{1-F(c^*)}{f(c^*)},$$

but leave when $c < c' = (T^* + b + c^*)/(1-k)$ where

$$\Delta(c')f(c') = (1-k)\lambda\Big(1 - F(c')\Big),$$

and $c' < c^*$. Staying firms ($c' \leq c$) receive T^*, but $T^* + c^*$ if they also cut ($c' \leq c \leq c^*$).
(iii) Countervailing incentives, $0 < k < 1$***:*** *firms leave when $c^* < c < c'$ where*

$$\Delta(c^*)f(c^*) = (k-1)\lambda F(c^*)$$
$$\Delta(c')f(c') = -k\lambda\Big(1 - F(c')\Big)$$

and $c^* = (T^* + b + c^*)/(1-k)$ and $c' = -(T' + b)/k$. Firms $c \leq c^*$ cut and receive $T^* + c^*$ while staying firms $c' \leq c$ pollute and receive T'.

The same principles that we saw in Proposition 2 for the two-dimensional case with no correlation between the parameters apply here as well. First, it is optimal to distort the intensive margin, that is, to deviate from the Pigouvian rule, to manage the rents of those who stay in the regime. Figure 4 helps to understand the elementary differences between the three reported cases. For *concurring incentives,* $k < 0$, the level of regulation is distorted downwards, $c^* < D/(1+\lambda)$, because this limits the rents of the low types and allows for targeted compensation to those with high types, following the logic in Figure 1a. Second, the extensive margin strikes a balance between relocation loss by marginal firms and over-compensating the remaining firms. For *concurring incentives*, the marginal excluded firm is a polluter (see Figure 4) and therefore there is no pollution leakage. However, there is a leakage of economic value-added and thus condition $\Delta(c')f(c') = k\lambda F(c')$ is a trade-off between pure location-specific economic losses and inframarginal compensations paid to the remaining firms.

For *conflicting incentives,* $k > 1$, the intensive margin is distorted upwards, $c^* > D/(1+\lambda)$. Quite intuitively, by setting a high price for emissions, the regulator wants to tax the high cost firms who pollute and stay in the regime but are nowhere close to leaving (see Figure 4); simultaneously, cutting firms create a large climate surplus and receive a generous compensation. This outcome conforms well with out analysis of the general case with two-dimensional types. The extensive margin decision, $\Delta(c')f(c') = (1-k)\lambda\big(1-F(c')\big)$, is now a trade-off including pollution damages, in addition to location-specific economic losses, since the marginal relocating firm in fact stops reducing emissions when it moves.

The case of *countervailing incentives* $0 < k < 1$ illustrates surprising outcomes that may arise under more general correlation structures. An interval of intermediate-cost firms relocate and there are no firms that would be indifferent between cutting and not cutting emissions. It follows that the optimal emissions price is not unique – any price between the costs determined by the two extensive margins in part *(iii)* of Proposition 5 leads to the desired outcome. Low-cost firms are socially valuable to include for their high climate surplus, and high-cost firms for their economic value-added and contribution to the public funds.

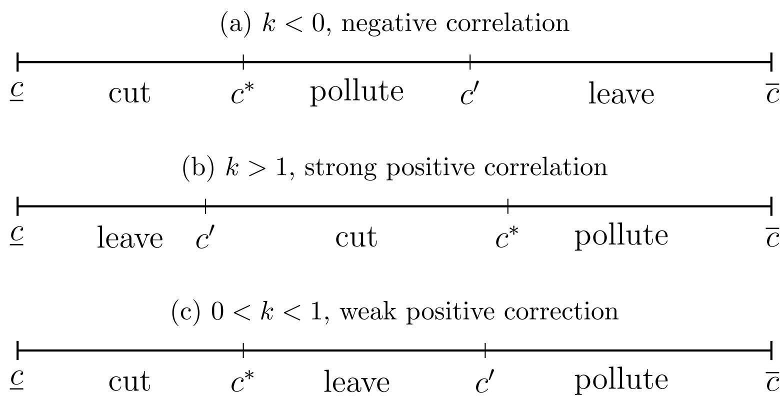

Figure 4: Optimal policies when c and θ are perfectly correlated.

4.2 The leakage game

We have assumed that firms move to a "pollution haven" without environmental policies in place. This does not have to be the case: the most likely destination regime may be one with equal stringency or at least some level of regulation.[22] We explore this possibility by analyz-

[22]This observation has been one of the reasons presented for why the "pollution haven effect" has proven difficult to demonstrate empirically: most trade takes place among industrialized countries sharing similarly high levels of environmental stringency (Ederington *et al.*, 2005).

ing the non-cooperative equilibrium environmental policies between two regimes.[23] Notably, our set-up with constant marginal externality cost guarantees that there is no traditional strategic common-pool interaction: if leakage is assumed away, policies in jurisdictions are independent. Intuitively, one region cannot manipulate the marginal damage faced by the other through its strategic choice of emissions (see, for example, Harstad and Liski 2013). However, the possibility of leakage together with asymmetric information about costs, it turns out, brings about another subtle strategic effect. We consider a Nash equilibrium between two countries, $n = \{i, j\}$, simultaneously setting compensation $T_n(c)$, emission reductions $X_n(c)$ and thus the total cost of regulation $C_n(c) = cX_n(c) - T_n(c)$ for each type c in their respective regimes. The Nash equilibrium is best described using our main formulation where there is no correlation between c and θ.[24] We also continue limiting attention to non-stochastic mechanisms. Policy makers in i and j face observationally similar firms; in particular, the firm type distributions are the same in the two locations. Firms first observe the policies in place, and then can choose with which mechanism to comply, that is, a firm in location i either reports to $(X_i(c), T_i(c))$ or pays the moving cost and reports to $(X_j(c), T_j(c))$. The relocation condition for firm c in i, now taking into account also the policies in the destination country, writes as:

$$cX_i(c) - T_i(c) \geq \theta + cX_j(c) - T_j(c) \tag{12}$$

A mass of firms that move depends now on the difference in the net cost between the two regimes, $\phi(C_i(c) - C_j(c))$. Firms may relocate even if the two regimes have identical carbon prices in place, as firms face lower cost of regulation in regimes with larger lump-sum compensations – such as emission trading schemes with more generous free permit allocations. Policies in the other regime also affect the net losses from relocation, and equation (7) for regime A can be rewritten as:

$$\Delta(X_i(c), X_j(c), T_i(c), c) = -\Big(\gamma + D\Big(X_i(c) - X_j(c)\Big) - cX_i(c) - \lambda T_i(c)\Big) \tag{13}$$

[23] We assume that only firms are privately informed, and both regimes share the same information regardless of firms' initial location. This is in contrast to studies by Helm and Wirl (2014) and Martimort and Sand-Zantman (2015) where the focus is on countries' private information. Our approach is conceptually closer to Lehmann *et al.* (2014) who develop the optimal Mirrlees income tax schedule in a game between two countries; yet our results are very different because of the global externality problem.

[24] As in the single country case, the correlation can shape the precise characterization but not the general lesson that, due to the strategic setting, one reason for distortions in the equilibrium mechanism is eliminated.

There are some key differences to the benchmark model. With $X_i(c) = X_j(c)$ there is no carbon leakage when a firm of type c moves. In fact, with $X_i(c) < X_j(c)$ firms' relocation causes *negative leakage* and regime i may even benefit from relocation.[25] We solve for the symmetric Nash equilibrium, where strategies are best responses to each others.[26]

Proposition 6. *For any given $G(\theta)$ and $F(c)$ satisfying our assumptions, in a symmetric non-cooperative equilibrium: (i) there is no net leakage, and (ii) both regimes distort the level of regulation by setting an emissions price below the Pigouvian level $c_n^* < D/(1+\lambda)$ for $n = \{i, j\}$.*

In the symmetric equilibrium firms face the same net cost in each location, $C_i(c) = C_j(c)$, and there is no leakage of pollution in the equilibrium.[27] Yet, the threat of off-equilibrium leakage justifies positive compensations. We see that the risk of leakage makes regional regulations strategic substitutes despite constant marginal damages from pollution. This result is intuitive, given the basic economics of the upward distortion that we saw in the single country case: the reason for choosing a higher than efficient emissions price is the targeted compensation to low-cost firms. When the other regime has a policy in place, this targeted compensation becomes unimportant as low-cost firms do not cause carbon leakage even if they move. This eliminates the upward distortion in the symmetric two country equilibrium.

4.3 Generalized functional forms

Convex pollution damages. So far our analysis has built on constant marginal damages from pollution. The assumption is well justified and commonly applied in the context of climate change (see Golosov *et al.* 2014; van den Bijgaart *et al.* 2016), but it is of interest to see if the results carry over to more general situations. Assume a convex damage function $D(X)$ ($D'(\cdot) > 0$, $D''(\cdot) > 0$) where $X = \hat{X} - \int_{\underline{c}}^{\overline{c}} \phi(C(c)X(c)dc$ depicts the total emissions,

[25] This resembles the "not in my backyard" result of strategic environmental policy literature with local pollutants (Hoel, 1997; Greaker, 2003) where the regulator may have an incentive to drive away some firms, in order to avoid the damages caused by their pollution. If the externality is global, this effect only arises if the other country has more stringent environmental regulation in place.

[26] In the Appendix we derive the best response for arbitrary policies in the other, not necessarily symmetric, country.

[27] In equilibrium there is reshuffling of firms that have low or even positive moving costs but there is no net relocation as the inflow matches outflow in both regions.

including both the counterfactual emissions $\hat{X}$ and the emission reduction by the regime. Social welfare function (6) now writes as:

$$\max_{X(c),T(c)} \int_{\underline{c}}^{\overline{c}} \Big(\gamma - cX(c) - \lambda T(c)\Big)\phi(C(c),c)dc + D(X) \tag{14}$$

subject to equations (3) and (4) holding. In this case, the marginal damages from emissions, and consequently the Pigouvian emissions price, are a function of total emissions $D'(X)$. Since firms are atomistic, they cannot affect the total pollution stock X and the level of regulation. It follows that the optimal policy still takes a bang-bang form where firms optimally reduce either $X(c) = 0$ or $X(c) = 1$. From the regulator's point of view, however, the convex damages lead to an important difference: the emissions price becomes a function of the mass of firms that relocate. Relocation of low-cost facilities shifts the aggregate cost curve upwards increasing the optimal emissions price.[28] But since the regulator foresees the aggregate mass of firms that will relocate given a policy, nothing essentially changes in the regulator's problem either:

Proposition 7. *For a strictly convex damage function, the mechanism follows Proposition 2 but with $D'(X)$ replacing the constant marginal damages D.*

Convex abatement costs. The main body of the analysis has assumed that production units can reduce at most one unit of pollution with linear abatement costs, and that units can relocate independently. Next, we generalize these assumptions and consider firms that can reduce more or less than one unit of pollution $X(c) \in \mathbb{R}^+$, with convex abatement costs $A(X(c), c)$ and privately known technology parameter c. More precisely, we assume $A_x(X(c), c) > 0$, $A_{xx}(X(c), c) \geq 0$, $A_{xc}(X(c), c) > 0$ and that $A(X(c), c)$ satisfies Inada conditions. Firms can only relocate as a single entity if the cost of regulation, when reporting truthfully, exceeds the privately known relocation cost:

$$C(c) = A(X(c), c) - T(c) > \theta \tag{15}$$

When $A(X(c), c)$ is linear in $X(c)$ for all c, we have equation (1). With convex $A(X(c), c)$, however, there is a major difference to the main model as it may be optimal for firms to choose interior levels of abatement. The social welfare function becomes:

[28] This effect is also mentioned in passing, but not explicitly analyzed, by Martin *et al.* (2014).

$$\max_{X(c),T(c)} \int_{\underline{c}}^{\overline{c}} \Big(\gamma - A(X(c), c) - \lambda T(c)\Big)\phi(C(c), c) - D\Big(1 - \phi(C(c), c)X(c)\Big) dc \qquad (16)$$

subject to equation (4), $X(c)$ non-increasing and the incentive-compatibility condition, $C(c) = \min_{\hat{c}}\{A(X(\hat{c}), c) - T(\hat{c})\}$. The convex abatement cost function allows the regulator to better screen firms: the lowest-cost firms are required to reduce more emissions, which reduces the high-cost firms' incentives to understate their type. In order to provide intuition to this case, we focus on cases where full separation is optimal, that is, the firm's abatement choice varies strictly with its unobservable abatement cost under the optimal contract.[29]

Proposition 8. *The optimal mechanism satisfies:*

$$A_x(X(c), c) = \frac{D}{1+\lambda} + \frac{\mu(c)}{(1+\lambda)\phi(C(c), c)} A_{xc}(X(c), c)$$

$$\mu(c) = \int_{\underline{c}}^{c} \Big(\Delta(\tilde{c})\phi'(C(\tilde{c}), \tilde{c}) - \lambda\phi(C(\tilde{c}), \tilde{c})\Big) d\tilde{c}$$

when $X(c)$ is strictly decreasing.

The optimal second-best policy with convex abatement costs has a similar structure as Proposition 2, which is a special case of Proposition 8 (with $A_{xc}(X(c), c) = 1$ for all $X(c)$ and c): marginal costs are set equal to marginal damages, expressed in terms of public funds, plus a distortion term $\mu(\cdot)$. Moreover, the distortion term consists of welfare effects of paying marginally more to all the low-cost firms ($\tilde{c} < c$) to satisfy the incentive compatibility condition. The main difference to Proposition 2 is the appearance of term $A_{xc}(X(c), c)$, making the marginal emissions prices different for firms with different abatement costs.

It should be noted that this model has heavy information requirement because the regulator must know the shape of abatement cost function $A(X(c), c)$ for each c. Moreover, the optimal mechanism in Proposition 8, unlike the main mechanism in Proposition 2, cannot be implemented by a simple linear taxes or an emissions trading market leading to a single carbon price for each industry. Instead, it calls for either the use of nonlinear emissions price

[29]The abatement levels provided by Proposition 8 may sometimes fail to be monotonic, so that the non-monotonicity condition for $X(c)$ is binding and bunching arises: some firms with different types c are offered the same mechanism. A detailed technical analysis of bunching is provided by Rochet and Stole (2002). Following e.g. Lehmann *et al.* (2014), we focus on the cases where full separation is optimal. Technically, we focus on cases where the non-monotonicity constraint for $X(c)$ never binds.

or offering polluting firms menus of emissions prices and lump-sum compensations. In permit markets where trading is allowed among polluters, such emissions price discrimination among firms within the same industry may not be possible.

5 Discussion

5.1 Policy recommendations

The mechanism introduced in Proposition 2 has attractive implementation properties as it has an indirect market interpretation (see Figure 1): the regulator sets an emissions price and a lump-sum compensation, after which some firms cut emissions, some choose to pay the permit price and some relocate their production elsewhere. The market mechanism can either take the form of a price-based regulation with emission tax and lump-sum rebates, or a quantity-based regulation such as a cap-and-trade scheme where compensation takes the form of free allowances. There is no aggregate uncertainty and the two instruments lead to identical outcomes in contrast to Weitzman (1974).

This indirect interpretation allows a number of lessons for the real-world planning of compensation policies. First, carbon tax refunds can be used to reduce the cost of regulations when a carbon tax is in place.[30] However, as refunds are claimed by firms that pay the tax, they effectively distort carbon price downwards and work as a targeted compensation to high-cost firms. Yet, the optimal mechanisms calls for a higher emissions price when leakage is introduced; therefore carbon tax refunds are not a good way to compensate industries under leakage risk, in comparison to appropriately targeted lump-sum rebates. Second, cap-and-trade schemes generally lead to a uniform carbon price, but many trading regimes allow a limited use of offsets, domestic or international, that tend to curb the effective carbon price downwards.[31] Offsets are valueless to low-cost firms and allowing their use becomes targeted

[30]Energy tax rebates are a common tool for subsidizing energy-intensive industries. For example, in Finland, firms with energy tax bill exceeding .5% of the value-added are entitled to participate in a rebate program. This together with the rebate rule targets the largest energy-consuming firms such that the top three receive 65% of refunds in iron and steel, 55% in chemical, and 75% in paper and pulp industry. See Tamminen *et al.* (2016).

[31]Firms in the EU ETS can use international credits generated through Clean Development Mechanism and Joint Implementation, up to a polluter-specific percentage, for compliance (EC, 2017). Historically, prices in the CDM markets have been 72% of the allowance price in the EU ETS between 2008-2012 and

compensations to firms with high costs. An implication of our analysis is that industries under leakage risk should be given free allowances instead of the possibility to use offsets for compliance. Third, a threat of leakage may lead to the exclusion of entire industries from the regulation.[32] In our model, this would resemble the extreme case where the carbon price, as well as the lump-sum compensation, is set to zero. This is never an optimal policy based on our model. Instead, a sector susceptible to leakage should be included in regulation and offered adequate lump-sum compensation.

Two widely discussed policies against carbon leakage are carbon tax adjustment, a combination of a border charge on imports and a rebate on exports with the aim of levelling the playing field between domestic and foreign production (see e.g. Hoel 1996 and Fischer and Fox 2012); and output-based allocation or rebating (see e.g. Böhringer and Lange 2005, Meunier *et al.* 2014 and Böhringer *et al.* 2015) that effectively work as production subsidies. These policies are mainly advocated as they reduce leakage via output price changes. However, our model points out other co-benefits: these policies increase firms' relocation cost parameters θ by either reducing the cost of staying in the regime, or by reducing profits after moving. As firms become less likely to relocate, the regulator can pay them less compensation and save on public funds.

5.2 Calibration: EU ETS

To illustrate the main mechanism quantitatively, we carry out an explorative calibration for the key sectors under leakage risk in the EU Emissions Trading Scheme: Cement, Iron and Steel, Chemical and Plastic, Wood and Paper; and Glass.[33] Together, these five sectors produce 355 $MtCO_2$, or 62 per cent of emissions from all industrial installations covered by the EU ETS (EEA, 2017). Our analysis builds on the survey data collected by Martin *et al.* (2014). The data contains firm-level assessments of the relocation probability conditional on receiving no free permits and receiving 80% for free.[34] From these responses, we construct

only 3% between 2013-2016 (https://www.theice.com/index).

[32] This was the case for aviation sector, which was exempt from emissions pricing until 2012, but has since been included although it is receiving vast majority of allowances for free (EC, 2017).

[33] To focus on the main mechanism presented in Section 3, we assume no correlation between c and θ, and no environmental policies in the destination country.

[34] In the survey, the firms were asked: "Do you expect that government efforts to put a price on carbon emissions will force you to outsource parts of the production of this business site in the foreseeable future, or to close down completely?" and "How would your answer to the previous questions change, if you received

Table 1: Descriptive statistics of the data used

	Total emissions in 2015 ($MtCO_2$)[1]	EBIT per emissions (€/tCO_2)[2]	Relocation probability[2] 0% compensation	80% compensation	No. firms[2]
Cement	113.8	32.73	0.46	0.20	46
Iron and Steel	120.6	80.52	0.60	0.21	25
Chemical and Plastic	74.9	177.96	0.24	0.06	64
Wood and Paper	27.1	89.31	0.14	0.03	61
Glass	18.2	120.56	0.14	0.05	24

[1]Data from EEA (2017), [2]Data from Martin *et al.* (2014).

emissions-weighted industry averages for the relocation probability, see Table 1. We choose social cost carbon to be $D = 25$ euros/tCO_2[35], and fit normal distributions for relocation costs, one for each industry, based on the responses. For instance, for "cement", we calibrate the two parameters of the normal distribution using the two relocation probabilities from Table 1: 46% percent of firms relocate if the full social cost is imposed, and 20% relocate if 80% of the true social cost D that they inflict is actually given back to firms.

Our estimate of parameter γ, the industry-specific value of a firm staying, is based on emissions-weighted average earnings before investment and tax (EBIT) per unit of pollution, expressed as €/tCO_2 in Table 1. Abatement cost estimates are hard to come by at the industry level. We use the marginal abatement cost estimates for the EU energy intensive industries from Böhringer *et al.* (2014a), which is almost linear and can be approximated by a uniform distribution.[36] The social cost of public funds is $\lambda = .6$.[37] With these assumptions, computing the optimal mechanism is a straightforward numerical exercise.

a free allowance for 80% of your current emissions?" Answers were given in a Likert scale between 1 and 5, where 1 was no impact (1 %), 3 was significant reduction in production (10 %) and 5 was complete close-down (99 %).

[35]The social cost of carbon 25 €/tCO_2 comes close to the median value in a distribution from integrated-assessment model outputs of 232 distinct studies (van den Bijgaart *et al.*, 2016).

[36]From Böhringer *et al.* (2014a), we obtain $100MtCO_2$ of reductions at 47 euros/tCO_2 which pins down the slope of the marginal cost curve. Abatement is allocated to sectors in proportion to their unrestricted emissions.

[37]Country-circumstances have a large impact on the real costs of taxation so one number cannot fit the entire EU. The chosen number is higher than those presented in Bovenberg and Goulder (1996) but closer to the survey of more recent estimates in Holtsmark and Bjertnæs (2015).

Table 2: Optimal mechanism for the EU ETS sectors

	Compensation per action			Industry-level	
	Cut (€/tCO$_2$)	Pollute (€/tCO$_2$)	Effective CO$_2$ price (€/tCO$_2$)	Optimal leakage	Optimal windfalls (M€)
Cement	5.3	-9.6	14.9	24.5 %	1,753
Iron and Steel	20.6	6.9	13.7	7.3 %	3,259
Chemical and Plastic	14.4	0.7	13.7	3.5 %	2,044
Wood and Paper	-2.2	-16.3	14.1	7.4 %	390
Glass	0.0	-14.5	14.5	8.2 %	188
Benchmark: observable c			15.6		
Benchmark: no leakage			11.3		

Notes: Optimal compensations in EU sectors, the implied marginal carbon tax rate, the total leakage and overcompensation under the optimal policy. The social cost of carbon is 25 €/tCO_2 and the social cost of public funds is $\lambda = .6$. Assumptions detailed in the text.

We report the optimal climate policies per sector in Table 2. The first column gives the compensation level for firms that cut emissions, the second column presents the transfer to firms that pollute, and third column gives the effective emissions price in the sector. The key sectors are treated very differently. At one extreme, Iron and Steel polluters receive a compensation, whereas the compensation to Wood and Paper is negative; that is, even after cutting emissions, firms in those sectors face a net tax. To further interpret the results, we rely on two benchmarks. First, the optimal emissions price in the absence of leakage but privately observed c is based on equation (9), that is, 11.3 euros/tCO_2. This helps us to identify the stand-alone impact of leakage on the emissions price. Second, the Pigouvian price when c is observable is obtained from Proposition 1, that is, 15.6 euros/tCO_2. This benchmark serves to identify the impact of non-observable c. We find that the distortions entailed by the mechanism are quantitatively significant: the effective CO_2 price is substantially distorted upwards, by $21 - 32$ per cent compared to the benchmark level with where leakage was assumed away (11.3 euros/tCO_2). Yet, abatement cost effect dominates for all the sectors and the emissions price falls short of the benchmark with observable c.

The last two columns in Table 2 show: the optimal trade-off between carbon leakage, the share of industry's emissions that will be relocated; and the total windfall profits, defined as the transfer in excess of relocation costs. The optimal carbon leakage is the highest (24.5 %) for the cement sector, which is the most emission-intensive and therefore has the lowest

EBIT per emissions (parameter γ). Keeping a big fraction of the cement firms would require large compensation per each unit of added value; tolerating a relatively high leakage is a way to reduce windfalls. The other key sectors have higher added-value per a unit of production (parameter γ), and the optimal leakage allowed by the regulator is lower, less than 10 per cent. Chemical and Plastics together with Iron and Steel are the two key sectors receiving the most windfalls under the optimal policy.[38]

6 Conclusions

> "High energy and electricity prices and unequal carbon pricing place the EU manufacturing sector in general – and the cement sector in particular – at risk of carbon leakage."
>
> — Lafrage (2013)

> "In theory [EU ETS] provides a cheap and efficient means to limit greenhouse gas reductions within an ever-tightening cap, but in practice it has rewarded major polluters with windfall profits[...]"
>
> — Carbon Trade Watch (2017)

Should firms be offered compensation for accepting regulations? As it is impossible to tailor compensation exactly to the firms who need it the most, any policy necessarily either involves excessive costs and firm relocation, or leads to excessive windfalls being paid to the polluting industries. The two quotes above, one expressed by the industry and another one by an NGO, illuminate this trade-off. The purpose of this study is to solve how an optimally designed policy can alleviate this conflict and strike an optimal balance between these two views. In particular, we argue that this disagreement is not just about distribution of surplus created by climate policies, but it also has direct welfare implications as it determines *who* will participate in the regulation.

As a first result, we find that carbon leakage is an essential part of an optimal second-best policy, as it decreases the overcompensation paid to the remaining firms. Observing carbon leakage is therefore not a sign of a failed environmental policy but, conversely, one

[38] In the Appendix, we analyze the sensitivity of our results to changes in key parameters: (i) for choosing a lower value for the social cost of public funds and (ii) for the case where the planner only cares about climate damages, and puts no direct benefit γ for keeping the firms.

can argue that EU ETS has failed exactly because no carbon leakage is observed; see studies by Dechezleprêtre *et al.* (2014) and Naegele and Zaklan (2017). Second, sectors exposed to leakage risk should never be compensated by rolling back environmental regulations; rather, leakage calls for stricter regulation accompanied by lump-sump compensations. In particular, the results of our model advise against certain real-world policies that are used to compensate industries under leakage risk, such as emission tax refunds, allowing those industries to use international offsets and exempting certain industries from regulation.

References

Ahlvik, L., Liski, M. and Martin, R. (2017). The association between the vulnerability score and compliance costs. *Mimeo*.

Baron, D. P. and Myerson, R. B. (1982). Regulating a monopolist with unknown costs. *Econometrica*, **50** (4), 911–930.

Böhringer, C., Bye, B., Fæhn, T. and Rosendahl, K. E. (2015). *Output-based rebating of carbon taxes in the neighbor's backyard*. Tech. rep., Oldenburg Discussion Papers in Economics.

—, Dijkstra, B. and Rosendahl, K. E. (2014a). Sectoral and regional expansion of emissions trading. *Resource and Energy Economics*, **37**, 201–225.

—, Fischer, C. and Rosendahl, K. E. (2014b). Cost-effective unilateral climate policy design: Size matters. *Journal of Environmental Economics and Management*, **67** (3), 318–339.

— and Lange, A. (2005). On the design of optimal grandfathering schemes for emission allowances. *European Economic Review*, **49** (8), 2041–2055.

—, Lange, A. and Rutherford, T. F. (2014c). Optimal emission pricing in the presence of international spillovers: Decomposing leakage and terms-of-trade motives. *Journal of Public Economics*, **110**, 101–111.

Bovenberg, A. L. and Goulder, L. H. (1996). Optimal environmental taxation in the presence of other taxes: general-equilibrium analyses. *The American Economic Review*, **86** (4), 985–1000.

Carbon Trade Watch (2017). Carbon trade watch: Cap and trade. http://www.carbontradewatch.org/issues/cap-and-trade2.html.

Dasgupta, P., Hammond, P. and Maskin, E. (1980). On imperfect information and optimal pollution control. *The Review of Economic Studies*, **47** (5), 857–860.

d'Aspremont, C. and Gérard-Varet, L.-A. (1979). Incentives and incomplete information. *Journal of Public Economics*, **11** (1), 25–45.

— and Gérard-Varet, L.-A. (1983). Regional externalities and efficient decentralization under incomplete information. *Jacques-Francois Thisse and Henry G. Zoller, eds., Location analysis of public facilities (North-Holland, Amsterdam)*, pp. 207–220.

Dechezleprêtre, A., Gennaioli, C., Martin, R., Muï, M. *et al.* (2014). *Searching for carbon leaks in multinational companies*. Tech. rep., Grantham Research Institute on Climate Change and the Environment.

EC (2017). The EU emissions trading system (EU ETS). https://ec.europa.eu/clima/policies/ets_en.

Ederington, J., Levinson, A. and Minier, J. (2005). Footloose and pollution-free. *The Review of Economics and Statistics*, **87** (1), 92–99.

EEA (2017). EU emissions trading system (ets) data viewer. https://www.eea.europa.eu/data-and-maps/dashboards/emissions-trading-viewer-1.

Fischer, C. and Fox, A. K. (2012). Comparing policies to combat emissions leakage: Border carbon adjustments versus rebates. *Journal of Environmental Economics and Management*, **64** (2), 199–216.

Gerlagh, R. and Liski, M. (2016). Carbon prices for the next hundred years. *The Economic Journal*, **doi:10.1111/ecoj.12436**.

Golosov, M., Hassler, J., Krusell, P. and Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, **82** (1), 41–88.

Greaker, M. (2003). Strategic environmental policy when the governments are threatened by relocation. *Resource and Energy Economics*, **25** (2), 141–154.

HARSTAD, B. (2012). Buy coal! a case for supply-side environmental policy. *Journal of Political Economy*, **120** (1), 77–115.

— and ESKELAND, G. S. (2010). Trading for the future: Signaling in permit markets. *Journal of Public Economics*, **94** (9), 749–760.

— and LISKI, M. (2013). Games and resources. *In: Shogren, J.F., (ed.) Encyclopedia of Energy, Natural Resource, and Environmental Economics, Vol. 2*, pp. 299–308.

HELLWIG, M. F. (2003). Public-good provision with many participants. *The Review of Economic Studies*, **70** (3), 589–614.

HELM, C. and WIRL, F. (2014). The principal–agent model with multilateral externalities: An application to climate agreements. *Journal of Environmental Economics and Management*, **67** (2), 141–154.

HOEL, M. (1996). Should a carbon tax be differentiated across sectors? *Journal of Public Economics*, **59** (1), 17–32.

— (1997). Environmental policy with endogenous plant locations. *The Scandinavian Journal of Economics*, **99** (2), 241–259.

HOLTSMARK, B. and BJERTNÆS, G. H. (2015). The size of the marginal cost of public funds – a discussion with speciel relevance to Norway. *Statistics Norway, Reports 2015/13.*

JULLIEN, B. (2000). Participation constraints in adverse selection models. *Journal of Economic Theory*, **93** (1), 1–47.

KIM, J.-C. and CHANG, K.-B. (1993). An optimal tax/subsidy for output and pollution control under asymmetric information in oligopoly markets. *Journal of Regulatory Economics*, **5** (2), 183–197.

LAFFONT, J.-J. and MARTIMORT, D. (2002). *The theory of incentives: the principal-agent model*. Princeton University Press.

— and TIROLE, J. (1996). Pollution permits and compliance strategies. *Journal of Public Economics*, **62** (1), 85–125.

Lafrage (2013). Lafarge answer to the public consultation on "A 2030 framework for climate and energy policies". *Comment*.

Lehmann, E., Simula, L. and Trannoy, A. (2014). Tax me if you can! optimal nonlinear income tax between competing governments. *The Quarterly Journal of Economics*, **129** (4), 1995–2030.

Levinson, A. and Taylor, M. S. (2008). Unmasking the pollution haven effect. *International Economic Review*, **49** (1), 223–254.

Lewis, T. R. (1996). Protecting the environment when costs and benefits are privately known. *The RAND Journal of Economics*, pp. 819–847.

— and Sappington, D. E. (1989). Inflexible rules in incentive problems. *The American Economic Review*, **27** (4), 69–84.

Maggi, G. and Rodriguez-Clare, A. (1995). On countervailing incentives. *Journal of Economic Theory*, **66** (1), 238–263.

Markusen, J. R. (1975). International externalities and optimal tax structures. *Journal of International Economics*, **5** (1), 15–29.

—, Morey, E. R. and Olewiler, N. D. (1993). Environmental policy when market structure and plant locations are endogenous. *Journal of Environmental Economics and Management*, **24** (1), 69–86.

Martimort, D. and Sand-Zantman, W. (2015). A mechanism design approach to climate-change agreements. *Journal of the European Economic Association*, **14** (3), 669–718.

Martin, R., Muûls, M., De Preux, L. B. and Wagner, U. (2014). Industry compensation under relocation risk: A firm-level analysis of the eu emissions trading scheme. *The American Economic Review*, **104** (8), 2482–2508.

Meunier, G., Montero, J.-P. and Ponssard, J.-P. (2016). Output-based allocations in pollution markets with uncertainty and self-selection. *Working paper*.

—, Ponssard, J.-P. and Quirion, P. (2014). Carbon leakage and capacity-based allocations: Is the EU right? *Journal of Environmental Economics and Management*, **68** (2), 262–279.

Montero, J.-P. (2000). Optimal design of a phase-in emissions trading program. *Journal of Public Economics*, **75** (2), 273–291.

— (2008). A simple auction mechanism for the optimal allocation of the commons. *The American Economic Review*, **98** (1), 496–518.

Motta, M. and Thisse, J.-F. (1994). Does environmental dumping lead to delocation? *European Economic Review*, **38** (3-4), 563–576.

Naegele, H. and Zaklan, A. (2017). *Does the EU ETS Cause Carbon Leakage in European Manufacturing?* Tech. rep., DIW Berlin, German Institute for Economic Research.

Norman, P. (2004). Efficient mechanisms for public goods with use exclusions. *The Review of Economic Studies*, **71** (4), 1163–1188.

Rochet, J.-C. and Stole, L. A. (2002). Nonlinear pricing with random participation. *The Review of Economic Studies*, **69** (1), 277–311.

Schmidt, R. C. and Heitzig, J. (2014). Carbon leakage: Grandfathering as an incentive device to avert firm relocation. *Journal of Environmental Economics and Management*, **67** (2), 209–223.

Seierstad, A. and Sydsæter, K. (1987). *Optimal control theory with economic applications*. Elsevier North-Holland, Inc.

Spulber, D. F. (1988). Optimal environmental regulation under asymmetric information. *Journal of Public Economics*, **35** (2), 163–181.

Tamminen, S., Ollikka, K. and Laukkanen, M. (2016). Suomen energiaverotus suosii energiaintensiivisiä suuryrityksiä. *VATT Policy Brief*, **2-2016**.

Ulph, A. and Valentini, L. (1997). Plant location and strategic environmental policy with inter-sectoral linkages. *Resource and Energy Economics*, **19** (4), 363–383.

van den Bijgaart, I., Gerlagh, R. and Liski, M. (2016). A simple formula for the social cost of carbon. *Journal of Environmental Economics and Management*, **77**, 75–94.

Weitzman, M. L. (1974). Prices vs. quantities. *The Review of Economic Studies*, **41** (4), 477–491.

APPENDIX FOR ONLINE PUBLICATION

A Appendix: Proofs

Proof of Proposition 1.
The policy problem can be written as:

$$\max_{X(c),C(c)} W = \int_{\underline{c}}^{\overline{c}} \Big(\gamma - cX(c) - \lambda(cX(c) - C(c))\Big)\phi(C(c), c) - D\Big(1 - \phi(C(c), c)X(c)\Big)dc$$

where we used $T(c) = cX(c) - C(c)$. The objective function is linear in $X(c)$ so the optimal solution takes a bang-bang form where either $X(c) = 0$ or $X(c) = 1$. The condition for $X(c) = 1$ is:

$$\frac{\partial W}{\partial X(c)} \geq 0 \Rightarrow (1 + \lambda)c \leq D \tag{A.1}$$

and $X(c) = 0$ otherwise. By Assumption 1, we show that the objective is monotonic in $C(c)$ and that there exists bounded optimal $C(c)$. The first-order condition for $C(c)$ is:

$$\frac{\partial W}{\partial C(c)} = 0 \Rightarrow \lambda\phi(C(c), c) = \underbrace{-\Big(\gamma + (D - c)X(c) - \lambda T(c)\Big)}_{=\Delta(X(c),T(c),c)} \phi'(C(c), c), \tag{A.2}$$

which can be rewritten as

$$\lambda\frac{1 - G(C(c))}{-g(C(c))} = \Delta(X(c), T(c), c).$$

By Assumption 1, the left-hand side is increasing while, by definition, $\Delta(.)$ is strictly decreasing in $C(c)$ for any given c. Thus, the solution to (A.2) exists and is unique. Q.E.D.

Proof of Corollary 1.
For *(i)*, note that when $X_B(c) = 0$, the first-order condition for transfers (A.2) becomes:

$$\lambda\phi(-T_B(c), c) = -\Big(\gamma - \lambda T_B(c)\Big)\phi'(-T_B(c), c)$$

that is independent of c. It immediately follows from this and (A.1) that $T_B'(c) = 0$ for $c > D/(1+\lambda)$. For *(ii)*, we evaluate the optimal compensation at both sides of $c_B^* = D/(1 + \lambda)$. Denote $T_B^-(c_B^*) = \lim_{c\to D/(1+\lambda)^-} T_B(c_B^*)$ and $T_B^+(c_B^*) = \lim_{c\to D/(1+\lambda)^+} T_B(c_B^*)$. At c_B^* the regulator is indifferent between $X(c) = 0$ and $X(c) = 1$ so condition (A.2) must hold for both of these actions:

$$-\Big(\gamma + (D - c_B^*) - \lambda T^-(c_B^*)\Big)\eta(c_B^* - T^-(c_B^*)) = -\Big(\gamma - \lambda T^+(c_B^*)\Big)\eta(-T^+(c_B^*))$$

where $\eta(\cdot) = \phi(\cdot)/\phi'(\cdot)$. Plug in $D = (1 + \lambda)c_B^*$ and simplify:

$$\Big(\gamma + \lambda(c_B^* - T^-(c_B^*))\Big)\eta(c_B^* - T^-(c_B^*)) = \Big(\gamma - \lambda T^+(c_B^*)\Big)\eta(-T^+(c_B^*))$$

It can be verified that this equation is satisfied for $T^{-}(c_B^*) - T^{+}(c_B^*) = c_B^*$.

For *(iii)*, using $C_B(c) = cX_B(c) - T_B(c)$ and $\eta(\cdot) = \phi(\cdot)/\phi'(\cdot)$ to write condition (A.2) as

$$\lambda\eta(C_B(c)) = -\Big(\gamma + (D - (1+\lambda)c)X_B(c) + \lambda C_B(c)\Big)$$

Then, differentiate both sides with respect to c:

$$\lambda\eta'(C_B(c))C_B'(c) = (1+\lambda)X_B(c) - \lambda C_B'(c)$$

$$\Rightarrow$$

$$C_B'(c) = \frac{(1+\lambda)X_B(c)}{\lambda(\eta'(C_B(c))+1)}.$$

By $C_B(c) = cX_B(c) - T_B(c)$, condition $T_B'(c) < 0$ is equivalent to $C_B'(c) > X_B(c)$ (recall that $X_B(c) = 1$ for all $c < \frac{D}{1+\lambda}$), so

$$T_B'(c) < 0$$

$$\Leftrightarrow$$

$$C_B'(c) = \frac{(1+\lambda)X_B(c)}{\lambda(\eta'(C_B(c))+1)} > X_B(c)$$

$$\Leftrightarrow$$

$$1 > \lambda\eta'(C_B(c))$$

Q.E.D.

Proof of Proposition 2. We begin by introducing a series of lemmas that characterize the optimal mechanism.

Lemma 1. *The transfer is constant when the policy $X(c)$ is constant.*

Proof. Proof by contradiction. Assume that there are c and c' with $T(c') > T(c)$ with $X(c') = X(c)$. Now firm c can get a lower net cost by reporting c':

$$cX(c') - T(c') < cX(c) - T(c)$$

However, this is in violation of the incentive compatibility condition in equation (3). Q.E.D.

Lemma 2. *$X(c)$ is nonincreasing in c.*

Proof. Proof by contradiction. If this is not true, there are types c and c', with $c < c'$ and $X(c') > X(c)$. Incentive compatibility requires $T(c)$ and $T(c')$ such that types do not want to report the other type:

$$-T(c) \leq cX(c') - T(c')$$

$$-T(c) \geq c'X(c') - T(c')$$

Combining these two inequalities leads to:

$$c'X(c') - T(c') \leq cX(c') - T(c') \Rightarrow c' \leq c$$

But this is a contradiction. Q.E.D.

Lemma 3. *Optimal abatement solution takes bang-bang form where $X(c) = \{0, 1\}$*

Proof. The problem can be stated as:

$$\max_{X(c),C(c)} W = \int_{\underline{c}}^{\overline{c}} \Big(\gamma - cX(c) - \lambda(cX(c) - C(c))\Big)\phi(C(c), c) - D\Big(1 - \phi(C(c), c)X(c)\Big)dc$$

s.t. $C'(c) = -X(c)$ holds for all c. The Hamiltonian for this problem reads:

$$\mathcal{H} = \Big(\gamma + (D - (1+\lambda)c)X(c) + \lambda C(c)\Big)\phi(C(c), c) - \mu(c)X(c) - D$$

The Hamiltonian is linear in control $X(c)$, and the necessary condition for optimality is that $X(c)$ takes a bang-bang form: $X^*(c) = \{0, 1\}$. Q.E.D.

There exists a solution to the problem as stated in Lemma 3 by Filippov-Cesari Theorem (Theorem 8, page 132, Seierstad and Sydsæter 1987). We find next a solution that satisfies the necessary conditions; it follows by Assumption 1 that there is a unique solution that satisfies the conditions.

Lemmas 1-3 above tell us that the optimal policy takes a threshold form, where $X(c) = 0$ for $c \leq c^*$ and $X(c) = 1$ for $c > c^*$. Transfers are either $T_1(c) = T$ for $c > c^*$ or $T_2(c) = T + c^*$ for $c \leq c^*$, guaranteeing indifference for type c^*: $-T_1(c^*) = c^* - T_2(c^*)$. In this formulation, the regulator is left to find c^* and T that maximize the social welfare:

$$\max_{c^*,T} W = \int_{\underline{c}}^{c^*} \Big(\gamma + (D - c) - \lambda(T + c^*)\Big)\phi(c - T - c^*, c)dc + \int_{c^*}^{\overline{c}} \Big(\gamma - \lambda T\Big)\phi(-T, c)dc \qquad \text{(A.3)}$$

Here the first integral covers all the firms below the threshold c^* cutting emissions, and the second term covers firms above the threshold that do not cut emissions. Begin by taking the first-order condition with respect to c^* using Leibniz's integral rule, we get:

$$\Big(\gamma+(D-c^*)-\lambda(T+c^*)\Big)\phi(-T, c^*)+\Big(\lambda T-\gamma\Big)\phi(-T, c^*)+\int_{\underline{c}}^{c^*} \Big(\Delta(c)\phi'(c-T-c^*, c)-\lambda\phi(c-T-c^*)\Big)dc = 0 \qquad \text{(A.4)}$$

where $\Delta(c) = -\Big(\gamma + (D-c)X(c) - \lambda T(c)\Big)$ denotes the net loss from relocation. Simplify and solve for c^* to get:

$$c^* = \frac{D}{1+\lambda} + \frac{\mu(c^*)}{(1+\lambda)\phi(-T, c^*)} \qquad \text{(A.5)}$$

where $\mu(c)$ is the integral term

$$\mu(c^*) = \int_{\underline{c}}^{c^*} \Big(\Delta(c)\phi'(C(c),c) - \lambda\phi(C(c),c)\Big)dc.$$

Then, find the first-order condition with respect to T:

$$\underbrace{\int_{\underline{c}}^{c^*} \Big(\Delta(c)\phi'(c-T+c^*,c) - \lambda\phi(c-T+c^*,c)\Big)dc}_{=\mu(c^*)} + \int_{c^*}^{\bar{c}} \Big(\Delta(c)\phi'(-T,c) - \lambda\phi(-T,c)\Big)dc = 0 \quad \text{(A.6)}$$

For $c > c^*$, $X(c) = 0$ and therefore the second integral does not depend on c apart from the term $f(c)$, see eq. (4), so that (A.6) can be written as:

$$\mu(c^*) + \Delta(c^*) \underbrace{g(-T)(-1)}_{=\phi'(-T,c^*)/f(c^*)} (1 - F(c^*)) - \lambda\Big(\underbrace{(1 - G(-T))}_{=\phi(-T,c^*)/f(c^*)} (1 - F(c^*))\Big) = 0 \quad \text{(A.7)}$$

$$\Rightarrow$$

$$\Delta(c^*)\phi'(-T,c^*) - \lambda\phi(-T,c^*) = -\frac{f(c^*)}{1-F(c^*)}\mu(c^*). \quad \text{(A.8)}$$

Proposition 2 follows directly from this. Q.E.D.

Proof of Proposition 3.

The distortion term writes as:

$$\mu(c^*) = \int_{\underline{c}}^{c^*} \Big[\underbrace{\Delta(c)\phi'(C(c),c) - \lambda\phi(C(c),c)}_{\equiv\Omega(C(c),c)}\Big]dc. \quad \text{(A.9)}$$

From Proposition 1, optimal benchmark transfer $C_B(c)$ under observable c solves

$$\Omega(C_B(c),c) = 0, \text{ for all } c \leq c_B^* = \frac{D}{1+\lambda}. \quad \text{(A.10)}$$

Continue assuming observable c and define

$$\widehat{C}_B(c) = cX_B(c) - T_B(c_B^*) \quad \text{(A.11)}$$

as the net cost of regulation for firm under a distorted mechanism with transfer $T_B(c_B^*)$ imposed on all firm types $c < c_B^*$.

The proof follows from:

$$\lambda\eta'(C_B(c)) < 1 \tag{A.12}$$

$$\Leftrightarrow$$

$$C_B'(c) > 1, \text{ for all } c < c_B^* = \frac{D}{1+\lambda} \tag{A.13}$$

$$\Rightarrow$$

$$\Omega(\widehat{C}_B(c), c) > 0, \text{ for all } c < c_B^* = \frac{D}{1+\lambda} \tag{A.14}$$

$$\Rightarrow$$

$$\mu(c_B^*) > 0. \tag{A.15}$$

Line (A.12) is the condition in the Proposition. Line (A.13) is item (iii) of Corollary 1 (note: $T_B'(c) < 0$ is equivalent to $C_B'(c) > X_B(c) = 1$ as $X_B(c) = 1$ for all $c < \frac{D}{1+\lambda}$). For line (A.14), $\widehat{C}_B(c) > C_B(c)$ for $c < c_B^*$ by construction of $\widehat{C}_B(c)$ together with line (A.13). Thus, compensation $T_B(c_B^*)$ is too low at $c < c_B^*$. Then, note that $\Omega(C(c), c)$ is monotonically increasing in $C(c)$ at each c, shown in the Proof of Proposition 1. These observations imply that it must be $\Omega(\widehat{C}_B(c), c) > 0$ for $c < c_B^*$. Line (A.15) then follows from

$$\mu(c_B^*) = \int_{\underline{c}}^{c_B^*} \Omega(\widehat{C}_B(c), c) dc > 0. \tag{A.16}$$

We have now evaluated the distortion term at c_B^*. Straightforwardly, the argument for $\mu(c_B^*) > 0$ implies $\mu(c) > 0$ for any $c < c_B^*$. We now ready to demonstrate the following:

$$\mu(c_B^*) > 0 \Rightarrow c^* > c_B^* \Rightarrow \mu(c^*) > 0 \tag{A.17}$$

The first implication in (A.17): assuming $c^* \leq c_B^*$ is a contradiction with the definition of c^*,

$$c^* = \frac{D}{1+\lambda} + \frac{\mu(c^*)}{(1+\lambda)\phi(C(c^*), c)} \tag{A.18}$$

$$= c_B^* + \frac{\mu(c^*)}{(1+\lambda)\phi(C(c^*), c)} \tag{A.19}$$

where $\mu(c^*) > 0$ for $c^* \leq c_B^*$ by our proof above. The second implication in (A.17) follows since $c^* > c_B^*$ can only hold with $\mu(c^*) > 0$.

Case $\lambda\eta'(C_B(c)) > 1$ follows the same reasoning but now $C_B'(c) < 1$ so that the net cost burden to the firm increases in type $c < c_B^*$ slower than abatement costs. It follows that the contract for the marginal type would lead to a net cost that is distorted downwards for the other types: $\widehat{C}_B(c) < C_B(c)$ for $c < c_B^*$. Transfer T_B to the last cutting type is too high if applied for $c < c_B^*$.

As a result, it must be that $\Omega(\widehat{C}_B(c), c) < 0$ for $c < c_B^*$, leading to $\mu(c_B^*) < 0$ and then $\mu(c^*) < 0$, following the steps from above. Q.E.D.

Proof of Proposition 4.

The condition for firm type c, θ to stay is

$$C(c) \leq \frac{\theta}{\sigma} + \hat{\theta},$$

or

$$\theta \geq \sigma\Big(C(c) - \hat{\theta}\Big)$$

where $\sigma > 0$ scales the spread around mean $\hat{\theta}$ for a given zero-mean risk θ. And the mass of firms for which this holds is $\phi(\sigma(C(c) - \hat{\theta}), c)$ at c. The optimal benchmark compensation, now including the scaling parameter σ becomes:

$$\sigma\Delta(c) = \lambda\eta\Big(\sigma(C_B(c) - \hat{\theta})\Big). \tag{A.20}$$

We prove the result by looking at how the benchmark schedule, $T_B(c)$ (or, $C_B(c)$), from Proposition 1 varies with marginal changes in parameters γ and σ. These variations in $C_B(c)$ under observable c translate into variations in the schedule that is optimal when c is not observable. First, differentiate both sides of equation (A.20) with respect to γ to get:

$$-1 - \lambda\frac{\partial C_B(c)}{\partial\gamma} = \lambda\frac{\partial C_B(c)}{\partial\gamma}\eta'(\cdot)$$

$$\Rightarrow$$

$$\frac{\partial C_B(c)}{\partial\gamma} = -\frac{1}{\lambda(1 + \eta'(\sigma(C_B(c) - \hat{\theta})))} < 0 \tag{A.21}$$

Higher γ always decreases the optimal net cost of regulation imposed on firms, that is, increases compensation $T_B(c)$ at any given c. As $T_B(c) = cX_B(c) - C_B(c)$, equation (A.21) means that $\partial T_B(c)/\partial\gamma > 0$. This proves the first part of result (i) in the Proposition. We proceed now, based on changes in $(X_B(c), T_B(c))$, to infer the changes in the mechanism $(X(c), T(c))$, optimal in the case of unobservable c and θ.

We begin by analyzing whether $C_B(c)$ is more sensitive to changes in γ for large or high values of c. Differentiating equation (A.21) with respect to c we get:

$$\frac{\partial^2 C_B(c)}{\partial\gamma\partial c} = \frac{\lambda\eta''(\cdot)C_B'(c)\sigma}{\Big(\lambda(1 + \eta'(\cdot))\Big)^2} < 0 \tag{A.22}$$

$$\Rightarrow$$

$$\frac{\partial C_B(c')}{\partial\gamma} > \frac{\partial C_B(c'')}{\partial\gamma} \text{ for } c' < c'' < c_B^* \tag{A.23}$$

where the last inequality of (A.22) follows as $C'_B(c) > 0$, and because we assumed convex hazard rate implying that $\eta(x) = \frac{1-G(\theta|c)}{-g(\theta|c)}$ is concave: $\eta''(x) < 0$ for all x. Equation (A.23) tells that the schedule $C_B(c)$ is more sensitive to changes in γ at high levels of net costs.

Then, based on this, we want to establish that for $\mu(c_B^*)$ defined in (A.16),

$$\frac{\partial \mu(c_B^*)}{\partial \gamma} = \int_{\underline{c}}^{c_B^*} \frac{\partial \Omega(\widehat{C}_B(c), c)}{\partial \gamma} dc < 0. \tag{A.24}$$

To see (A.24), note that $\Omega(C_B(c), c)$ is monotonic in $C_B(c)$ for each c (from the proof of Proposition 1). Then, we need to distinguish two cases. First, if $T_B(c) > T_B(c^*)$, the compensation to type c_B^* is lower than to $c < c_B^*$, so that $C_B(c) < \widehat{C}_B(c)$ where $\widehat{C}_B(c)$ is defined in (A.11). Because the gap between $C_B(c)$ and $\widehat{C}_B(c)$ decreases in γ by (A.23) for each type $c < c_B^*$, $\Omega(\widehat{C}_B(c), c)$ must decrease for each $c < c_B^*$:

$$\partial \Omega(\widehat{C}_B(c), c)/\partial \gamma < 0 \tag{A.25}$$

Formally, $\Omega(\widehat{C}_B(c), c)$ measures how much the first-order condition for setting the optimal cost level for each type deviates from zero. Second, if $T_B(c) < T_B(c_B^*)$, the compensation to type c_B^* is higher than to $c < c_B^*$, so that $C_B(c) > \widehat{C}_B(c)$. Now, a change in γ widens the gap $C_B(c) - \widehat{C}_B(c) > 0$ since $C_B(c)$ is more sensitive to changes in γ at high levels of c. Then, by monotonicity of $\Omega(.)$, value $\Omega(\widehat{C}_B(c), c) < 0$ becomes more negative, that is, we have again $\partial \Omega(\widehat{C}_B(c), c)/\partial \gamma < 0$. The proof is completed by replacing $c_B^* = c^*$ in the above. This proves the first part of part (ii) of the proposition.

The proof for σ follows the same reasoning. Begin by differentiating both sides of equation (A.20) with respect to σ to get:

$$\Delta(c) - \lambda\sigma \frac{\partial C_B(c)}{\partial \sigma} = \lambda \eta'(\cdot)\Big[(C_B(c) - \hat{\theta}) + \sigma \frac{\partial C_B(c)}{\partial \sigma}\Big]$$

$$\frac{\partial C_B(c)}{\partial \sigma} = \frac{-\lambda \eta'(\cdot)\Big(C_B(c) - \hat{\theta}\Big) + \Delta(c)}{\lambda\sigma\Big(1 + \eta'(\cdot)\Big)} \tag{A.26}$$

In general, the sign of right-hand side of (A.26) is ambiguous, meaning that the sign of $\partial T_B(c)/\partial \sigma$ is ambiguous. To continue, we want to analyze whether $\partial C_B(c)/\partial \sigma$ is higher for high or low values of c. To do this, we differentiate (A.26) with respect to c and simplify:

$$\frac{\partial^2 C_B(c)}{\partial \sigma \partial c} = -\frac{1+\lambda}{\lambda}\Big[(C_B(c) - \hat{\theta}) + \sigma \frac{\partial C_B(c)}{\partial \sigma}\Big]\eta''(\cdot)\Big(1 + \eta'(\cdot)\Big)^{-2} \tag{A.27}$$

As we have assumed $\eta''(c) < 0$, the sign of this term depends on the sign of the term in square brackets:

$$C_B(c) - \hat{\theta} + \sigma \frac{\partial C_B(c)}{\partial \sigma}$$

Plug in $\partial C_B(c)/\partial\sigma$ from equation (A.26):

$$= C_B(c) - \hat{\theta} + \sigma \frac{-\lambda\eta'(\cdot)\Big(C_B(c) - \hat{\theta}\Big) + \Delta(c)}{\lambda\sigma\Big(1 + \eta'(\cdot)\Big)}$$

$$\Rightarrow$$

$$= \frac{\Big(C_B(c) - \hat{\theta}\Big)\Big(\lambda\sigma\Big(1 + \eta'(\cdot)\Big)\Big) - \sigma\lambda\eta'(\cdot)\Big(C_B(c) - \hat{\theta}\Big) + \sigma\Delta(c)}{\lambda\sigma\Big(1 + \eta'(\cdot)\Big)}$$

$$\Rightarrow$$

$$= \frac{-\hat{\theta}\lambda + \Big((1+\lambda)c - D - \gamma\Big)}{\lambda\Big(1 + \eta'(\cdot)\Big)} < 0$$

$$\Rightarrow$$

$$\frac{\partial C_B(c')}{\partial\sigma} > \frac{\partial C_B(c'')}{\partial\sigma} \text{ for } c' < c'' < c_B^* \tag{A.28}$$

where the last inequality follows from the fact that $c \leq D/(1+\lambda)$ so that the numerator is negative. It follows that $\frac{\partial C_B'(c)}{\partial\sigma} > 0$. The remaining of the proof follows steps that are identical to the sensitivity analysis done with respect γ from equations (A.23) onwards. Q.E.D.

Proof of Proposition 5.

Lemmas 1-3 continue to hold when correlation between c and θ is introduced. Based on the Lemmas, the regulator adopts a threshold policy, paying $T(c) = T^*$ for $c > c^*$ and $T(c) = T^* + c^*$ for $c \leq c^*$. Firms leave the regime if their cost of staying exceeds the cost of relocation. With perfect correlation, the private information is rendered one-dimensional and we have two thresholds: (i) the intensive margin determining which firms cut emissions, and (ii) the extensive margin determining which firms leave the regime. Firms that do not cut emissions leave if

$$-T^* > b + kc$$

Leading to

$$c < -\frac{b + T^*}{k} \text{ for } k > 0 \tag{A.29a}$$

$$c > -\frac{b + T^*}{k} \text{ for } k < 0 \tag{A.29b}$$

Firms cutting emissions leave if:

$$c - T^* - c^* > b + kc$$

Leading to

$$c > \frac{b + T^* + c^*}{1-k} \text{ for } k < 1 \tag{A.30a}$$

$$c < \frac{b + T^* + c^*}{1-k} \text{ for } k > 1 \tag{A.30b}$$

We can find three separate cases: (i) when $k < 0$ only (A.29b) binds, (ii) when $k > 1$ only (A.30b) and (iii) when $0 \leq k < 1$ both (A.29a) and (A.30a) bind (see Figure 4 for a graphical illustration).

Concurring incentives, $k < 0$: The social welfare function can be divided in three intervals. Firms cut emissions if $c \leq c^*$, firms stay but not cut emissions when $c^* < c = c' \leq -\frac{b+T^*}{k}$ and firms leave when $c = c' > -\frac{b+T^*}{k}$. The social welfare function can be written as:

$$\max_{c^*,T^*} \int_{\underline{c}}^{c^*} \Big(\gamma + (D-c) - \lambda(T^* + c^*)\Big) f(c)dc + \int_{c^*}^{-\frac{b+T^*}{k}} \Big(\gamma - \lambda T^*\Big) f(c)dc$$

Take the first order conditions with respect to c^*:

$$\Big(D - (1+\lambda)c^*\Big) f(c^*) - \int_{\underline{c}}^{c^*} \lambda f(c)dc = 0$$

$$D - (1+\lambda)c^* - \lambda \frac{F(c^*)}{f(c^*)} = 0$$

$$c^* = \frac{D}{1+\lambda} - \frac{\lambda}{1+\lambda}\frac{F(c^*)}{f(c^*)} \tag{A.31}$$

Take the first order conditions with respect to T^*:

$$-\frac{1}{k}\underbrace{\Big(\gamma - \lambda T^*\Big)}_{-\Delta(c')} f(c') - \lambda \int_{c^*}^{-\frac{b+T^*}{k}} f(c)dc - \lambda \int_{\underline{c}}^{c^*} f(c)dc = 0$$

$$\Delta(c')f(c') = k\lambda F(c') \tag{A.32}$$

where the last step follows from solving the integrals and using $F(\underline{c}) = 0$. Equation (A.31) shows that emissions price is distorted downwards below the first-best level $D/(1+\lambda)$ when $\lambda > 0$. Equation (A.32) shows that compensation is chosen so that it balances the marginal damages from relocation and the overcompensation paid to all the remaining firms.

Conflicting incentives, $k > 1$: Firms leave if $c = c' < \frac{b+T^*+c^*}{1-k}$, stay and cut emissions if $c' \leq c \leq c^*$ and stay without cutting emissions if $c > c^*$. The social welfare function can be written as:

$$\max_{c^*,T^*} \int_{\frac{b+T^*+c^*}{1-k}}^{c^*} \Big(\gamma + (D-c) - \lambda(T^*+c^*)\Big) f(c)dc + \int_{c^*}^{\overline{c}} \Big(\gamma - \lambda T^*\Big) f(c)dc$$

Take the first order conditions with respect to c^*:

$$\Big(D-(1+\lambda)c^*\Big) f(c^*) - \frac{1}{1-k} \underbrace{\Big(\gamma + (D-c') - \lambda(T^*+c^*)\Big)}_{-\Delta(c')} f(c') - \int_{c'}^{c^*} \lambda f(c)dc = 0$$

$$\Big(D-(1+\lambda)c^*\Big) f(c^*) = -\frac{1}{1-k}\Delta(c') f(c') + \lambda\Big(F(c^*) - F(c')\Big) \quad \text{(A.33)}$$

Take the first order conditions with respect to T^*:

$$-\frac{1}{1-k} \underbrace{\Big(\gamma + (D-c') - \lambda(T^*+c^*)\Big)}_{-\Delta(c')} f(c') - \lambda \int_{c'}^{\overline{c}} f(c)dc = 0$$

$$\Delta(c') f(c') = (1-k)\lambda\Big(1-F(c')\Big) \quad \text{(A.34)}$$

Compensation is chosen so that it balances the marginal damages when a firm of type c' moves (left hand side) and the overcompensation to all the remaining firms (right-hand side). Using (A.34), simplifying and solving for c^* we can write (A.33) as follows:

$$c^* = \frac{D}{1+\lambda} + \frac{\lambda}{1+\lambda}\frac{1-F(c^*)}{f(c^*)} \quad \text{(A.35)}$$

The emissions price is distorted upwards above the first-best level $D/(1+\lambda)$ when $\lambda > 0$.

Countervailing incentives, $0 < k < 1$**:** Firms leave when $\frac{b+T^*+c^*}{1-k} = c'' < c < c' = -\frac{b+T^*}{k}$. Firms cut emissions if $c \leq c''$ and stay but not cut when $c \geq c'$. The social welfare function can be written as:

$$\max_{c^*,T^*} \int_{\underline{c}}^{\frac{b+T^*+c^*}{1-k}} \Big(\gamma + (D-c) - \lambda(T^*+c^*)\Big) f(c)dc + \int_{-\frac{b+T^*}{k}}^{\overline{c}} \Big(\gamma - \lambda T^*\Big) f(c)dc$$

where the first integral denotes firms that cut emissions, and the second part firms that stay but do not cut. Take the first order conditions with respect to c^*:

$$\frac{1}{1-k} \underbrace{\Big(\gamma + (D-c'') - \lambda(T^*+c^*)\Big)}_{-\Delta(c'')} f(c'') - \lambda \int_{\underline{c}}^{c''} f(c)dc = 0$$

$$\Delta(c'')f(c'') = (k-1)\lambda F(c'') \tag{A.36}$$

That is, the extensive margin for cutting firms is found at the point where marginal damages when a cutting firm moves (left-hand side) are set equal to the overcompensation paid to the remaining firms that cut (right-hand side), located left from c'. Take the first order conditions with respect to T^*:

$$\frac{1}{1-k}\Big(\gamma + (D - c'') - \lambda(T^* + c^*)\Big)f(c'') - \lambda\int_{\underline{c}}^{c''} f(c)dc + \frac{1}{k}\underbrace{\Big(\gamma - \lambda T^*\Big)}_{-\Delta(c')} f(c') - \lambda\int_{c'}^{\overline{c}} f(c)dc = 0$$

The sum of two first terms is equal to zero by the first-order condition for c^*. Simplify to write:

$$\Delta(c')f(c') = -k\lambda\Big(1 - F(c')\Big) \tag{A.37}$$

Again, the optimal external margin is set by the trade-off between damages when a marginal non-cutting firm moves (left-hand side) and the overcompensation to all the mass of non-cutting firms, located right of point c' (right-hand side).

Proof of Proposition 6.

Regimes $n = \{i, j\}$ choose their environmental policies non-cooperatively. Regimes set cut-offs c_i^* and base compensation T_i^* so that the firms that cut receive $T_i + c_i^*$, firms that do not cut receive T_i. Without loss of generality, denote the regime with lower level of regulation as the regime j so that $c_j^* < c_i^*$. We find first the best response by i to such lower regulation by j. Taking (T_j, c_j^*) by regime j as given, regime i finds the optimal mechanism (T_i, c_i^*) by solving:

$$\begin{aligned}\max_{c_i^*, T_i} W_i = &\int_{\underline{c}}^{c_j^*} \Big[\gamma - c - \lambda(T_i + c_i^*)\Big]\phi\Big(T_j + c_j^* - T_i - c_i^*, c\Big)dc + \\ &\int_{c_j^*}^{c_i^*} \Big[\gamma + (D - c) - \lambda(T_i + c_i^*)\Big]\phi\Big(c - T_i - c_i^* + T_j, c\Big)dc + \\ &\int_{c_i^*}^{\overline{c}} \Big[\gamma - \lambda T_i\Big]\phi\Big(T_j - T_i, c\Big)dc\end{aligned}$$

The first line ($c \leq c_j^*$) covers the mass of firms that cut emissions in both regimes. The second line ($c_j^* < c \leq c_i^*$) covers the mass of firms that cut emissions in regime i only. The third line ($c > c_i^*$) is the mass of firms that do not cut emissions in either regime. First, use Leibniz' rule to maximize with respect to c_i^*:

$$\Big[D-(1+\lambda)c_i^*\Big]\phi(T_j-T_i,c)+\int_{\underline{c}}^{c_j^*}\left(\Delta_1(c)\phi'\Big(c-T_i-c_i^*+T_j,c\Big)-\lambda\phi\Big(c-T_i-c_i^*+T_j,c\Big)\right)dc$$
$$+\int_{c_j^*}^{c_i^*}\left(\Delta_2(c)\phi'(T_j-T_i,c)-\lambda\phi(T_j-T_i,c)\right)dc=0$$

where $\Delta_1=-(\gamma-c-\lambda(T_i+c_i^*))$ is the damage from relocation by firm c when it would cut emissions in both regimes ($X_i(c)=X_j(c)$), and $\Delta_2=-(\gamma+(D-c)-\lambda(T_i+c_i^*))$ are the damages from relocation when the regime only cuts in regime i ($X_i(c)=1,X_j(c)=0$). Similarly, the first-order condition for the optimal compensation T_i can be written as:

$$\Delta_3\phi'(T_j-T_i,c)-\lambda\phi(T_j-T_i,c)+$$
$$\frac{f(c_i^*)}{1-F(c_i^*)}\int_{\underline{c}}^{c_j^*}\Delta_1(c)\phi'(\cdot,c)-\lambda\phi(\cdot,c)dc+\int_{c_j^*}^{c_i^*}\Delta_2(c)\phi'(\cdot,c)-\lambda\phi(\cdot,c)dc=0$$

where $\Delta_3=-(\gamma-\lambda T_i)$ are the losses from relocation for firm c when neither regime implements cuts in emissions ($X_i(c),X_j(c)=0$). Similarly, we can find the best responses in a symmetric equilibrium, $c_i^*=c_j^*$ and $T_i=T_j$. The equilibrium condition for $c_i^*=c_i^*=c_j^*$ becomes:

$$X_i(c)=\begin{cases}1 & \text{if } c_i\le c_i^*=\frac{D}{1+\lambda}+\frac{\mu(c^*)}{(1+\lambda)\phi(0,c)}\\ 0 & \text{otherwise}\end{cases}$$

$$\Delta_3\phi'(0,c)=\lambda\phi(0,c)-\mu(c_i^*)$$

where
$$\mu(c_i^*)=\int_{\underline{c}}^{c_i^*}\Delta_1(\tilde{c})\phi'(0,c)-\lambda\phi(0,c)d\tilde{c}$$
with $\Delta_1(c)=-\Big[\gamma-c_i-\lambda(T_i+c_i^*)\Big]$ and $\Delta_3(c)=-\Big[\gamma-\lambda T_i\Big]$.

We are interested in the sign of the distortion term $\mu(c^*)$. To solve for the sign, we evaluate the program without the distortion (setting $\mu(c^*)=0$)) and see if that satisfies the first-order conditions. If not, we solve for the sign of the distortion. In the benchmark level, the compensation for firms with $X_i(c)=0$ is:

$$-\Big(\gamma-\lambda T_i\Big)\phi'(0)=\lambda\phi(0)$$

Plug this into $\mu(c^*)$:

$$\mu(c^*) = \int_{\underline{c}}^{c^*} -\Big[\gamma - c_i - \lambda(T_i + c_i^*)\Big]\phi'(0,c) - \lambda\phi(0,c)dc$$

$$\mu(c^*) = \int_{\underline{c}}^{c^*} -\Big((D-c) - \lambda c^*\Big)dc < 0$$

Where the last inequality follows from the fact that $c^* = D/(1+\lambda)$ for $c = c^*$ and $c < D/(1+\lambda)$ for all $c < c^*$. It follows that we must have downward distortion at the equilibrium. Given the action of regime i, regime j best response is to stick to the conjectured symmetric action, establishing that there is an equilibrium in symmetric strategies. Q.E.D.

Proof of proposition 7.

Next, we will extend the analysis to the more general case where the damages from emissions may be convex. We will assume a damage function $D(X)$ ($D'(\cdot) > 0$, $D''(\cdot) > 0$) where $X = \hat{X} - \int_{\underline{c}}^{\overline{c}}(\phi(C(c)X(c))dc$ depicts the total emissions. The first term $\hat{X}$ is the counterfactual pollution (without emission reduction) and the second term in D is the emission reduction by the regime. Social welfare function now becomes:

$$\max_{X(c),T(c)} \int_{\underline{c}}^{\overline{c}} \Big(\gamma - cX(c) - \lambda T(c)\Big)\phi(C(c),c)dc + D(X) \tag{A.38}$$

By Lemmas 1-3 we know that without loss of generality we can focus on threshold policies where $X(c) = 1$ for $c \le c^*$ and $X(c) = 0$ otherwise. Firms that cut emissions are paid $T + c^*$. Firms that do not cut receive T. When damages are convex, the problem can be written as a choice of the threshold c^* and a payment T to all the firms. Maximization problem corresponding (A.3) can now be written as:

$$\begin{aligned}\max_{c^*,T} W = &\int_{\underline{c}}^{c^*} \Big(\gamma - c - \lambda(T + c^*)\Big)\phi(c - T - c^*)dc + \int_{c^*}^{\overline{c}} \Big(\gamma - \lambda\Big)T\phi(-T)dc \\ &- D\Big(\hat{X} - \int_{\underline{c}}^{c^*} \phi(c - T - c^*)dc\Big)\end{aligned} \tag{A.39}$$

Solve FOC with respect to c^* using the Leibniz's rule:

$$\Big((D'(\hat{X}) - c^*) - \lambda(T + c^*)\Big)\phi(-T) + \underbrace{\int_{\underline{c}}^{c^*} \Big(\Delta_1(c)\phi'(c - T - c^*)\Big) - \lambda\phi(c - T - c^*)dc}_{\mu(c^*)} \tag{A.40}$$

where $\Delta(c) = -(\gamma + (D'(\hat{X}) - c) - \lambda(T + c^*))$ for firms with $c \le c^*$ cutting emissions. Next, solve with respect to $T^*(c)$:

$$-\int_{c^*}^{\bar{c}}\Big(\Delta_2(c)\phi'(-T)-\lambda\phi(-T)\Big)dc-\underbrace{\int_{\underline{c}}^{c^*}\Big(\Delta_1(c)\phi'(c-T-c^*)-\lambda\phi(c-T-c^*)\Big)dc}_{\mu(c^*)}=0 \quad \text{(A.41)}$$

where $\Delta(c) = -(\gamma - \lambda T)$ for firms $c > c^*$ not cutting emissions. We see that the optimal policy remains similar: Equations (A.40)-(A.41) are identical to equations (A.4)-(A.6), with one difference: the marginal damage term $D'(\hat{X})$ replaces the constant marginal damage D in Proposition 2.

Proof of Proposition 8.

Unlike in the main section, the optimal solution no longer takes a bang-bang form (Lemma 3 does no longer hold). The incentive compatibility can be written as $C'(c) = A_c(X(c), c)$, holds (proof is standard and is omitted, see for instance Baron and Myerson (1982)). Here, we focus on the cases where full separation is optimal, that is, where the non-monotonicity condition for $X(c)$ does not bind. Hamiltonian for the problem is:

$$\mathcal{H}=\Big(\gamma-(1+\lambda)A(X(c),c)+\lambda C(c)\Big)\phi(C(c),c)-D\Big(1-\phi(C(c),c)X(c)\Big)-\mu(c)A_c(X(c),c) \quad \text{(A.42)}$$

where $\mu(c)$ is the co-state variable of the incentive compatibility constraint. We assume that $X(c)$ is differentiable. Using Pontryagin's principle, the necessary conditions for the optimum are:

$$\Big(\gamma+D-(1+\lambda)A_x(X(c),c)+\lambda C(c)\Big)\phi(C(c),c)-\mu(c)A_{xc}(X(c),c)=0 \quad \text{(A.43)}$$

$$\mu'(c)=\Delta(c)\phi'(C(c))-\lambda\phi(C(c),c) \quad \text{(A.44)}$$

$$\mu(\underline{c})=0 \quad \text{(A.45)}$$

Here $\Delta(c) = -(\gamma + DX(c) - (1+\lambda)A(X(c), c) - \lambda T(c))$ denote the net losses from relocation. From (A.43) we can solve:

$$A_x(X(c),c)=\frac{D}{1+\lambda}+\frac{\mu(c)}{(1+\lambda)\phi(C(c),c)}A_{xc}(X(c),c) \quad \text{(A.46)}$$

Integrating over (A.44), and fixing the lower bound by using the transversality condition (A.45), we get:

$$\mu(c)=\int_{\underline{c}}^{c}\Big(\Delta(\tilde{c})\phi'(C(\tilde{c}),\tilde{c})-\lambda\phi(C(\tilde{c}),\tilde{c})\Big)d\tilde{c} \quad \text{(A.47)}$$

where $\Delta(X(c), T(c), c) = -\Big(\gamma + DX(c) - A(X(c), c) - \lambda T(c)\Big)$ is the net loss of relocation by firm of type c.

B Appendix: Extensions to the calibration

We carry out two extensions to the numerical quantification in order to see how sensitive our results are to changes in the key parameters; specifically on the assumptions of industry-specific fixed values γ and the social cost of public funds λ. We begin by assuming that the regulator cares only about environmental damages and puts no direct value on industries ($\gamma = 0$). This sensitivity analysis helps us to distinguish the effect of "carbon leakage" from regulator's motives to support the industry for other reasons. The results are shown below in Table 3. Three things follow from this analysis. First, expectedly, all industries are being heavily taxed as the net losses from relocation become smaller. Second, the effective CO_2 price is distorted heavily upwards, (35-42 %) above the no-leakage benchmark and even above the observable-c benchmark for all the sectors except for Wood and Paper. This result is in line of the sensitivity analysis, where it was found that a lower value of γ increases the optimal level of regulation. Intuitively, the regulator wants to compensate low-cost firms for emission reduction and tax the high-cost firms, and this is done by choosing a high emissions price. Third, the trade-off between leakage and overcompensation is tilted in favor of the latter: As firms' relocation becomes less bad, more emphasis is given on avoiding windfall profits.

Table 3: The optimal policy without industry-specific direct benefits

	Compensation per action			Industry-level	
	Cut (€/tCO_2)	Pollute (€/tCO_2)	Effective CO_2 price (€/tCO_2)	Optimal leakage	Optimal windfalls (M€)
Cement	-14.6	-30.7	16.1	53.4 %	550
Iron and Steel	-5.9	-21.9	16.0	52.1 %	768
Chemical and Plastic	-20.6	-36.1	15.7	39.7 %	355
Wood and Paper	-27.7	-43.0	15.3	35.1 %	71
Glass	-39.0	-54.7	15.6	24.4 %	45
Benchmark: observable c			15.6		
Benchmark: no leakage			11.3		

Notes: Optimal compensations in EU sectors and the implied marginal tax rate on emissions when the social cost of carbon is 25 €/tCO_2; the social cost of public funds is $\lambda = 0.6$; and $\gamma = 0$ for all the sectors.

In the next extensions, we analyze the effect of choosing a low social cost of public funds. For this analysis, we choose a value $\lambda = 0.2$. As the results in Table 4 show, the firms are compensated heavily, expectedly so, because the regulator values public funds less. In all of the sectors, except for Wood and Paper, even the polluting firms receive compensation. The effective CO_2 price is distorted upwards 7 - 11 % compared to the benchmark with no leakage; this distortion is relatively smaller than in the main analysis with higher social cost of public funds ($\lambda = 0.6$). Because the regulator cares less about overcompensating the firms and puts an emphasis on preventing leakage, the optimal leakage is less than 10 % for all the sectors, but the total windfalls under the optimal policy become as high as 11.4 billion Euros for the five key sectors.

Table 4: The optimal policy with a low social cost of public funds

	Compensation per action			Industry-level	
	Cut (€/tCO_2)	Pollute (€/tCO_2)	Effective CO_2 price (€/tCO_2)	Optimal leakage	Optimal windfalls (M€)
Cement	28.3	8.4	19.9	8.6 %	3,299
Iron and Steel	37.1	18.0	19.1	2.0 %	4,559
Chemical and Plastic	31.7	12.5	19.2	1.0 %	2,750
Wood and Paper	16.5	-2.8	19.3	2.3 %	612
Glass	25.2	5.5	19.7	2.5 %	345
Benchmark: observable c			20.8		
Benchmark: no leakage			17.9		

Notes: Optimal compensations in EU sectors and the implied marginal tax rate on emissions when the social cost of carbon is 25 €/tCO_2; sector-specific γ are as in Table 1 and the social cost of public funds is $\lambda = .2$.