CEEPR
MIT Center for Energy and Environmental Policy Research

Working Paper Series

Global Externalities, Local Policies, and Firm Selection

Lassi Ahlvik and Matti Liski

DECEMBER 2021

CEEPR WP 2021-019

Working Paper Series.

Global externalities, local policies, and firm selection

Lassi Ahlvik and Matti Liski*

October 18, 2021

Abstract

How to fight global problems with local tools? When only firms know what externality-producing activities can be relocated, policies shape the location distribution of firm types with different social values. We find that, because of this selection effect, the optimal local policies confront firms' mobility with elevated corrective externality prices, in contrast with the common remedies for the relocation risk. Our mechanism incentivizes also moving firms to limit the externality, and it influences strategically the distribution of moving firms that comply with policies elsewhere. The magnitude of these effects is illustrated by a quantification for the key sectors in the EU emissions trading system.

Keywords: Externalities; mechanism design; private information; climate change; emissions trading; carbon leakage
JEL codes: D82; L51; Q54; Q58

*Ahlvik (lassi.ahlvik@helsinki.fi): Department of Economics and Management at the University of Helsinki. Liski (matti.liski@aalto.fi): Economics department of Aalto University. We thank seminar participants at presentations in Athens (EAERE), Bergen, Gothenburg, Helsinki, Paris, Oslo, Torino (IEARE keynote), Toulouse (11th TSE conference on climate and energy), and Zurich for valuable comments and discussion. We also thank Chang-Koo Chi, Baran Doda, Bård Harstad, Daniel Hauser, Torfinn Harding, Johannes Hörner, Yuanhao Li, Jussi Lintunen, Juan-Pablo Montero, Matias Pietola, Knut Einar Rosendahl, Julia Salmi, Thomas Sterner, Morten Sæthre, Juuso Välimäki, Ulrich Wagner, Po Yin Wong, and Andrey Zhukov for valuable comments and discussions. Matti is grateful for funding from the Academy of Finland program New Energy, and from the Aalto University project SAGA.

1 Introduction

How to fight global problems without hurting the local welfare? Economists are increasingly confronting this question: Whether it relates to financial sector regulation, virus outbreaks, labor market standards, or cross-border pollutants, policy makers are often left with only local tools for dealing with global spill-overs. Local policies are commonly opposed on the grounds that policies force businesses out to non-regulated regimes, thereby undermining their effectiveness. For example, what is the benefit of a stricter capital requirement on a bank if, after its cross-border relocation, the systemic risk remains the same?

Environmental regulation is a particularly prominent case. The U.S. Congress passed a resolution opposing a carbon tax on the basis that, among other things, it *"will lead to more jobs and businesses moving overseas"* (H. Con. Res. 119, 2018). In the European Union, industries have argued that, in the absence of a global climate policy, strengthening the Emissions Trading Scheme would force businesses to leave *"without any environmental need"* (Fagan-Watson, 2015). In response to such concerns, policies routinely compromise on the externality price: Rebates of environmental taxes are used to subsidize energy-intensive industries,[1] emissions trading regimes allow the use of cheaper offsets for selected firms or industries,[2] and a threat of relocation is used as a reason to exclude entire sectors from regulations.[3].

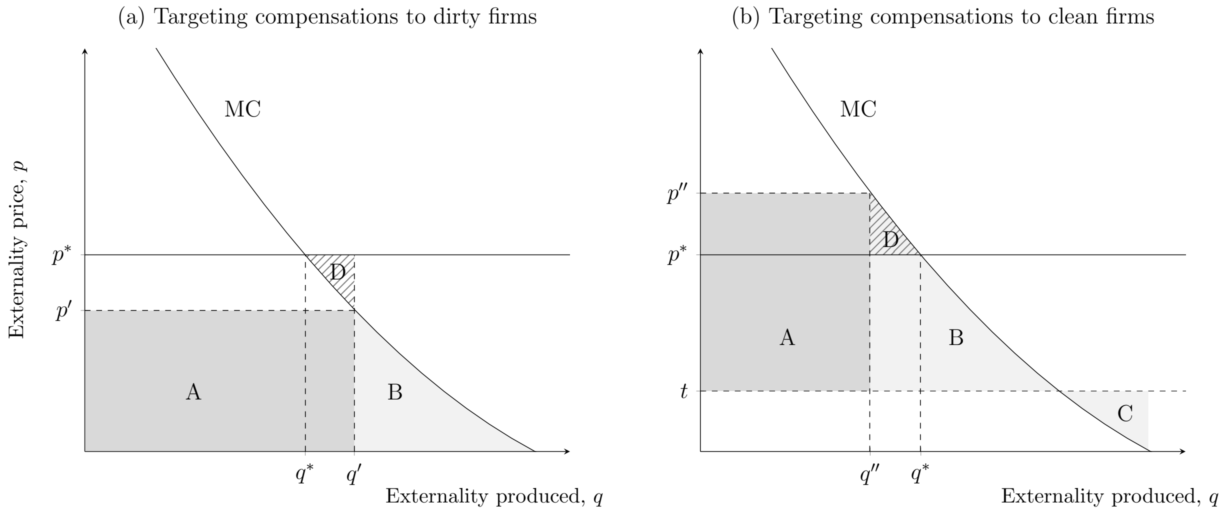
This paper argues that such common policy responses to industry relocation are misguided if the policy maker is armed not only with externality prices but also with transfers. For global problems, some firms can do more at home than others and are thus more valuable to keep: Transfers from scarce public funds should reach those firms first. But because firms' available options are privately known, the policies must incentivize firms to self-select

[1]In the Nordic countries, known for their high CO_2 taxes, energy-intensive industries commonly receive rebates of carbon and energy taxes, see Bragadóttir *et al.* (2014). Rebates curb the effective carbon prices because they are only received by the firms paying the tax.

[2]In the Trudeau government's recent carbon pricing initiative in Canada, emissions-intensive and trade-exposed facilities are regulated by the Output-Based Pricing System which allows the use of offsets (see Environment Canada 2018). Selected firms in the EU Emissions Trading System were given the option to use cheap international credits generated through the Kyoto Mechanisms, up to a polluter-specific percentage, for compliance (EC, 2019b). As offsets are valueless to low-cost firms, their use effectively distorts the carbon prices.

[3]A case in point is the international aviation sector that is excluded from the emission pricing, apart from flights within the European Economic Area (EEA) included in the EU Emissions Trading System since 2012. Airlines are receiving 95% of their historical emissions as free allowances. The current plan is to include also flights to and from EEA in 2024 (Larsson *et al.*, 2019)

Figure 1: Illustration of the selection effect



Notes: The graph plots marginal cost (MC) for limiting the externality when the externality price is (a) decreased ($p' < p^*$), leading to a higher externality level ($q' > q^*$) or (b) increased together with compensation ($p'' > p^* = p'' - t$), leading to a lower externality level ($q'' < q^*$). Area A: compliance cost for firms that pay the externality price. Area B: compliance cost for firms that eliminate the externality. Area C: benefits to the low-cost compliant firms. Area D: deadweight loss.

the desired action and location. This selection effect calls for higher externality prices, not lower.

Figure 1 illustrates this idea. The marginal cost (MC) curve captures the cost of reducing the externality by aggregating unit costs over small individual firms or plants; it increases to the left starting from the unconstrained externality level. Consider externality price p^* deemed optimal in the absence of firm relocation. High-cost firms pay this price and remain dirty, while firms with costs lower than p^* eliminate the externality. When all firms have some risk of moving, choosing externality price $p' < p^*$, as shown in Figure 1a, lowers the compliance cost of dirty firms (area A) and incentivizes them to stay, while creating a deadweight loss (area D). Yet, the location of these dirty firms is irrelevant for the global problem as they produce the externality regardless of their location. Therefore, the rollback of the externality price, as in Figure 1a, can never be justified by the global externality problem alone. In contrast, the problem calls for targeting compensations to low-cost firms that can limit the externality at home, as in Figure 1b: Choosing a higher externality price $p'' > p^*$ accompanied by lump-sum compensation $t = p'' - p^*$ would reduce the cost to the clean firms (area B) and even make regulation profitable for some firms (area C), without affecting the cost for the inframarginal dirty firms (area A).

Conceptually, the combination of global externalities and mobile firms presents a novel mechanism design problem where the principal cares about the agents' types and actions even when they choose not to participate in the mechanism. As standard in mechanism design approach, the policy maker is constrained by asymmetric information, not by the set of available instruments; in our setting, the two-part tariff (in Figure 1) is indeed the optimal local mechanism. In standard incentive mechanisms, however, the actions of non-participating agents have no direct bearing on the principal's welfare. For example, it would be absurd to offer contracts to privately-informed customers who choose not to buy a product (Mussa and Rosen, 1978; Rochet and Stole, 2002). In contrast, with a global externality, the moving agents' actions continue to matter for welfare, and thus shaping the actions of these agents becomes part of the design problem. The action taken in another location depends on the firm's type, so the designer cares not only about firms 'kept' but also those 'lost' to other locations.

The policy makers value public funds, creating a taxation motive that is common from the literature (Lewis, 1996). If firms are immobile, then they are taxed to raise public funds and a lower than socially optimal (Pigouvian) price for the externality is chosen: This allows for a higher general tax base, thereby raising more revenues from all firms, including those that become clean. In contrast, when firms can move, the self-selection of staying firms justifies moving the tax burden towards dirty firms, following the logic outlined in Figure 1 above. This in itself always increases the local externality price, and, under certain conditions, raises it even above the Pigouvian first-best levels. These findings are in sharp contrast with the general line of results from the incentive regulation theory where the downward distortion in the regulatory stringency is the theory's bread and butter (Laffont and Tirole, 1993).

It is generally optimal to offer regulatory contracts also to the relocating firms. With information on firms' compliance costs, it turns out, it would be optimal to impose a single global (Pigouvian) externality price for the staying and leaving firms, and influence location choices by transfers. Typically, however, such information is not available and the incentive problems in the optimal mechanism break the result of a uniform global price. We show that, when sovereignty of jurisdictions prevents taxation of firms in other locations, the optimal policy is neither purely local nor purely global but always implements two distinct externality prices: A higher local price for firms that stay, and a lower global price at which reductions are bought from firms that relocate.

The firm selection effect, our main result, calls for elevated externality prices when firms can move to avoid policies. But what if also the alternative locations have some policies in place? In fact, more than 20% of global carbon emissions are currently subject to some form of carbon pricing,[4] resulting from unilateral policies in the absence of a binding cooperative climate treaty. We analyze such a bottom-up approach, where countries choose strategically their incentive mechanisms for mobile firms. Rather surprisingly, in contrast with the pure unilateral policy design, it follows that in a symmetric equilibrium the externality price can never rise above the Pigouvian level. If the alternative location has some externality price in place, it becomes less important to subsidize the clean firms at home.

Last, we provide an illustrative quantification of the optimal carbon leakage policy for the key sectors in the EU emissions trading system (EU ETS) based on the firm-level data on relocation propensities from Martin *et al.* (2014a). The data allow us to draw representative relocation risk distributions for five sectors forming together 62% of the industry emissions covered by the trading program. With representative values for the social cost of carbon emissions and public funds, we quantify the optimal mechanism with results on carbon leakage, distortions in the emissions price, and the fraction of the sectoral cost that is optimally covered from public funds. The main theoretical results turn out to be also economically significant. The optimal local carbon prices are increased upwards by $17-29$ per cent compared to the benchmark without firm relocation. The optimal global price is generally less than a third of the local price. The higher carbon prices also translate into larger cuts, even after the leakage of emissions ($2-17$% per sector) is taken into account: The threat of relocation, in itself, calls for 9.6 $MtCO_2$ additional emission reductions (13% higher than in the benchmark, an amount roughly equal to total manufacturing emissions in Sweden), and the optimal global mechanism supplements this by reducing additional 1.2 $MtCO_2$ abroad (2% compared to the benchmark). Finally, in this quantification, the outcome is more or less unaffected if we restrict attention to mechanisms that set a uniform externality price for all sectors but keep the transfers differentiated.

Literature. Our study contributes to the literature on environmental regulation under privately informed polluters, but the combination of global externalities and mobile firms introduce new aspects to the classical problem. The polluter's privately observed opportunity to move introduces a participation constraint that is different from those typically analyzed in the literature, including zero-profit conditions (Spulber, 1988; Kim and Chang, 1993) or voluntary participation of firms (Lewis, 1996; Montero, 2000; van Benthem and Kerr,

[4] Source: https://carbonpricingdashboard.worldbank.org/map_data

2013). All such participation constraints act similarly in that they only bind for high-cost polluters who are left with no information rents. Therefore, the stringency of the regulation (intensive margin) is distorted to limit information rents. In contrast, in our setting with global externalities and privately known relocation costs, policy-driven information rents have social value as such because they act as targeted compensations to those who take actions; in addition, policy-driven relocation (extensive margin) is a tool for limiting the rents of those who don't take actions.[5] Conceptually, the setting leads to a self-selection model with random participation, but the main results, upward distortion in the regulatory stringency and incentives to non-participating agents, do not arise in applications to firm competition (Rochet and Stole, 2002) or optimal income taxation (Lehmann *et al.*, 2014).

The global externality problems have inspired economists to look for Pareto improving regulations that extend beyond national borders; for instance, van Benthem and Kerr (2013) discuss the design of an optimal international offset program, and Harstad (2012) shows that climate-friendly countries benefit from buying and conserving foreign fossil-fuel deposits. While our global mechanism is similar in spirit, it turns out that firm relocation with private information leads to a policy problem with unique features: Regulation creates information rents at home and abroad, and the mechanism needs to optimally manipulate the rents mobile firms can expect by moving.

Our study is the first to take a mechanism design approach in the literature on the so-called "carbon leakage" problem. Given that firms' private information on the relocation propensity is an indisputably essential feature of the problem, it is surprising that it has received little attention to date.[6] Most of the earlier literature has restricted the set of policy instruments at one's disposal to carbon taxes (Markusen *et al.*, 1993; Motta and Thisse, 1994; Hoel, 1997; Ulph and Valentini, 1997; Petrakis and Xepapadeas, 2003; Greaker, 2003). A few studies have focused on limiting firm relocation with lump-sum compensations as the only instrument (Schmidt and Heitzig, 2014; Martin *et al.*, 2014a). With such limitations, the policy maker is forced to solve two problems, managing both externalities and relocation, with one instrument so the outcome depends on how exactly the available policy instruments are introduced. By taking a mechanism design approach, we avoid ad-hoc restrictions on

[5] The latter channel of limiting information rents is similar to excluding consumers from using a public good (Hellwig, 2003; Norman, 2004), or preventing natural monopolies from serving a market (Baron and Myerson, 1982).

[6] Greaker (2003) and Martin *et al.* (2014a) note the information problem but leave the mechanism design problem open for future research.

the set of admissible policies; rather, the policy maker is left only with constraints stemming from the private information held by firms.

2 The set-up

Consider a continuum of firms with unit mass, each characterized by cost $\beta \in [\underline{\beta}, \overline{\beta}]$ (with $\underline{\beta} \geq 0$) of reducing one unit of a negative externality, which we refer to as emissions.[7] We take the viewpoint of country i, and the alternative location is denoted by j. The mass of firms in location i (in location j, resp.) is characterized by density distribution function $\phi_i(C(\beta), \beta)$ that depends on firm's type β and also on cost difference $C(\beta) = C_i(\beta) - C_j(\beta)$, where $C_k(\beta)$ is the cost for firm of type β in location $k = i, j$.

The policy maker in i chooses a mechanism, denoted by $M_i(\beta)$, implementing, as explained below, costs and actions for each firm type β. Formally, $M_i(\beta) = \{C_i(\beta), X_i(\beta)\}$, where $X_i(\beta) \in [0, 1]$ is the fraction of firm's externality reduced, and $C_i(\beta) = \beta X_i(\beta) - T_i(\beta)$ is the net cost of compliance in location i that depends on transfer $T_i(\beta)$. $M_j(\beta) = \{C_j(\beta), X_j(\beta)\}$ captures multiple interpretations, with the main ones being: location j is a pollution haven ($X_j(\beta) = 0$) that may attract firms with subsidies (Section 3); or mechanism $M_j(\beta)$ might be offered by home location i to attract voluntary participation in j (Section 4); or a setting where locations (i, j) play a policy design game (Section 5.1).

Local policy maker in i cares about the *local* welfare impacts of *global* emissions, firms' value-added at home, and also the costs of transferring public funds to the firms. The payoff function, W_i, captures these elements through avoided damages per unit of pollution, $D > 0$, firms' location-specific value-added, $\gamma \geq 0$, and the cost of public funds, $\lambda > 0$:

$$W_i = \int_{\underline{\beta}}^{\overline{\beta}} \Big(\gamma + DX_i(\beta) - C_i(\beta)\Big)\phi_i(C(\beta), \beta) + DX_j(\beta)\phi_j(C(\beta), \beta) - (1 + \lambda)T(\beta)d\beta, \quad (1)$$

where $T(\beta)$ is the total transfer: $T(\beta) = T_i(\beta)\phi_i(C(\beta), \beta) + T_j(\beta)\phi_j(C(\beta), \beta)$.[8] It is useful

[7] What we define as a "firm" can be interpreted more broadly as a unit of production such that abatement costs are independent and identically distributed across production units that can be relocated individually. A real-world company can therefore consist of several of such production units and reduce multiple units of pollution with increasing marginal costs. We consider a model of convex costs within each "unit" in Section 5.3.

[8] Our main analysis assumes linear externality damages and purely global externalities, but these assumptions are not critical for our results as we show in Sections 5.3 and 5.4. Moreover, our setting abstracts away from distributional goals and "fairness" considerations, such as those analyzed by Harstad and Eskeland

to define the *welfare effect of relocation* as the change in social welfare at i when a firm of type β relocates to j:

$$\Delta(M_i(\beta), M_j(\beta), \beta) = -\Big(\gamma + D\big(X_i(\beta) - X_j(\beta)\big) - C_i(\beta) - (1+\lambda)\big(T_i(\beta) - T_j(\beta)\big)\Big), \quad (2)$$

or $\Delta(\beta)$ for shorthand. Some insights follow from just observing this definition. First, if a firm cuts emissions in neither regime ($X_i = 0, X_j = 0$), there is no "leakage" of pollution when a firm moves although relocation may still be socially undesirable due to loss of value-added γ and the firm's possible contribution to public funds. Second, a firm that cuts emissions only when staying ($X_i = 1, X_j = 0$) creates surplus $D - \beta$. Another key observation is that a firm's contribution to the social welfare depends on its privately known cost parameter. Third, when a firm cuts the same in both locations ($X_i = 1, X_j = 1$), relocation has no effect on the global externality but the firm's social value still depends on its costs β: All else equal, the local policy maker prefers to keep firms with low costs.[9]

We micro-found the location distribution of firms, $\phi_k(C(\beta), \beta)$, $k = i, j$, by assuming, first, that firms have two-dimensional private information and, second, by explaining how the policy maker can infer the location distribution from the cost difference, $C(\beta)$.

First, in addition to the emission reduction cost, β, each firm has specific privately-known relocation cost, $\theta \in (-\infty, \overline{\theta}]$, that creates a preference for location i. Finite $\overline{\theta} > 0$ puts a limit on how much firms can be taxed at home and ensures that the problem is well-defined.[10] The density and the cumulative distribution for the relocation costs are $g(\theta)$ and $G(\theta)$, respectively.[11] The distribution of abatement costs follows a continuous density function $f(\beta)$, with $F(\beta)$ denoting the respective cumulative distribution. We make the standard

(2010).

[9]Note that the firm's value added, γ, does not vary in its abatement cost β. If high-cost firms had systematically lower ($\gamma'(\beta) < 0$) or higher ($\gamma'(\beta) > 0$) value added, the preference for keeping low-cost firms would be reinforced or reduced, respectively.

[10]Some mass of firms, however small, has extreme $\theta < 0$ and will always move, which leads to interior outcomes. The assumption is reconcilable with all firms having positive relocation cost when the firms that end up in i come from both i and j (see fn. 13). While relocation cost is the most natural interpretation of θ, it could be more generally seen as the firm-specific preference for a particular location. For instance, it captures the cost of physically moving and the expected decrease in profits due to choosing another location. Our model thus incorporates both firms actively relocating existing production units to other regimes (analyzed by Martin *et al.* 2014a) and investment leakage when regulation causes multinational firms to expand into another location for new production (analyzed by Hanna 2010). The latter case would be captured by a timing in which firms choose their first locations only after the policies set by countries. We assume that cost θ is paid after the firm chooses an alternative location; it does not directly enter country i's welfare in equation (1).

[11]Here, we assume no correlation between relocation and compliance costs $G(\theta|\beta) = G(\theta)$ but in our work-

regularity assumptions on the distributions:

Assumption 1. *Distribution $F(\beta)$ satisfies the hazard rate assumption*[12]

$$\frac{d}{d\beta}\frac{F(\beta)}{f(\beta)} \geq 0 \geq \frac{d}{d\beta}\frac{1-F(\beta)}{f(\beta)},$$

and likewise for distribution $G(\theta)$.

Second, we argue that one can focus on β as the primitive type of each firm. Denote $C_i(\beta, \hat{\beta})$ from reporting some $\hat{\beta}$ for type β in mechanism i (respectively in j); in actuality, firms do not have to "report" but, as is usual, the direct approach facilitates exposition. The total cost depends on the required action and compensation received: $\beta X_i(\hat{\beta}) - T_i(\hat{\beta})$, where the transfer can also be negative. An incentive-compatible mechanism gives truthful reporting:

$$\beta = \arg\min_{\hat{\beta}} \left\{\beta X_i(\hat{\beta}) - T_i(\hat{\beta})\right\} \text{ for all } \beta, \tag{3}$$

defining $C_i(\beta, \beta) = C_i(\beta)$ as the net cost of compliance at location i. Thus, effectively, we can write $M_i(\beta) = \left\{T_i(\beta), X_i(\beta)\right\}$. In addition to the net costs of compliance in each location, the decision to move is also affected by the firm's relocation cost θ. The firm chooses to report in country i if:

$$C_i(\beta) \leq \theta + C_j(\beta). \tag{4}$$

Importantly, the relocation cost enters this condition additively and $X_i(\beta)$ cannot be used to screen on θ. Because the relocation decision is binary, firms' direct report of θ is not needed because the policy maker can infer the mass of firms that stay from equation (4). This approach follows Rochet and Stole (2002) and rules out randomized treatments: If, additionally, firms were asked to report $\hat{\theta}$, the regulator could arrange a lottery to treat differently firms reporting the same types (see also Rochet and Stole 2003). Intuitively, the optimal mechanism could screen on θ by randomly assigning permits to operate in the region of interest, with probabilities depending on reports. This threat of exclusion could be used

ing paper version we extend to correlated private information. Another key assumption is that information is purely private, that is, firms have no information about each other which is not available to the policy maker. This prevents the use of mechanisms such as those studied by Cremer and McLean (1988), Varian (1994) and Duggan and Roberts (2002).

[12] The condition on the hazard rates is standard (Jullien, 2000) and satisfied for a long list of commonly used distributions (Bagnoli and Bergstrom, 2005).

for limiting information rents that firms have in the relocation dimension. Our approach buys tractability and is without loss of generality if discrimination of observationally similar firms' is not allowed; quite realistically, the same action taken by the firms leads to the same treatment received.

This microstructure gives the mass of firms of type β in locations i and j:[13]

$$\begin{aligned}\phi_i(C(\beta),\beta) &= \big(1-G(C(\beta))\big)f(\beta)\\ \phi_j(C(\beta),\beta) &= G(C(\beta))f(\beta)\end{aligned} \tag{5}$$

We use $\phi_i'(C(\beta),\beta)$ as shorthand for $d\phi_i(C(\beta),\beta)/dC(\beta)$ and define the inverse hazard rate as

$$\eta(C(\beta)) \equiv \frac{\phi_i(C(\beta),\beta)}{\phi_i'(C(\beta),\beta)}.$$

Note that $\eta(C(\beta))$ is negative and increasing in $C(\beta)$ and, because $f(\beta)$ cancels out in its definition, η only depends on type β through $C(\beta)$. We make one final assumption that allows focusing on interior outcomes:

Assumption 2. *The upper and lower bounds of distribution $F(\beta)$ satisfy:*

$$D-(1+\lambda)\underline{\beta} > 0 > D-(1+\lambda)\overline{\beta}.$$

3 Local mechanism for global externalities

We begin with the optimal local mechanism, where the policy maximizes welfare (1) such that equations (3) and (5) hold, and $T_j(\beta)=0$ for all β. The last constraint immediately implies that the location j is a "pollution haven", as no reductions are incentivized and thus not made in j, $X_j(\beta)=0$.[14] The optimal local policy in i comes down to deciding how much and which firms should limit the externality, and how much of the private cost of regulations

[13] Note that this setting incorporates the case where countries have different distributions for relocation costs: $G_i(C(\beta))$ in i and $G_j(-C(\beta))$ in j. In that case, the mass of firms ending up in i is $(1-G_i(C(\beta)))f(\beta)+G_j(-C(\beta))f(\beta)$. By interpreting $G(C(\beta)) \equiv G_i(C(\beta)) - G_j(-C(\beta))$, the net leakage, we get equation (5).

[14] Even though location j does not limit externalities, it could be active in attracting firms by offering subsidy T_j', not conditional on actions. Then, the relocating firms would be the ones with $\theta \le C_i(\beta)+T_j'$, where transfer T_j' is received by the firms that relocate. We can suppress T_j' by interpreting θ' as the net relocation cost $\theta - T_j'$. With this interpretation, the moving firm margin becomes simply $\theta' = C_i(\beta)$, implying no material change in the analysis; a "pollution haven" does not change the standard participation constraint in equation (4).

are covered from the public funds.

We show in the Appendix that the optimal local mechanism takes a simple form. All staying firms will receive base compensation T_i^* that becomes a tax if it is negative. Firms with costs below the threshold, $\beta \leq \beta_i^*$, cut emissions ($X_i(\beta) = 1$) and firms above the threshold, $\beta > \beta_i^*$, pollute ($X_i(\beta) = 0$) and pay an additional β_i^*, which can be interpreted as the emission price.

Lemma 1. ***(Two-part tariff)*** *Optimal local mechanism $M_i(\beta)$ sets two constants (T_i^*, β_i^*):*

$$\begin{cases} T_i(\beta) = T_i^*, & X_i(\beta) = 1 \quad \text{for } \beta \leq \beta_i^* \\ T_i(\beta) = T_i^* - \beta_i^*, & X_i(\beta) = 0 \quad \text{for } \beta > \beta_i^*. \end{cases} \tag{6}$$

To understand how much a marginal increase in the compensation to a firm of type β is worth, we define *marginal surplus* $MS_i(C(\beta), \beta) \equiv \Delta(\beta)\phi_i'(C(\beta), \beta) - \lambda\phi_i(C(\beta), \beta)$. This term represents the usual trade-off between efficiency and rent-extraction: Compensating more at i incentivizes firms to stay and the country gains surplus $\Delta\phi_i'$, but this increases the mass of firms receiving compensations $-\lambda\phi_i$. These marginal surpluses from all firms

$$\mu_i(\underline{\beta}, \overline{\beta}) = \int_{\underline{\beta}}^{\overline{\beta}} MS_i(C(\beta), \beta) d\beta, \tag{7}$$

guide the optimal base transfer, T_i^*. At optimum, small changes in T_i^* should not lead to welfare gains: $\mu_i(\underline{\beta}, \overline{\beta}) = 0$. By this, the base transfer optimally balances the marginal surpluses from those firms that limit the externality and those that do not, $\mu_i(\underline{\beta}, \beta_i^*) + \mu_i(\beta_i^*, \overline{\beta}) = 0$. In fact, all staying firms that do not cut ($\beta > \beta_i^*$) look alike as they contribute the same to the welfare, and thus the marginal surplus from them becomes just

$$\mu_i(\beta_i^*, \overline{\beta}) = MS_i(C(\beta_i^*), \beta_i^*) \frac{1 - F(\beta_i^*)}{f(\beta_i^*)}. \tag{8}$$

The second part of the tariff is the externality price, β_i^*. It is set by the trade-off

$$\Big(D - (1 + \lambda)\beta_i^*\Big)\phi_i(C(\beta_i^*), \beta_i^*) = \mu_i(\beta_i^*, \overline{\beta}), \tag{9}$$

where the left side gives the marginal social gain from increasing the threshold for cuts, β_i^*, and the right side is the marginal surplus from all $\beta > \beta_i^*$: All else equal, a unilateral increase in the externality price leads to higher cost burden for polluting firms and thus to more relocation. However, the optimal base transfer adjusts for this through $\mu_i(\underline{\beta}, \beta_i^*) = -\mu_i(\beta_i^*, \overline{\beta})$ and thus links the choice of β_i^* to the marginal surpluses from both sets of firms. This interconnection between externality price β_i^* and base transfer T_i^* is the key in our analysis.

Two benchmarks are useful for comparison. First, if firms' β was known, the Planner would set the efficient Pigouvian cut-off, $\beta_i^* = \beta_P = D/(1+\lambda)$, and ask reductions from all $\beta \leq \beta_P$.[15] Note that when private resources are spent on reductions instead of taxes the actions are costly in terms of public funds, explaining term $1+\lambda$. The second benchmark is one where firms are completely immobile. This can be captured by a degenerate distribution with all mass at $\theta = \overline{\theta}$, and $C(\beta) \leq \overline{\theta}$. Let us denote the optimal externality price in this benchmark case by β_B.

Theorem 1. ***(Local Mechanism)*** *Optimal $M_i(\beta)$ is characterized by (T_i^*, β_i^*) where*

(i) *For a degenerate distribution $G(\theta)$ where all mass is at $\theta = \overline{\theta}$ (immobile firms), $T_i^* = \beta_i^* - \overline{\theta} \equiv T_B$*

$$\beta_i^* = \frac{D}{1+\lambda} - \frac{\lambda}{1+\lambda}\frac{F(\beta_i^*)}{f(\beta_i^*)} \equiv \beta_B < \beta_P$$

(ii) *For a non-degenerate distribution $G(\theta)$ where $\theta \in (-\infty, \overline{\theta}]$ (mobile firms), T_i^* is given by $\mu_i(\underline{\beta}, \beta_i^*) = -\mu_i(\beta_i^*, \overline{\beta})$ and*

$$\beta_i^* = \frac{D}{1+\lambda} - \frac{\mu_i(\beta_i^*, \overline{\beta})}{(1+\lambda)\phi_i(C(\beta_i^*), \beta_i^*)} > \beta_B.$$

Proof. See Appendix. □

The first part of the theorem is the canonical incentive regulation result where the least efficient agents receive no surplus; the upper bound $\overline{\theta}$, which can be arbitrarily large, can be interpreted as the firms' zero-profit condition (as in Spulber 1988; Kim and Chang 1993; Lewis 1996). The emission price is distorted below the Pigouvian level, $\beta_i^* < \beta_P$, by the familiar tradeoff between allocative efficiency and rent extraction. Lowering the externality price allows higher taxes to all, including the firms that can comply with very low β thus reducing their information rents.

[15] If β is known the planner can set $MS_i(C(\beta), \beta) = 0$ type by type, so $\mu_i(\beta_i^*, \overline{\beta}) = 0$ and the Pigouvian cut-off follows by (9). The term $D/(1+\lambda)$ is familiar from the earlier literature, such as Bovenberg and van der Ploeg (1994).

The second part of the theorem states our main result: The optimal externality price for mobile firms is strictly above the immobile-firm benchmark, $\beta_i^* > \beta_B$, and can even rise above the first-best level, β_P. We develop intuition for this argument in stages. Consider a two-part tariff fixed at the benchmark level, (T_B, β_B). How would this tariff change if agents' outside options improved so that true θ values come from $(-\infty, \overline{\theta}]$? Consider first that the policy maker can adjust only the externality price, not the transfer:

Lemma 2. ***(One-part tariff)*** *Holding transfer T_B constant, the externality price β_1, is reduced as an optimal response to mobility:*

$$\beta_i^* = \frac{D}{1+\lambda} - \frac{1}{1+\lambda}\frac{1-F(\beta_i^*)}{f(\beta_i^*)}\Big(\frac{\Delta(\beta_i^*)}{\eta(C(\beta_i^*))} - \lambda\Big) \equiv \beta_1 < \beta_B$$

where $\Delta(\beta_i^*) = -\big(\gamma - \lambda(T_B - \beta_i^*)\big)$.

We show in the Appendix that $\beta_i^* \geq \beta_B$ would mean that all firms that stay in i cut, but this cannot be optimal because polluting firms have social value through their γ and contribution to public funds. Thus, it is optimal to compensate them with a lower externality price, $\beta_1 < \beta_B$. The common intuition that policies need to be relaxed to keep firms from relocating builds on the assumption that the policy maker's hands are tied to use only one instrument in line with Figure 1a.

The second part of Theorem 1 shows that the optimal mechanism uses two instruments, one for compensations and another for the externality. The policy maker can now prevent firm relocation by increasing the base compensation and, in addition, use the externality price for efficient firm selection. The externality price moves in the opposite direction in comparison to the one-part tariff, $\beta_i^* > \beta_B$, because this allows increasing the transfer to the low-cost compliers whose relocation causes the greatest social loss. This is exactly the intuition given in Figure 1b.

Whether the optimal externality price rises above the Pigouvian level, $\beta_i^* > \beta_P$, depends on the characteristics of the industry, such as the location-preference distribution, $G(\beta)$. Recall that if the policy maker could observe firm types β but not θ, there would be no reason to deviate from the efficient cut-off, $\beta_i^* = \beta_P$. At this cut-off, the accompanying transfer, T_P, would be dictated by the social value of the firms, $MS_i(C(\beta), \beta) = 0$ at $\beta = \beta_P$. After some inspection, this latter condition can be written as $\lambda\eta(C(\beta)) = \Delta(\beta)$ which allows us to state the following result:

Proposition 1. ***(Distortion relative to first-best)*** *The policy is stricter than the Pigouvian reference, $\beta_i^* > \beta_P = D/(1+\lambda)$, if*

$$\lambda\eta'(C) < 1 \qquad \textit{for all } \underline{\beta} < C + T_P < \beta_P. \tag{10}$$

Low-cost firms are valuable because of their ability to effectively cut emissions at home and, intuitively, emission price is set above β_P if they are "valuable enough". Consider firms that have lower net costs than $\beta_P - T_P$. The social value of keeping firms is decreasing in β with slope 1, see equation (2). But the net cost of compliance is increasing in β and the high-cost firms need more compensation to stay; this effect given by $\lambda\eta'(C)$. The inequality in Proposition 1 compares these two effects.[16]

Note that upward distortion is not standard in models of random participation. In our setting the planner directly takes into account firms' private costs β as part of the social surplus. In a direct mechanism firms have an incentive to emphasize their social value by understating β and the incentive-compatible regulation is optimally distorted upwards to reduce such understatements. Similar effect does not arise in applications to nonlinear pricing (Rochet and Stole, 2002) where a selling firm is not concerned about buyers' valuation as such, or optimal income taxation Lehmann *et al.* (2014) where the government is not concerned about agents' productivity *per se*. In those settings, a non-decreasing inverse hazard rate assumption, $\eta'(C) \geq 0$, is sufficient to eliminate any upward distortions.

4 Global mechanism for global externalities

Until now, we have ruled out cross-boundary transfers and assumed that the policy maker has lost all opportunities to regulate firms that move. This strict focus on purely local policies comes with a loss of generality: After moving, firms' actions continue to impact welfare at home, so the policy maker should try to influence those actions. While taxing firms in other countries is not possible, it may be possible to incentivize moving firms to cut emissions voluntarily. We now look for welfare improvements by designing one incentive-compatible mechanism for the leaving firms $M_j(\beta) = \{T_j(\beta), X_j(\beta)\}$ and another one for the staying

[16] The condition $\lambda\eta'(C) < 1$ can be relaxed: it is enough that the derivative condition holds at cost levels implemented by the type-tailored mechanism where β is observable but θ remains private. Yet, we prefer to state the condition on primitives.

firms $M_i(\beta) = \{T_i(\beta), X_i(\beta)\}$, to maximize the welfare in (1) such that equations (3) and (5) hold and the voluntary participation constraint $C_j(\beta) \leq 0$ holds for all β.

The global mechanism has characteristics distinct from the purely local one. First, it is only possible to tax firms at home, but not in other sovereign jurisdictions, reflected by the voluntary participation constraint. Second, cross-border transfers have social cost $1+\lambda$ and thus come with a welfare-loss even if $\lambda = 0$. In contrast, the domestic transfers are effectively evaluated with social cost λ as, intuitively, the transfer circulates within the economy. Last, the outside options of home firms can now be manipulated by the simultaneous offering of the treatments $\{M_i(\beta), M_j(\beta)\}$. In particular, a firms' relocation incentive depends on depends on actions and transfers in both locations: $C(\beta) = \beta(X_i(\beta) - X_j(\beta)) - (T_i(\beta) - T_j(\beta))$.

We show in the Appendix that the optimal global mechanism takes the form of a two-part tariff at home location and a one-part tariff at the foreign location:

Lemma 3. ***(One-part tariff abroad)*** *The optimal global mechanism defines $M_i(\beta)$ by two constants as in Lemma 1 and $M_j(\beta)$ by one constant β_j^*:*

$$\begin{cases} T_j(\beta) = \beta_j^*, & X_j(\beta) = 1 \qquad \text{for } \beta \leq \beta_j^* \\ T_j(\beta) = 0, & X_j(\beta) = 0 \qquad \text{for } \beta > \beta_j^*. \end{cases} \tag{11}$$

Thus, it is optimal to set the base transfer to zero for leaving firms and pay only for cuts. Emission price β_j^* is not paid by the firms, but it is rather a payment that firms forgo if they do not reduce emissions. The global mechanism offers, in principle, great opportunities. If $X_i(\beta) = X_j(\beta) = 1$ for all low-cost firms with β below some β^*, then one implements global price β^* and thereby a global emissions cap. This eliminates the "leakage" problem altogether as global emissions become independent of firms' location. In fact, if β was observable, the optimal policy would set only one price at the Pigouvian level, $\beta^* = D/(1+\lambda)$, both for staying and moving firms, together with differentiated transfers across types and actions.[17]

But when also β is unobservable it is not possible to achieve $C_j(\beta) = 0$ abroad by differentiated transfers $T_j(\beta)$. To introduce the design problem for managing the rents at home and abroad in a stepwise manner, consider first policies that cannot discriminate firms' emissions based on their location and is thus constrained to set the same emissions price to

[17] With observable β for each $\beta \leq \beta^*$ moving, transfer $T_j(\beta)$ would satisfy $C_j(\beta) = 0$ because there is no need to pay more than necessary to "buy" the socially valuable abatement action from abroad. And, for each $\beta \leq \beta^*$ staying, T_i^* would be set to satisfy $MS_i(C(\beta), \beta) = 0$.

all.[18]

Proposition 2. ***(Uniform global price)*** *If the implemented global emission price is constrained to be uniform, $\beta^* = \beta_i^* = \beta_j^*$, it is optimally set at:*

$$\beta^* = \frac{D}{1+\lambda} - \frac{G(C(\beta^*)) + \lambda}{1+\lambda} \frac{F(\beta^*)}{f(\beta^*)} < \beta_P$$

While Theorem 1 gives the optimal *purely local* mechanism, Proposition 2 is the *purely global* mechanism counterpart. As the relocating firms cannot avoid the emission price, their relocation does not influence the total amount of externality produced globally. Recall that the reason for choosing a higher-than-Pigouvian emissions price is to target compensation to low-cost firms. This targeted compensation becomes less important when the low-cost firms will cut emissions even if they move. As a result, the uniform global price is distorted below the Pigouvian level.

It turns out that the policy maker can always do better than implementing either the purely local (Theorem 1) or the purely global mechanism (Proposition 2). Consider a local β_i^* and a smaller global cut-off $\beta_j^* < \beta_i^*$, satisfying

$$\big(D - (1+\lambda)\beta_i^*\big)\phi_i(C(\beta_i^*), \beta_i^*) = \mu_i(\beta_i^*, \overline{\beta}), \tag{12}$$

$$\big(D - (1+\lambda)\beta_j^*\big)\phi_j(C(\beta_j^*), \beta_j^*) = \mu_j(\underline{\beta}, \beta_j^*), \tag{13}$$

where μ_i and μ_j measure marginal surpluses from firms at locations i and j, respectively. When $\beta_j^* < \beta_i^*$, equation (12) takes exactly the same form as the one for the local mechanism in (9): It captures the loss from firms with $\beta > \beta_i^*$ that move to avoid the penalty on the externality altogether. The logic for equation (13) is different: marginally increasing β_j^* does not directly impact firms' costs at home but it makes relocation more attractive. Therefore,

$$\mu_j(\underline{\beta}, \beta_j^*) = \int_{\underline{\beta}}^{\beta_j^*} MS_j(C(\beta), \beta) d\beta = \int_{\underline{\beta}}^{\beta_j^*} -\Delta(\beta)\phi_j'(C(\beta_j^*), \beta) + (1+\lambda)\phi_j(C(\beta_j^*), \beta) d\beta, \tag{14}$$

where $-\Delta(\beta) = \gamma - \lambda T_i^* - \beta + (1+\lambda)\beta_j^*$. The higher global price, β_j^*, attracts movers which

[18] As noted before, a real-world company may consist of several of the units that we have called "firms", with headquarters at location i. The policy changes the distribution of activities across locations, and the interpretation of "no discrimination" means that all emissions from the same company is brought under the same cap.

is costly due to the lost value added, γ, but there is no impact on the externality: D does not appear. Cross-border transfers, as noted earlier, have social cost $1+\lambda$, rather than just λ.[19]

The two policies are intricately interconnected; see Appendix for the full characterization. We can make a few general statements. First, introducing a very small global externality price is welfare improving, because it has first-order welfare effects due to reduced emissions, $D-(1+\lambda)\underline{\beta} > 0$, but only second-order effects to increased relocation, $\mu_j(\underline{\beta},\underline{\beta}) = 0$.[20] Therefore, a small global price is better than the purely local mechanism. Second, paying a high β_j^* (but yet $\beta_j^* < \beta_i^*$) leads to non-negligle costs as captured by $\mu_j(\underline{\beta},\beta_j^*) > 0$. To deal with this negative effect, the global emissions price is always set below the local one.

Theorem 2. ***(Global Mechanism)*** *The optimal $M_i(\beta)$ and $M_j(\beta)$ implement $(\beta_i^*, T_i^*, \beta_j^*)$:*

(i) *A transfer T_i^* determined by: $\mu_i(\underline{\beta},\beta_i^*) = -\mu_i(\beta_i^*,\overline{\beta})$*

(ii) *A strictly positive but downward distorted global price*

$$0 < \beta_j^* = \frac{D}{1+\lambda} - \frac{\mu_j(\underline{\beta},\beta_j^*)}{(1+\lambda)\phi_j(C(\beta_j^*),\beta_j^*)} < \beta^*$$

(iii) *together with a strictly higher domestic price*

$$\beta^* < \beta_i^* = \frac{D}{1+\lambda} - \frac{\mu_i(\beta_i^*,\overline{\beta})}{(1+\lambda)\phi_i(C(\beta_i^*),\beta_i^*)}$$

where β^* is the uniform-price benchmark defined in Proposition 2.

Under Assumptions 1 and 2, the optimal mechanism always has both a global ($\beta_j^* > 0$) and a local ($\beta_i^* > \beta_j^*$) component. Our findings emphasize why local policies have a higher priority despite the fact that marginal damages are equal across locations. Intuitively, as local firms are valuable as such, the planner tolerates higher information rents at home than

[19]It may seem surprising that the left-hand sides of (12)-(13) look so similar despite the social cost $1+\lambda$ for cross-border transfers. Equation (12) includes the abatement cost of the marginal cutting firm, β_i^*, and transfers β_i^* from domestic firms weighted by λ. Equation (13) includes of the cross-border transfer β_j^* weighted by $1+\lambda$.

[20]Note that the cost is second-order, because we assume that firms do not abate abroad without active policies by country i. If country j had implemented a positive emission price, then the foreign price would be paid to a mass of firms and it would not be guaranteed that the optimal subsidy for foreign polluters is strictly positive, see Section 5.1.

abroad for the same cuts in emissions. In fact, Theorem 2 gives a stronger result: The price differentiation leads to a local price that is distorted upwards from the uniform-price benchmark in Proposition 2.

5 Extensions

5.1 Symmetric equilibrium

So far, the focus has been on how to *unilaterally* implement policies by a single country or a coalition of cooperating countries. But if other countries also start setting policy targets, such as those determined in the 2015 Paris Agreement, the carbon pricing policies will have a wider global coverage. How does the optimal mechanism in i respond to the increasing policy coverage? The mechanism, it turns out, manipulates the efficacy of policy outcomes elsewhere through the location distribution of firm types.[21]

We consider a game between two policy makers deviating minimally from our one-country analysis. We assume two symmetric regions that have the same D, γ, λ, F, and the distribution G that is symmetric around 0. First, home (i) and abroad (j) simultaneously set their local two-part tariffs (T_k, β_k), $k \in \{i, j\}$. Second, firms observe the mechanisms in place, and then choose the mechanism to comply with according to the self-selection condition (4).[22] The best-responses are interesting as such: They detect the strategic role of firm selection. The implications for a symmetric Nash equilibrium follow readily from these.[23]

Think first how to optimally choose (T_k, β_k) holding the mechanism selected by the other country as given. Consider $\beta_i^* > \beta_j^*$ and note that T_i^* should be set so that marginal surpluses $MS_i(C(\beta), \beta) = \Delta(\beta)\phi_i'(C(\beta), \beta) - \lambda\phi_i(C(\beta), \beta)$, integrated over all firms give $\mu_i(\underline{\beta}, \overline{\beta}) = 0$,

[21] This setting allows for richer strategic interactions than what we study in this extension. For example, if the destination of the mobile firms introduces a fixed externality price, such as a carbon tax, the total externality produced abroad can be manipulated strategically through the distribution of relocating firms. In contrast, when the destination country introduces a quantity-based regulation, such as an emissions trading scheme, the total externalities produced abroad are unaffected by firm relocation but the foreign externality price, and thereby firms' outside options, becomes endogenous to the mechanism.

[22] 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.

[23] Our assumption of a constant marginal damage guarantees that there is no traditional strategic common-pool interaction: without leakage, policies are independent between jurisdictions as one region cannot manipulate the marginal damage faced by the other through the strategic choice of emissions (see e.g. Van Long 2010).

otherwise the compensation should be increased or reduced. While this general reasoning is similar in the one-country situation, the make up of marginal surpluses is different in the game: $\mu_i(\underline{\beta}, \overline{\beta}) = \mu_i(\underline{\beta}, \beta_j^*) + \mu_i(\beta_j^*, \beta_i^*) + \mu_i(\beta_i^*, \overline{\beta})$, where $\mu_i(\underline{\beta}, \beta_j^*)$ covers $MS_i(C(\beta), \beta)$ from firms that cut in both regimes, $\mu_i(\beta_j^*, \beta_i^*)$ is for firms that cut only in i, and lastly $\mu_i(\beta_i^*, \overline{\beta})$ captures firms who cut in neither regime.[24] The cut-off β_i^* impacts the last part of this breakdown and thus solves

$$\big(D - (1+\lambda)\beta_i^*\big)\phi_i(C(\beta_i^*), \beta_i^*) = \mu_i(\beta_i^*, \overline{\beta}),$$

where, just like in the one-country situation, $\mu_i(\beta_i^*, \overline{\beta})$ measures the social loss from having a marginally higher β_i^* that increases costs for polluting firms who then locate more to j to pay the lower β_j^*. On the other hand, the country setting β_j^* solves

$$\big(D - (1+\lambda)\beta_j^*\big)\phi_j(C(\beta_j^*), \beta_j^*) = \mu_j(\beta_j^*, \beta_i^*) + \mu_j(\beta_i^*, \overline{\beta}),$$

where the interpretation of the right-hand side is similar as for i but β_j^* impacts the selection of firms differently from country i: It impacts those who cut only in i and those who avoid cuts in both locations. If the equilibrium is symmetric, we have $\mu_i(\beta_j^*, \beta_i^*) = 0$ and, quite surprisingly, countries evaluate the social cost of a lost firm, as if there was no damage: When the firm chooses the other location, it cuts the same as at home.[25]

Proposition 3. ***(Symmetric equilibrium)*** *In any symmetric non-cooperative equilibrium of the game,*
(i) *there is a global cap on emissions,*
(ii) *local externality prices are set below the Pigouvian level:*

$$\beta_i^* = \beta_j^* < \frac{D}{1+\lambda}.$$

The symmetric equilibrium creates a global cap on emissions, similar to that in Proposition 2. Low-cost firms will cut emissions in both countries and their relocation does not affect the

[24]The expanded expressions are in the Appendix. Here subindices i and j refer to local policies set by i and j, respectively, and ij and ji refer to global policies.

[25]Cramton *et al.* (2017) promote the idea of climate negotiations centered around commitments to a minimum carbon price. Our model offers one interpretation for the price emerging in a non-cooperative interaction. This interpretation is not in conflict with Cramton *et al.* (2017), but the insights on firm selection and leakage warrant a further investigation of reciprocal collaboration under the scheme.

level of the global externality. The second part of the proposition states that the upward-distortion, which was possible in the case of unilateral policies, never happens in a symmetric equilibrium. This result follows closely the intuition presented Proposition 2: The reason for choosing a higher than Pigouvian emissions price is to target compensation to low-cost firms. This targeted compensation becomes less important as low-cost firms do not cause emission leakage if they move to a regime where they also face regulation, eliminating the possibility of an upward distortion in the symmetric two country equilibrium.

What if both countries are given the opportunity to offer contracts also to firms in other locations, as in the global mechanism of Theorem 2? Country i would pay $\beta^*_{ij} \geq 0$ for firms that cut emissions in the other country and likewise for j. Firms in country j cut emissions if their cost is less than the effective carbon price, $\beta \leq \beta^*_{ij} + \beta^*_j$, that both countries can control. The top-up, β^*_{ij}, is optimally set at:

$$\big(D - (1+\lambda)\beta^*_{ij}\big)\phi_j(C(\beta^*_j), \beta^*_j) \leq \mu_{ij}(\underline{\beta}, \beta^*_{ij} + \beta^*_j)$$

holding with equality when $\beta^*_{ij} > 0$ and with strict inequality when $\beta^*_{ij} = 0$, and where μ_{ij} is as in (14). While in Section 4 it was always optimal to set a positive global price, this is no longer true in the symmetric equilibrium. Consider a small foreign price $\beta^*_{ij} \approx 0$. It has a positive effect due to reduced emissions, $D - (1+\lambda)\beta^*_{ij} \approx D > 0$, but the top-up needs to be paid to the entire mass of firms cutting in country j, $\mu_{ij}(\underline{\beta}, \beta^*_{ij} + \beta^*_j)$ which does not become zero when $\beta^*_{ij} \approx 0$. In the appendix we characterize the symmetric equilibrium and leave possible extensions open for future research.

5.2 Multi-sector mechanism

One may interpret our main mechanisms, in Theorems 1 and 2, as ones that are implemented sector by sector; after all, sectors differ in their primitive distributions F and G, and also value-added γ (Martin *et al.*, 2014a,b). This leads to differentiated externality pricing across sectors as proposed by Hoel (1996), but with an important difference: Sectors exposed to carbon leakage should face higher, not lower, externality prices (by our Theorem 1). This may, however, not be feasible as the policy maker might be restricted to use only a single corrective price for all the sectors and rely only on sector-specific transfers; indeed, this is the design adopted in the EU emissions trading system. With a slight abuse of notation, we now let $i = 1, ..., N$ to denote the sectors and consider welfare

$$\max_{\beta^*, T_1, \ldots, T_N} \sum_{i=1}^{N} \int_{\underline{\beta}}^{\beta^*} \Big(\gamma_i + D - \beta - \lambda T_i^*\Big)\phi_i(C(\beta), \beta)d\beta + \int_{\beta^*}^{\overline{\beta}} \Big(\gamma_i + \lambda(\beta^* - T_i^*)\Big)\phi_i(C(\beta), \beta)d\beta,$$

where β^* is the one price for all sectors and T_i^* is the sector-specific transfer.

Proposition 4. ***(Multisector extension of Theorem 1)*** *If the implemented emission price is constrained to be uniform across sectors, β^*, it is optimally set at:*

$$\beta^* = \frac{D}{1+\lambda} + \sum_{i=1}^{N} \frac{\mu_i(\beta^*, \overline{\beta})}{(1+\lambda)\phi_i(C(\beta^*), \beta^*)},$$

and T_i^* satisfies $\mu_i(\underline{\beta}, \beta^*) + \mu_i(\beta^*, \overline{\beta})=0$ sector by sector.

The transfer balances marginal surpluses from the two sets of firms, exactly as in Theorem 1 but with the cutoff β^* being "off" as it is optimal for the full group of N sectors. The potential loss from this adherence to one price is a quantitative question that we address below in Section 6.2.

5.3 General functional forms

The functional forms used in our main analysis are simplifying but they turn out not to be very restrictive. First, it would be possible to introduce an increasing and convex function $D(X)$ where $X = \int_{\underline{\beta}}^{\overline{\beta}}(1 - X_i(\beta))\phi_i(C(\beta), \beta) + (1 - X_j(\beta))\phi_j(C(\beta), \beta)d\beta$ gives the total emissions. In this setting, $D'(X)$ naturally replaces the constant marginal damages, denoted by D.[26] Since firms are atomistic, they cannot affect the level of regulation through the total pollution stock X. From the policy maker's point of view, however, the convex damage function leads to one difference: Relocation of low-cost facilities shifts the marginal cost curve upwards increasing the optimal emissions price as well as the Pigouvian reference. But since the policy maker foresees the aggregate mass of firms that will stay in i given a policy, nothing essential changes in the problem.

Next, we have assumed that the economy consists of numerous production units, referred to as "firms", each having a unit cost of emission reductions. As noted earlier, assuming

[26]It should be noted that constant damages can well approximate the predictions of the comprehensive climate-economy models (Golosov *et al.*, 2014; van den Bijgaart *et al.*, 2016).

no economies of scale we can interpret "firms" as units of a larger company with independently distributed abatement costs. An alternative modelling approach would be to consider abatement costs that, instead of being independently distributed, depend on a firm-specific privately known technology parameter β so that costs follow a convex function $A(X_k(\beta), \beta)$, with $X_k(\beta) \in \mathbb{R}^+$, $k = i, j$.[27] Then, where $X_i(\beta)$ and $X_j(\beta)$ are strictly decreasing, the optimal mechanism sets:[28]

$$\big(D - (1+\lambda)A_x(X_i(\beta), \beta)\big)\phi_i(C(\beta), \beta) = \mu_i(\beta, \overline{\beta})A_{xc}(X_i(\beta), \beta), \tag{15}$$

$$\big(D - (1+\lambda)A_x(X_j(\beta), \beta)\big)\phi_j(C(\beta), \beta) = \mu_j(\underline{\beta}, \beta)A_{xc}(X_j(\beta), \beta), \tag{16}$$

where $\mu_i(\beta, \overline{\beta}) = \int_{\beta}^{\overline{\beta}} \Big(\Delta(\tilde{\beta})\phi_i'(C(\tilde{\beta}), \tilde{\beta}) - \lambda\phi_i(C(\tilde{\beta}), \tilde{\beta})\Big) d\tilde{\beta}$ and $\mu_j(\underline{\beta}, \beta) = \int_{\underline{\beta}}^{\beta} \Big(-\Delta(\tilde{\beta})\phi_j'(C(\tilde{\beta}), \tilde{\beta}) + (1+\lambda)\phi_j(C(\tilde{\beta}), \tilde{\beta})\Big) d\tilde{\beta}$. These conditions closely resemble the ones in (12)-(13) in Section 4, with one important difference: while the main mechanism creates two prices, one local and one global (see Lemma 3), condition (15) and (16) offer each firms different effective prices depending on their β-type through function A. It follows that, unlike the main mechanisms in Theorems 1 and 2, the outcome cannot be implemented by a simple linear tax or an emissions trading market. Although the policy maker can now screen firms better by second-degree price discrimination to save on the public funds, the approach has heavy information requirements that may render it impractical for real-life policy-making: the regulator must be informed about the shape of the abatement cost functions $A(X_k(\beta), \beta)$ for each β.

5.4 Partly local externalities

Some global externality problems have a local element. For example, the reduction of greenhouse gases is typically associated with other jointly produced local pollutants, as emphasized by recent empirical work (e.g., Wagner and De Preux 2016; Holland *et al.* 2018). We can

[27] More precisely, we assume $A_x(X, \beta) > 0$, $A_{xx}(X, \beta) \geq 0$, $A_{xc}(X, \beta) > 0$ and that $A(X, \beta)$ satisfies the Inada conditions. The derivations are given in the Appendix.

[28] The abatement levels provided by equations (15) and (16) may sometimes fail to be monotonic, so that the non-monotonicity condition for $X_i(\beta)$ or $X_j(\beta)$ is binding and bunching arises: some firms with different types β 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), in this extension we focus on the cases where full separation is optimal. Technically, we focus on cases where the non-monotonicity constraints for $X_i(\beta)$ and $X_j(\beta)$ never bind.

include this effect by letting the total damage D to be a sum of global αD and local $(1-\alpha)D$ components, with $\alpha \in [0,1]$. With this, the net loss from relocation (2) becomes

$$\Delta(\beta) = -\Big(\gamma - (1-\alpha)D + D(X_i(\beta) - \alpha X_j(\beta)) - C_i(\beta) - (1+\lambda)(T_i(\beta) - T_j(\beta))\Big).$$

Take $\gamma' = \gamma - (1-\alpha)D$ as a new definition of the value-added, and the role of Δ in the analysis remains practically unchanged. The observation has nevertheless interesting consequences. Note first that the immobile firms benchmark (β_B, T_B) from Theorem 1 remains independent of α. Clearly, the total social gain from reductions remains at D, so the marginal trade-off for emission reductions remains the same for any division between local and global damages. This same reasoning holds when firms are mobile but, because the firms' social value is now lower ($\gamma' < \gamma$), transfer T_i^* changes, and this alters the socially optimal emissions price β_i^*. Yet, we can show that the result of Theorem 1 remains:[29]

Proposition 5. ***(Local-damage extension of Theorem 1)*** *For any division $\alpha = [0,1]$ between local and global damages, the immobile firm benchmark remains at*

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

and the optimal local mechanism price is strictly higher,

$$\beta_i^* > \beta_B.$$

The result is in contrast with the main line of results from the literature starting with Oates and Schwab (1988) where mobile capital and competition between jurisdictions leads to downward distortions in environmental policies even when damages are purely local. As in our main result, the driver of the difference to the literature is the firm selection effect that motivates targeting transfers to those firms that have the lowest abatement costs.

[29] Variations in α translate into variations in the optimal (β_i^*, T_i^*), but the full comparative statics is not possible without stronger assumptions on distributions F and G. We study numerically the relationship between γ and the optimal mechanism in the quantification Section (see Figure 2).

6 Application to the EU ETS sectors

6.1 Practical implementation strategies

We look at the magnitude of the results by providing a quantification of the optimal mechanism for the key sectors in the EU emissions trading system. Putting numbers first aside, it is useful to spell out how the theory mechanism can be mapped into policy instruments with a practical meaning.

The local mechanism (Theorem 1) can be implemented by a combination of externality pricing and base compensations. The externality price, β_i^*, can be implemented by a cap-and-trade scheme or by a carbon tax and, as there is no aggregate uncertainty, these two policy instruments lead to identical outcomes. The price can be differentiated between sectors, for example, by differentiated carbon taxes (Hoel, 1996), compliance cost rebates or provisions to use cheaper offsets for compliance. The base compensation, T_i^*, can take the form of direct monetary compensation, free allowances, or lump-sum rebates.[30]

The global mechanism (Theorem 2), in turn, can be implemented by cross-border permit trading, by allowing relocating firms to sell permits to the local policy maker for verified emission reductions in their new location for price β_j^*. This is against the main principles in the EU emissions trading system, where moving firms are not permitted to continue trading with the market. In our optimal mechanism, these voluntary cross-border trades resemble an "opt-in" scheme where relocating firms with high abatement cost choose not to participate, receiving their outside option. While cross-border transfers may seem politically difficult to implement, they are not new in the international arena; in linked emission trading systems firms are allowed to trade permits across country borders (CARB, 2017; EC, 2019a).

6.2 Quantification

To illustrate the economic significance of the results, we carry out an exploratory quantification for the key sectors under relocation risk in the EU emissions trading system (EU ETS): cement, iron and steel, chemical and plastic, wood and paper, and glass. Together, these

[30] For example, in the EU 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, 2019b).

Table 1: Descriptive statistics of the data used

	Total emissions in 2015 ($MtCO_2$)[1]	EBIT per emissions (€/tCO_2)[2]	Relocation probability[2]		No. firms[2]	Parameters[3]	
			0% compen-sation	80% compen-sation		Mean	Variance
Cement	113.8	32.73	0.46	0.20	46	27.98	716.6
Iron and Steel	120.6	80.52	0.60	0.21	25	20.38	363.3
Chemical and Plastic	74.9	177.96	0.24	0.06	64	41.26	525.3
Wood and Paper	27.1	89.31	0.14	0.03	61	53.05	672.9
Glass	18.2	120.56	0.14	0.05	24	65.32	1389.8
Aggregate	354.6	88.49	0.42	0.15	220	30.08	591.0

[1]Data from EEA (2017), [2]Data from Martin *et al.* (2014a), [3]Mean and variance of standard distribution, calibrated separately for each sector. Aggregate is calculated based on the sum (columns 1-5) and emission-weighted averages (columns 2-4) of individual sectors, and by calibrating a distribution based on relocation probabilities (columns 6-7).

five sectors produce 355 $MtCO_2$, or 62 per cent of emissions from all industrial installations covered by the EU ETS (EEA, 2017).

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. Bayer and Aklin (2020) find that the EU ETS, with an average price of 10.2 €/tCO_2, has reduced 11.5% of the regulated emissions between 2008 and 2016. We assume that abatement costs are distributed uniformly, and calibrate the distribution based on the numbers presented above.[31] The social cost of public funds is assumed to be $\lambda = .6$.[32] Marginal damages are chosen to be $D =$40 €/tCO_2, in line with Nordhaus (2017).[33]

We calibrate the relocation cost distributions based on the survey data collected by Martin *et al.* (2014a). The data contains firm-level assessments of the relocation probability

[31]The average European Union Allowance (EUA) price from April 2008 to December 2016 was 10.2 €(https://ember-climate.org/carbon-price-viewer/). Bayer and Aklin (2020) use generalized synthetic control and find that the EU ETS has reduced emissions by 11.5% on average over 2008-2016, corresponding to 40.8 $MtCO_2$ reduction in the sectors shown in Table 1, roughly in line with studies that use firm- or plant-level data (Martin *et al.*, 2016; Dechezleprêtre *et al.*, 2018). These numbers pin down the parameters of the uniform distribution, $\underline{\beta} = 0$ and $\overline{\beta} = 10.2/0.115 = 88.7$ €/tCO_2. The authors look into four sectors: energy, metals, minerals and chemicals, and find similar responses across sectors. We allocate abatement to sectors in proportion to their initial emissions. The assumption of uniform distribution is in line with general equilibrium modelling, where the abatement cost curve is approximately linear (Böhringer *et al.*, 2014).

[32]Country-specific circumstances have a large impact on the real costs of taxation so one number cannot fit the entire EU. Our number $\lambda = 0.6$ represents a median value of those calculated for the EU countries by computable general equilibrium models (Barrios *et al.*, 2013).

[33]Nordhaus's estimated marginal damages for 2020 are \$36.7-\$37.3 in \$2010, converted to 2020 euros these numbers are 39.8-40.5 €/ton CO_2. The value also comes close to the social cost of carbon used by the Obama Administration in the U.S. (EPA, 2017).

Table 2: Optimal mechanism for the EU ETS sectors

	Implementation of the mechanism			Implied emission reductions		
	Base compensation (€/tCO_2)	Local CO_2 price (€/tCO_2)	Global CO_2 price (€/tCO_2)	Local reductions (MtCO_2)	Global reductions (MtCO_2)	Emission leakage (MtCO_2)
Panel A - Local mechanism						
Cement	12.7	23.4	-	25.60	-	4.41
Iron and Steel	27.2	21.4	-	28.09	-	0.96
Chemical and Plastic	20.3	21.3	-	17.68	-	0.29
Wood and Paper	4.2	22.0	-	6.45	-	0.28
Glass	6.3	22.8	-	4.42	-	0.26
Total				82.24	-	6.20
Panel B - Uniform-price mechanism						
Cement	11.7	22.1	-	24.14	-	4.23
Iron and Steel	27.8	22.1	-	29.08	-	0.97
Chemical and Plastic	21.0	22.1	-	18.38	-	0.29
Wood and Paper	4.2	22.1	-	6.47	-	0.28
Glass	5.8	22.1	-	4.28	-	0.26
Total				82.35	-	6.03
Panel C - Global mechanism						
Cement	12.0	23.0	7.5	24.83	1.08	3.60
Iron and Steel	27.1	21.3	3.3	28.06	0.05	0.92
Chemical and Plastic	20.3	21.3	1.6	17.68	0.01	0.28
Wood and Paper	4.1	22.0	4.1	6.43	0.02	0.27
Glass	6.2	22.7	6.2	4.40	0.03	0.24
Total				81.40	1.19	5.31
Immobile firm benchmark	-	18.2	-	72.69	-	-

Notes: Optimal base compensations, implied marginal carbon taxes (columns 1-3) and the effects on emission reductions and leakage (columns 4-6) for the local mechanism (Panel A), the uniform-price mechanism between sectors (Panel B) and the global mechanism (Panel C). The social cost of carbon is 40 €/tCO_2 and the social cost of public funds is $\lambda = .6$ leading to β_P =25.0 €/tCO_2. Assumptions detailed in the text.

conditional on receiving no free permits and receiving 80% for free.[34] From these responses, we construct emissions-weighted industry averages for the relocation probability, see Table 1. We 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 carbon price is imposed, and 20% relocate if 80% of the carbon price is given back to firms. We use the Pigouvian carbon price $D/(1+\lambda) = 25$ €$/tCO_2$ in these calculations.

We report the optimal policies per sector in Table 2. Panel A gives the optimal local mechanism (Theorem 1). The first column gives the base compensation level, and the sec-

[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 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 %).

ond column presents the effective local emissions price per sector. All the sectors receive compensation, with Iron and Steel, Chemical and Plastic and Cement sector receiving the most. The optimal CO_2 prices are differentiated between sectors and vary between 21.3-23.4 €/tCO_2. The key take-away result from the quantification is that the main theoretical results turn out to be also economically significant: the effective CO_2 price is substantially elevated, by $17 - 29$ per cent compared to the benchmark level where leakage was assumed away, 18.2 €/tCO_2 (Theorem 1, immobile firms).[35] Yet, in all the sectors the emissions price falls short of the Pigouvian benchmark (25 €/tCO_2), so the condition for upward distortion (Proposition 1) does not hold for these parameters. In columns 4-6 we show that these higher local prices translate into larger global emission reductions even when firm relocation is taken into account.[36] For a benchmark, if all the sectors considered would be immobile by assumption, the total emission reductions would be 72.69 MtCO_2. A key observation is that the higher carbon price also translate into larger cuts (82.24MtCO_2), even after emission leakage (6.20MtCO_2) is taken into account; mobile firms increase the total reduction by 13 per cent.

Panel B shows the restricted uniform-price mechanism where the corrective price is constrained to be the same across all the sectors as in Proposition 4. The resulting uniform price is 22.1€/tCO_2, or 21 per cent above the no-leakage benchmark. A key observation is that the policy maker benefits relatively little from differentiating carbon prices between sectors. The levels of compensation, emission reductions and leakage change little when the emission price changes.

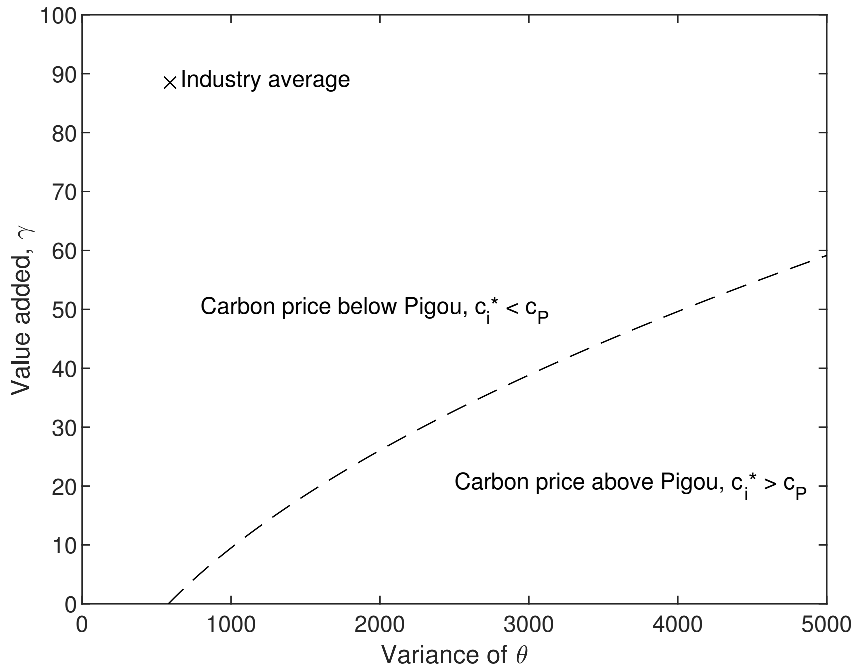
Finally, Panel C presents the optimal global mechanism per sector (Theorem 2). The key result is that, because of the effects identified in Section 4, the optimal global price is well below the local one: only $8 - 33$ per cent of the local price. This global emission cap leads to additional reductions of 1.19 MtCO_2 abroad for a total of 8.5 million euros used in cross-border transfers. In this quantification, the base compensation, local price and the local reductions are relatively robust to the introduction of the global mechanism.

Under what parameter assumptions would it be optimal to set an emission price above

[35] Note that the benchmark price for immobile firms, β_B, can be computed without knowing the distribution of the relocation cost. For the optimal mechanism, this distribution is essential and its support can be extended to include $\overline{\theta} = +\infty$, as in the case of a normal distribution.

[36] Columns 1-3 of Table 2 are based on the sample of surveyed medium-sized manufacturing firms with 50-5,000 employees, see Martin *et al.* (2014a). In Columns 4-6 we use the total emissions per industry (Table 1, column 1) and assume that the sample is representative for the entire industry. Figure 2 presents a sensitivity analysis for different values of γ and the variance of G.

Figure 2: Carbon price distortion and parameter choices



Notes: The Figure depicts the results from a simulation that finds the parameter combinations for which the optimal carbon price exceeds or falls below the Pigouvian level (dashed line). The social cost of carbon is 40 €/tCO_2 and the social cost of public funds is $\lambda = .6$. The emissions-weighted industry average value for the value added is $\gamma = 88.7$ and the variance is 591.0.

the Pigouvian level? In Figure 2 we have two key parameters used in the quantification, the value added γ (horizontal axis) and the dispersion of relocation cost θ as measured by the variance. The depicted locus gives the parameter combinations for which the optimal carbon price exceeds or falls below the Pigouvian level. When the Planner knows the relocation cost rather precisely (low variance of θ), the externality price is down-distorted as is expected; a higher variance means more dispersion in θ and thus a greater fraction of low β firms will move, which is mitigated by a higher externality price and larger transfers. When staying firms have a high value added (high γ), then keeping the low-cost firms becomes relatively less important and carbon price is set below the Pigouvian level, to compensate the firms who face the highest cost of regulation. But when value added is less significant (low γ), relatively more weight is given on the global externality. It becomes particularly important to keep the low-cost firms that can cut emissions effectively by a combination of an upward-distortion and lump-sum compensations. This illustration is obtained for a representative industry, with parameters reported under the Figure. The black mark denotes the average values for industries in Table 2.

7 Conclusions

A hundred years after the first proposal for corrective externality prices, the economics profession continues to believe in the approach to solve the global commons problems (Cramton *et al.*, 2017). When firms' *costs of limiting the externality* are not observed, externality pricing conveniently incentivizes low-cost firms to cut. We show that externality prices, if suitably designed, have another advantage: When firms' *relocation costs* are not observed, high externality prices incentivize low-cost firms, that can efficiently contribute to the commons problem, to stay. This selection effect elevates the optimal corrective price where the risk of firm relocation is present. Along the same lines of reasoning, self-interested decisions justify payments that, effectively, implement externality prices also for moving firms.

These results advise against regulatory rollbacks and other forms of routinely-used compensation policies that effectively curb carbon prices, including emission tax refunds, the use of cheap offsets, and exemptions of certain industries from regulation. As firm relocation serves to limit overcompensation paid to industries, observing carbon leakage is not a sign of a failed policy but an essential feature of the information-constrained optimal mechanism. On the contrary, one can argue that the EU emissions trading system has failed exactly because no relocation is observed; see studies by Dechezleprêtre *et al.* (2019) and Naegele and Zaklan (2019). Finally, instead of the current practice where moving firms stop being part of the regulation, there is a well-founded justification for cross-border transfers that, effectively, allow moving firms to sell their emission reductions to the local policy maker.

References

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, **26** (2), 445–469.

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

Barrios, S., Pycroft, J. and Saveyn, B. (2013). The marginal cost of public funds in the EU: the case of labour versus green taxes. *Fiscal Policy and Growth*, **403**.

Bayer, P. and Aklin, M. (2020). The european union emissions trading system reduced CO2 emissions despite low prices. *Proceedings of the National Academy of Sciences*, **117** (16), 8804–8812.

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

Bovenberg, A. L. and van der Ploeg, F. (1994). Environmental policy, public finance and the labour market in a second-best world. *Journal of Public Economics*, **55** (3), 349–390.

Bragadóttir, H., von Utfall Danielsson, C., Magnusson, R., Seppänen, S., Stefansdotter, A. and Sundén, D. (2014). *The Use of Economic Instruments: In Nordic Environmental Policy 2010-2013*. Nordic Council of Ministers.

CARB (2017). Facts about the linked cap-and-trade programs. *https://ww2.arb.ca.gov/sites/default/files/classic//cc/capandtrade/linkage/linkage_fact_sheet.pdf*.

Cramton, P., MacKay, D. J., Ockenfels, A. and Stoft, S. (2017). *Global carbon pricing: the path to climate cooperation*. The MIT Press.

Cremer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, pp. 1247–1257.

Dechezleprêtre, A., Gennaioli, C., Martin, R., Muûls, M. and Stoerk, T. (2019). Searching for carbon leaks in multinational companies. *Centre for Climate Change Economics and Policy Working Paper No. 187 ISSN 2515-5709*.

—, Nachtigall, D. and Venmans, F. (2018). The joint impact of the european union emissions trading system on carbon emissions and economic performance.

Duggan, J. and Roberts, J. (2002). Implementing the efficient allocation of pollution. *American Economic Review*, **92** (4), 1070–1078.

EC (2019a). Agreement on linking the emissions trading systems of the EU and Switzerland. *https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6708*.

— (2019b). EU Emissions Trading System (EU ETS). https://ec.europa.eu/clima/policies/ets_en, online; accessed 2019-03-01.

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

Environment Canada (2018). *Regulatory proposal for the Output-Based Pricing System Regulations under the Greenhouse Gas Pollution Pricing Act*. Environment and Climate Change Canada, Public Inquiries Centre.

EPA (2017). The social cost of carbon. *https://19january2017snapshot.epa.gov/climatechange/social-cost-carbon_.html*.

Fagan-Watson, B. (2015). *Big business using trade groups to lobby against EU climate policy*. The Guardian, April 15, 2015.

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.

H. Con. Res. 119 (2018). *Expressing the sense of Congress that a carbon tax would be detrimental to the United States economy.*

Hanna, R. (2010). US environmental regulation and FDI: evidence from a panel of US-based multinational firms. *American Economic Journal: Applied Economics*, **2** (3), 158–89.

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.

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.

Holland, S. P., Mansur, E. T., Muller, N. and Yates, A. J. (2018). Decompositions and policy consequences of an extraordinary decline in air pollution from electricity generation. *NBER Working Paper 25339*.

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 Tirole, J. (1993). *A theory of incentives in procurement and regulation.* MIT press.

Larsson, J., Elofsson, A., Sterner, T. and Åkerman, J. (2019). International and national climate policies for aviation: a review. *Climate Policy*, **19** (6), 787–799.

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

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

Markusen, J. R., 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. (2014a). Industry compensation under relocation risk: A firm-level analysis of the EU emissions trading scheme. *The American Economic Review*, **104** (8), 2482–2508.

—, —, — and Wagner, U. J. (2014b). On the empirical content of carbon leakage criteria in the eu emissions trading scheme. *Ecological Economics*, **105**, 78–88.

—, — and Wagner, U. J. (2016). The impact of the european union emissions trading scheme on regulated firms: what is the evidence after ten years? *Review of environmental economics and policy*, **10** (1), 129–148.

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

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

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, **18** (2), 301 – 317.

Naegele, H. and Zaklan, A. (2019). Does the EU ETS cause carbon leakage in european manufacturing? *Journal of Environmental Economics and Management*, **93**, 125–147.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, **114** (7), 1518–1523.

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

Oates, W. E. and Schwab, R. M. (1988). Economic competition among jurisdictions: efficiency enhancing or distortion inducing? *Journal of public economics*, **35** (3), 333–354.

Petrakis, E. and Xepapadeas, A. (2003). Location decisions of a polluting firm and the time consistency of environmental policy. *Resource and energy Economics*, **25** (2), 197–214.

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

— and — (2003). The economics of multidimensional screening. *Econometric Society Monographs*, **35**, 150–197.

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.

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 Benthem, A. and Kerr, S. (2013). Scale and transfers in international emissions offset programs. *Journal of Public Economics*, **107**, 31–46.

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.

Van Long, N. (2010). *A survey of dynamic games in economics*, vol. 1. World Scientific.

Varian, H. R. (1994). A solution to the problem of externalities when agents are well-informed. *The American Economic Review*, pp. 1278–1293.

Wagner, U. J. and De Preux, L. (2016). The co-benefits of climate policy: Evidence from the EU Emissions Trading Scheme. *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: International Climate Policy, No. D15-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg.*

A Appendix: Proofs

Proof of Lemma 1. (Two-part tariff)

The two-part form is shown by Lemma A.4 below.

Proof of Theorem 1. (Local mechanism)

We begin by introducing a series of lemmas characterizing the optimal mechanism.

Lemma A.1. *In a given mechanism i, transfer $T_i(\beta)$ is constant when policy $X_i(\beta)$ is constant.*

Proof. Proof by contradiction. Assume that there are β and β' with $T_i(\beta') > T_i(\beta)$ with $X_i(\beta') = X_i(\beta)$. Now firm β can get a lower net cost by reporting β':

$$\beta X_i(\beta') - T_i(\beta') < \beta X_i(\beta) - T_i(\beta)$$

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

Lemma A.2. *In a given mechanism i, $X_i(\beta)$ is nonincreasing in β.*

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

$$\beta X_i(\beta) - T_i(\beta) \leq \beta X_i(\beta') - T_i(\beta')$$

$$\beta' X_i(\beta) - T_i(\beta) \geq \beta' X_i(\beta') - T_i(\beta')$$

Combining these two inequalities leads to:

$$\beta'\big(X_i(\beta') - X_i(\beta)\big) - T_i(\beta') \leq \beta\big(X_i(\beta') - X_i(\beta)\big) - T_i(\beta') \Rightarrow \beta' \leq \beta$$

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

Lemma A.3. *In the optimal local mechanism ($T_j(\beta) = 0$, $X_j(\beta) = 0$), actions take a bang-bang form: $X_i(\beta) = \{0, 1\}$*

Proof. The objective function (1) can be written as:

$$\max_{X_i(\beta), C_i(\beta)} \int_{\underline{\beta}}^{\overline{\beta}} \Big(\gamma + DX_i(\beta) - (1+\lambda)\beta X_i(\beta) + \lambda C_i(\beta)\Big)\phi_i(C_i(\beta), \beta)d\beta$$

s.t. $C_i'(\beta) = -X_i(\beta)$ holds for all β. Denoting the co-state variable by $\nu_i(\beta)$, the Hamiltonian for this problem reads:

$$\mathcal{H} = \Big(\gamma + DX_i(\beta) - (1+\lambda)\beta X_i(\beta) + \lambda C_i(\beta)\Big)\phi_i(C_i(\beta),\beta) + \nu_i(\beta)X_i(\beta)$$

The Hamiltonian is linear in the controls $X_i(\beta)$, and the necessary conditions for optimality imply that $X_i(\beta)$ takes a bang-bang form: $X_i(\beta) = \{0,1\}$. Q.E.D.

Lemma A.4. *The optimal policy takes the two-part tariff form in Lemma 1*

Proof. There exists a solution to the problem as stated in Lemma A.3 by Filippov-Cesari Theorem (Theorem 8, page 132, Seierstad and Sydsæter 1987). Lemmas A.1-A.3 above tell us that the optimal policy that satisfies the incentive compatibility conditions takes a threshold form, where $X_i(\beta) = 1$ for $\beta \leq \beta_i^*$ and $X_i(\beta) = 0$ for $\beta > \beta_i^*$. Transfers are $T_i^1(\beta) = -\beta_i^* + T_i^*$ for $\beta > \beta_i^*$ and $T_i^2(\beta) = T_i^*$ for $\beta \leq \beta_i^*$, guaranteeing indifference for type β_i^*: $-T_i^1(\beta_i^*) = \beta_i^* - T_i^2(\beta_i^*)$. Q.E.D.

By Lemma 1, policy maker i is left to find β_i^* and T_i^* that maximize the social welfare, from equation (1):

$$\max_{\beta_i^*, T_i^*} W_i = \int_{\underline{\beta}}^{\beta_i^*} \Big(\gamma + D - \beta - \lambda T_i^*\Big)\phi_i(\beta - T_i^*, \beta)d\beta + \int_{\beta_i^*}^{\overline{\beta}} \Big(\gamma + \lambda(\beta_i^* - T_i^*)\Big)\phi_i(\beta_i^* - T_i^*, \beta)d\beta$$

Here the first integral covers all the firms below the threshold β_i^* cutting emissions, and the second term covers firms above the threshold that do not cut emissions (see Figure A.1 for a graphical illustration). Note that, by Lemma 1, incentive compatibility constraints are captured by the fact that the government is restricted to offer the same β_i^* and T_i^* to all agents. Begin by taking the first-order condition with respect to β_i^*. Using Leibniz's integral rule, we can derive equation (9):

$$\Big(D - (1+\lambda)\beta_i^*\Big)\phi_i(\beta_i^* - T_i^*, \beta_i^*) - \underbrace{\int_{\beta_i^*}^{\overline{\beta}} \Big(\Delta(\beta)\phi_i'(\beta_i^* - T_i^*, \beta) - \lambda\phi_i(\beta_i^* - T_i^*, \beta)\Big)d\beta}_{=\mu_i(\beta_i^*, \overline{\beta})} = 0 \qquad \text{(A.1)}$$

where $-\Delta(\beta) = \gamma + (D - \beta)X_i(\beta) - \lambda T_i(\beta)$ denotes the net welfare effect of relocation by type β. Simplify and solve for β_i^*:

$$\beta_i^* = \frac{D}{1+\lambda} - \frac{\mu_i(\beta_i^*, \overline{\beta})}{(1+\lambda)\phi_i(\beta_i^* - T_i^*, \beta_i^*)}. \qquad \text{(A.2)}$$

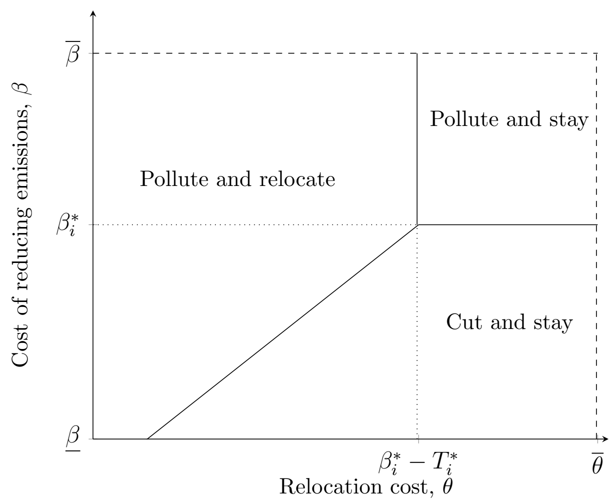


Figure A.1: Graphical illustration of the two-dimensional type space in the local mechanism.

Then, find the first-order condition with respect to T_i^*:

$$\underbrace{\int_{\underline{\beta}}^{\beta_i^*} \Big(\Delta(\beta)\phi_i'(\beta - T_i^*, \beta) - \lambda\phi_i(\beta - T_i^*, \beta)\Big) d\beta}_{=\mu_i(\underline{\beta}, \beta_i^*)} + \underbrace{\int_{\beta_i^*}^{\overline{\beta}} \Big(\Delta(\beta)\phi_i'(\beta_i^* - T_i^*, \beta) - \lambda\phi_i(\beta_i^* - T_i^*, \beta)\Big) d\beta}_{=\mu_i(\beta_i^*, \overline{\beta})} = 0 \tag{A.3}$$

Consider $\beta_i^* = \underline{\beta}$: No firm is required to cut. Then, $\mu_i(\beta_i^*, \overline{\beta})$ in (A.1) becomes $\mu_i(\underline{\beta}, \overline{\beta})$ which is zero by condition (A.3). By Assumption 2, the left side of (A.1) is strictly positive, a contradiction. We must have $\beta_i^* > \underline{\beta}$. Consider then $\beta_i^* = \overline{\beta}$: All firms are required to cut. $\mu_i(\beta_i^*, \overline{\beta})$ in (A.1) becomes $\mu_i(\overline{\beta}, \overline{\beta}) = 0$. By Assumption 2, the left side of (A.1) is now strictly negative, a contradiction with optimality. We thus must have $\beta_i^* < \overline{\beta}$. These arguments show that the optimal policy is a two-part tariff implementing interior $\beta_i^* \in (\underline{\beta}, \overline{\beta})$.

Combine (A.2) and (A.3) to write:

$$\beta_i^* = \frac{D}{1+\lambda} + \frac{\mu_i(\underline{\beta}, \beta_i^*)}{(1+\lambda)\phi_i(\beta_i^* - T_i^*, \beta_i^*)}. \tag{A.4}$$

Immobile firms. In this case we have the participation constraint of the form $C(\beta) \leq \overline{\theta}$. Following the usual arguments for the mechanism design literature (Laffont and Tirole, 1993), it is optimal to set the base transfers such that the participation constraint is binding for high-β firms that receive no information rents, $C(\beta_i^*) = \overline{\theta}$:

$$\beta_i^* - T_i^* = \overline{\theta}.$$

The high-β firms are indifferent between moving and staying while low-β firms receive information rents. Consequently there is no leakage and we have $\phi_i'(C(\beta), \beta) = 0$, $\phi_i(C(\beta), \beta) = f(\beta)$ and $\phi_j(C(\beta), \beta) = 0$ for all β. Now the integral $\mu_i(\underline{\beta}, \beta_i^*)$ can be simplified to:

$$\mu_i(\underline{\beta}, \beta_i^*) = \int_{\underline{\beta}}^{\beta_i^*} -\lambda f(\beta) d\beta = -\lambda F(\beta_i^*)$$

Plug this into equation (A.4) for the optimal externality price with immobile firms:

$$\beta_i^* = \frac{D}{1+\lambda} + \frac{\mu_i(\underline{\beta}, \beta_i^*)}{(1+\lambda)\phi_i(-T_i^*, \beta_i^*)} = \frac{D}{1+\lambda} - \frac{\lambda}{1+\lambda} \frac{F(\beta_i^*)}{f(\beta_i^*)} \equiv \beta_B \tag{A.5}$$

Mobile firms. We show that when distribution $G(\theta)$ is defined on $(-\infty, \overline{\theta}]$, it is optimal to set the externality price strictly above β_B defined in (A.5). The proof is by contradiction. Assume that $\beta_i^* \leq \beta_B$. Use the definitions $\phi_i(C(\beta), \beta) = (1 - G(\beta))f(\beta)$ and $\phi_i'(C(\beta), \beta) = -g(\beta)f(\beta)$, equation (5), where we denote $G(C(\beta)) = G(\beta)$ and $g(C(\beta)) = g(\beta)$ for shorthand. Rewrite condition (A.4) to obtain:

$$D - (1+\lambda)\beta_i^* + \int_{\underline{\beta}}^{\beta_i^*} -\Delta_L(\beta) \frac{g(\beta)}{1 - G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)} - \lambda \frac{1 - G(\beta)}{1 - G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)} d\beta = 0. \tag{A.6}$$

Here, we have denoted the welfare effect of relocation for firms that cut emissions as: $-\Delta_L(\beta) = \gamma - \lambda T_i^* + D - \beta$, where the subindex L refers to "Low" costs β. Use the assumption $\beta_i^* \leq \beta_B$, implying $D - (1+\lambda)\beta_i^* \geq \lambda \frac{F(\beta_i^*)}{f(\beta_i^*)}$, to write first-order condition (A.6) as the following inequality:

$$\int_{\underline{\beta}}^{\beta_i^*} -\Delta_L(\beta) \frac{g(\beta)}{1 - G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)} \underbrace{-\lambda \frac{1 - G(\beta)}{1 - G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)}}_{\equiv A} d\beta \leq -\lambda \frac{F(\beta_i^*)}{f(\beta_i^*)} \tag{A.7}$$

Integrate term A by parts (note, that $C'(\beta) = 1$), and use $F(\underline{\beta}) = 0$ to write:

$$A = -\int_{\underline{\beta}}^{\beta_i^*} \lambda \frac{1 - G(\beta)}{1 - G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)} d\beta = -\lambda \frac{F(\beta_i^*)}{f(\beta_i^*)} - \int_{\underline{\beta}}^{\beta_i^*} \frac{g(\beta)}{1 - G(\beta_i^*)} \frac{F(\beta)}{f(\beta_i^*)} d\beta$$

Using this, inequality (A.7) becomes:

$$\int_{\underline{\beta}}^{\beta_i^*} \frac{g(\beta)}{1-G(\beta_i^*)} \frac{f(\beta)}{f(\beta_i^*)} \underbrace{\left(-\Delta_L(\beta) - \lambda \frac{F(\beta)}{f(\beta)} \right)}_{\equiv B} d\beta \leq 0 \tag{A.8}$$

The inequality holds true if term B, defined above, is nonpositive. Using the definition of $-\Delta_L$:

$$-\Delta_L(\beta) = \gamma - \lambda T_i^* + D - \beta > D - \beta - \lambda \beta_i^* > D - (1+\lambda)\beta_i^* \geq \lambda \frac{F(\beta_i^*)}{f(\beta_i^*)}$$

The first inequality follows from the fact that $\beta_i^* \leq \beta_B$ implies $-\mu_i(\beta_i^*, \overline{\beta}) \leq -\lambda F(\beta_i^*) < 0$, which by the definition of $\mu_i(\beta_i^*, \overline{\beta})$ in eq. (8) implies $\gamma - \lambda T_i^* + \lambda \beta_i^* > 0$. The second inequality follows from the fact that β_i^* is the upper integral bound in (A.8), and therefore $\beta \leq \beta_i^*$. The third inequality follows from $\beta_i^* \leq \beta_B$, implying $D - (1+\lambda)\beta_i^* \geq \lambda \frac{F(\beta_i^*)}{f(\beta_i^*)}$. Term B in (A.8) then writes as:

$$B = -\Delta_L(\beta) - \lambda \frac{F(\beta)}{f(\beta)} > \lambda \frac{F(\beta_i^*)}{f(\beta_i^*)} - \lambda \frac{F(\beta)}{f(\beta)} \geq 0 \tag{A.9}$$

where the last inequality follows from the hazard rate assumption (Assumption 1), and the fact that β_i^* is the upper bound of the integral in (A.8). However, inequality (A.9) leads to a contradiction with (A.8). It must therefore be that, with mobile firms, $\beta_i^* > \beta_B$.

Note that the monotone hazard rate assumption (Assumption 1) does not guaranteed uniqueness in our two-dimensional mechanism design problem. Yet, Lemma A.4 tells us that the optimal mechanism takes the form of a two-part tariff, and under Assumption 2 conditions (A.1) and (A.3) characterize the optimal mechanism. Therefore, this proof tells us that the result in Theorem 1 must hold for the optimal mechanism.

This completes the proof of Theorem 1. Q.E.D.

Proof of Lemma 2

We show first that it cannot be that $\beta_i^* \geq \beta_B$. Assume $\beta_i^* \geq \beta_B$. This assumption implies that all types with $\beta \geq \beta_B$ move, because $C(\beta) \geq \overline{\theta}$, with equality only if $\beta = \beta_B$ or $\beta_i^* = \beta_B$. Thus, we are left with the welfare from firms $\beta < \beta_B$:

$$W_i = \int_{\underline{\beta}}^{\beta_B} \big[\gamma + D - \beta - \lambda T_B\big](1 - G(C(\beta))f(\beta)d\beta.$$

Note that all staying firms cut because $\beta_i^* \geq \beta_B$ by assumption. This W_i is unaffected if we reduce β_i^* to β_B, as all firms staying still cut and firms with $\beta \geq \beta_B$ move. But we can obtain a discrete improvement in the payoff by reducing the cutoff even further to $\beta_i^* = \beta_B - \epsilon$ where ϵ is small, giving welfare

$$W_i = \int_{\underline{\beta}}^{\beta_B-\epsilon} \big[\gamma + D - \beta - \lambda T_B\big](1 - G(C(\beta))f(\beta)d\beta + \int_{\beta_B-\epsilon}^{\overline{\beta}} \big[\underbrace{\gamma - \lambda(T_B - \beta_i^*)}_{\equiv E}\big](1 - G(C(\beta_i^*)))f(\beta)d\beta$$

where the first term is approximately the same as our first welfare expression for W_i as $\epsilon \approx 0$. The second integral measures the payoff from a mass of polluting firms. Note that this second term, the value of staying firms, is positive because $G(C(\beta_i^*)) < 1$ and

$$E = \gamma - \lambda(T_B - \beta_i^*) = \gamma - \lambda(\beta_B - \overline{\theta} - \beta_B + \epsilon) \approx \gamma + \lambda\overline{\theta} > 0$$

where, for the first step we have used the definition of $T_B = \beta_B - \overline{\theta}$ and $\beta_i^* = \beta_B - \epsilon$, and in the second step we simplify and use $\epsilon \approx 0$. This proves that $\beta_i^* \geq \beta_B$ cannot be optimal.

Given $\beta_i^* < \beta_B$, we can focus on the interior solution in the firm type distribution and set emission price β_i^* set according to first-order condition (A.1):

$$\big(D-(1+\lambda)\beta_i^*\big)\big(1-G(\beta_i^*-T_B)\big)f(\beta_i^*) - \int_{\beta_i^*}^{\overline{\beta}} \Big(\big[\gamma-\lambda(T_B-\beta_i^*)\big]g(\beta_i^*-T_B)f(\beta) - \lambda\big(1-G(\beta_i^*-T_B)\big)f(\beta)\Big)d\beta = 0.$$

Solving the integral and using the definition of η and $-\Delta(\beta_i^*) = \gamma - \lambda(T_B - \beta_i^*)$ this can be written:

$$\beta_i^* = \frac{D}{1+\lambda} - \frac{1}{1+\lambda}\frac{1-F(\beta_i^*)}{f(\beta_i^*)}\Big(\frac{\Delta(\beta_i^*)}{\eta(C(\beta_i^*))} - \lambda\Big) \equiv \beta_1$$

Q.E.D

Proof of Proposition 1

The proof proceeds as follows. For all $\beta \leq \beta_i^*$ the marginal surplus, by definition, is:

$$MS_i(C(\beta),\beta) = [\gamma + D - \beta - \lambda T_i^*]g(C(\beta))f(\beta) - \lambda(1 - G(C(\beta)))f(\beta)$$

$$= g(\beta - T_i^*)f(\beta)\big[\gamma + D - \beta - \lambda T_i^* + \lambda\eta(\beta - T_i^*)\big] \tag{A.10}$$

where we have used the definition for $\eta = -(1-G)/g$ and the cost for cutting firms: $C(\beta) = \beta - T_i^*$. Pigouvian transfer T_P is defined based on the marginal surplus for the last cutting type, $\beta = \beta_P$:

$$MS_i(C(\beta_P),\beta_P) = g(\beta)f(\beta_P)\big[\gamma + D - \beta_P - \lambda T_P + \lambda\eta(\beta - T_P)\big] = 0$$

$$\Rightarrow$$

$$\big[\gamma + \lambda\beta_P - \lambda T_P + \lambda\eta(\beta - T_P)\big] = 0. \tag{A.11}$$

For the last step we have used $\beta_P = D/(1+\lambda)$. The proof is by contradiction. We assume that (i) $\lambda\eta'(C) < 1$ holds for $\underline{\beta} - T_P < C < \beta_P - T_P$ and (ii) $\beta_i^* \leq \beta_P$. Then we show that this leads to a positive $MS_i(C(\beta),\beta)$, given in (A.10), for all $\beta \leq \beta_P$, and show contradiction.

Consider the optimal transfer, T_i^*. From equation (A.2) we observe that assumption (ii): $\beta_i^* \leq \beta_P$ is equivalent to $\mu_i(\beta_i^*, \overline{\beta}) \geq 0$:

$$\int_{\beta_i^*}^{\overline{\beta}} \big[\gamma + \lambda(\beta_i^* - T_i^*)\big]g(\beta_i^* - T_i^*)f(\beta) - \lambda(1 - G(\beta_i^* - T_i^*))f(\beta)d\beta \geq 0$$

$$\Rightarrow$$

$$\gamma + \lambda C_i^* + \lambda\eta(C_i^*) \geq 0$$

For $C_i^* = C_P = \beta_P - T_P$ this holds with strict equality by equation (A.11). As the left-hand side is strictly increasing in $C_i^* = \beta_i^* - T_i^*$, we must have $C_i^* \geq C_P$. This together with $\beta_i^* < \beta_P$ implies $T_i^* \leq T_P$. Use this observation to rewrite equation (A.10):

$$MS_i(C(\beta),\beta) = g(\beta)f(\beta)\big[\gamma + D - \beta \underbrace{-\lambda T_i^*}_{\geq -\lambda T_P} + \underbrace{\lambda\eta(\beta - T_i^*)}_{\geq \lambda\eta(\beta - T_P)}\big]$$

$$\geq g(\beta)f(\beta)\big[\underbrace{\gamma + D - \beta - \lambda T_P + \lambda\eta(\beta - T_P)}_{\equiv L}\big] > 0 \quad \text{for all } \underline{\beta} \leq \beta < \beta_P. \tag{A.12}$$

Where the last inequality follows from observing that term L is zero at $\beta = \beta_P$ and, by assumption (i): $\lambda\eta'(C) < 1$, decreasing in β for all $\beta \leq \beta_P$. Inequality (A.12) leads to $\mu_i(\underline{\beta}, \beta_i^*) = \int_{\underline{\beta}}^{\beta_i^*} MS_i(C(\beta),\beta)d\beta > 0$, implying $\beta_i^* > \beta_P$ by equation (A.4); a contradiction to assumption (ii), $\beta_i^* \leq \beta_P$. Q.E.D.

Proof of Proposition 2.

The policy maker is constrained to use a purely global mechanism, that is, the same externality price in both countries: $\beta_i^* = \beta_j^* = \beta^*$. The transfer to foreign firms is $T_j = \beta^*$ if they cut $(\beta \leq \beta^*)$, and $T_j = 0$ otherwise $(\beta > \beta^*)$ (see Lemma 3). Firms under the global cap $(\beta \leq \beta^*)$ face the same cost difference across locations regardless of their type: $C(\beta) = C_i(\beta) - C_j(\beta) = (\beta - T_i^*) - (\beta - \beta^*) = \beta^* - T_i^*$. The regulator is left to find T_i^* and β^* to maximize the social welfare:

$$W_i = \int_{\underline{\beta}}^{\beta^*} \Big(\gamma + D - \beta - \lambda T_i^*\Big)\phi_i(\beta^* - T_i^*, \beta) + \Big(D - (1+\lambda)\beta^*\Big)\phi_j(\beta^* - T_i^*, \beta)d\beta +$$
$$\int_{\beta^*}^{\overline{\beta}} \Big(\gamma + \lambda(\beta^* - T_i^*)\Big)\phi_i(\beta^* - T_i^*, \beta)d\beta.$$

The first line captures firms that cut either home or abroad, and their mass is just $\phi_i(C(\beta), \beta) + \phi_j(C(\beta), \beta) = f(\beta)$. Everything with $C^* \equiv \beta^* - T_i^*$ appears in both optimality conditions $\frac{\partial W_i}{\partial \beta^*} = 0$ and $\frac{\partial W_i}{\partial T_i^*} = 0$ but with opposite signs. Using this observation, write the first-order condition with respect to the global price, β^*:

$$\frac{\partial W_i}{\partial \beta^*} = \Big(D - (1+\lambda)\beta^*\Big)\phi_i(\beta^* - T_i^*, \beta) + \Big(D - (1+\lambda)\beta^*\Big)\phi_j(\beta^* - T_i^*, \beta) +$$
$$\int_{\underline{\beta}}^{\beta^*} -(1+\lambda)\phi_j(\beta^* - T_i^*, \beta)d\beta + \frac{\partial W_i}{\partial C^*} = 0.$$

Similarly, after some scrutiny, we obtain

$$\frac{\partial W_i}{\partial T_i^*} = \int_{\underline{\beta}}^{\beta^*} -\lambda\phi_i(\beta^* - T_i^*, \beta)d\beta - \frac{\partial W_i}{\partial C^*} = 0$$

Merge this with the integral on the second line of $\frac{\partial W_i}{\partial \beta^*} = 0$ to write this second line as

$$\int_{\underline{\beta}}^{\beta^*} \big(G(\beta^*) + \lambda\big) f(\beta)d\beta,$$

where we used $\phi_j(C(\beta), \beta) = G(C(\beta))f(\beta)$. Use this to write $\frac{\partial W_i}{\partial \beta_i^*} = 0$ as

$$\big(D - (1+\lambda)\beta^*\big)f(\beta^*) - \int_{\underline{\beta}}^{\beta^*} \big(G(\beta^*) + \lambda\big) f(\beta)d\beta = 0$$

$$\Rightarrow \beta^* = \frac{D}{1+\lambda} - \frac{G(\beta^*) + \lambda}{1+\lambda} \frac{F(\beta^*)}{f(\beta^*)} < \beta_P$$

This is the uniform emission price in the purely-global mechanism. Q.E.D.

Proof of Lemma 3. (Two-part tariff)

The two-part form is shown by Lemma A.6 below.

Proof of Theorem 2 (Global mechanism).

Lemmas A.1 and A.2 continue to hold. We begin by deriving a global counterpart for Lemma A.3:

Lemma A.5. *In the optimal global mechanism ($C_j(\beta) \leq 0$), actions take a bang-bang form where $X_i(\beta) = \{0, 1\}$ and $X_j(\beta) = \{0, 1\}$*

Proof. Objective function (1) can be written as:

$$\max_{X_i(\beta), C_i(\beta), X_j(\beta), C_j(\beta)} \int_{\underline{\beta}}^{\overline{\beta}} \Big(\gamma + DX_i(\beta) - (1+\lambda)\beta X_i(\beta) + \lambda C_i(\beta)\Big)\phi_i\big(C_i(\beta) - C_j(\beta), \beta\big) + \Big(DX_j(\beta) - (1+\lambda)\beta X_j(\beta) + (1+\lambda)C_j(\beta)\Big)\phi_j\big(C_i(\beta) - C_j(\beta), \beta\big)d\beta$$

s.t. $C'_k(\beta) = -X_k(\beta)$ holds for all β and $k = i, j$. Denoting the two co-state variables by $\nu_i(\beta)$ and $\nu_j(\beta)$, the Hamiltonian for this problem reads:

$$\mathcal{H} = \big(\gamma + DX_i(\beta) - (1+\lambda)\beta X_i(\beta) + \lambda C_i(\beta)\big)\phi_i(C_i(\beta) - C_j(\beta), \beta) + \big(DX_j(\beta) - (1+\lambda)\beta X_j(\beta) + (1+\lambda)C_j(\beta)\big)\phi_j\big(C_i(\beta) - C_j(\beta), \beta\big) + \nu_i(\beta)X_i(\beta) + \nu_j(\beta)X_j(\beta)$$

The Hamiltonian is linear in the controls $X_i(\beta)$ and $X_j(\beta)$, and the necessary conditions for optimality imply that $X_i(\beta)$ and $X_j(\beta)$ take bang-bang forms: $X_i(\beta) = \{0, 1\}$ and $X_j(\beta) = \{0, 1\}$. Q.E.D.

Lemma A.6. *The optimal policy takes the two-part tariff form in Lemma 3*

Proof. There exists a solution to the problem as stated in Lemma A.3 by Filippov-Cesari Theorem (Theorem 8, page 132, Seierstad and Sydsæter 1987). Based on Lemmas A.1, A.2 and A.5, the mechanism that satisfies the incentive compatibility conditions takes a threshold form and boilds down to finding a tuple $(\beta_i^*, \beta_j^*, T_i^*, T_j^*)$ that defines $M_i(\beta)$ and $M_j(\beta)$ through base compensations and thresholds for cuts at home and abroad. It can be observed that the optimal transfer to moving firms takes the form: $T_j(\beta) = \beta_j^*$ if they cut ($\beta \leq \beta_j^*$), and $T_j(\beta) = 0$ otherwise ($\beta > \beta_j^*$): It is welfare-reducing to pay firms in j more than necessary to incentivize emission reductions. Q.E.D

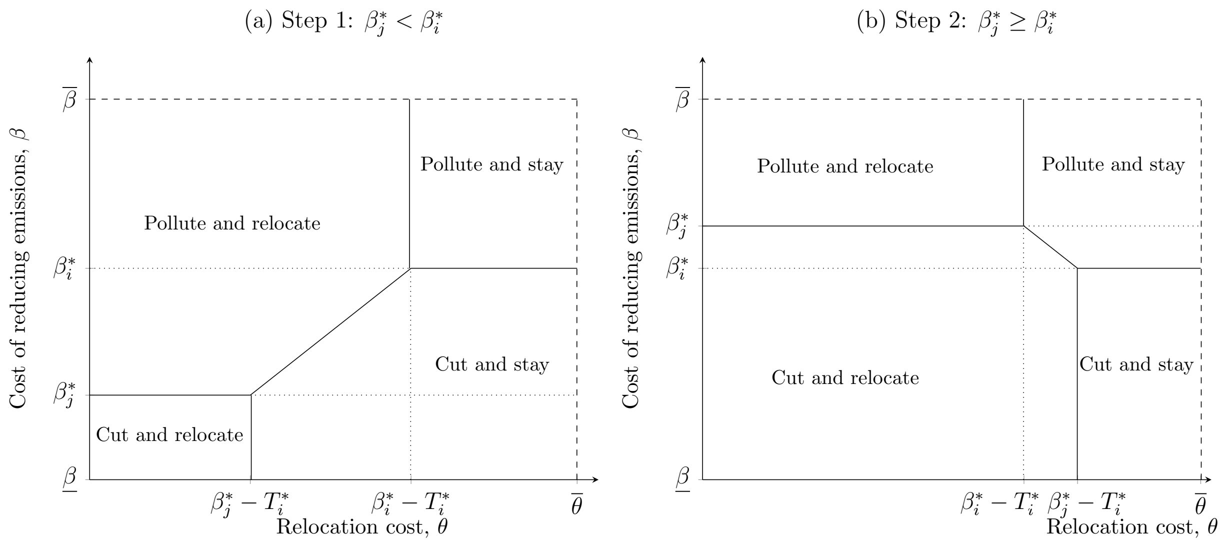


Figure A.2: Graphical illustration of the two-dimensional type space in the global mechanism

This form for the transfer to moving firms gives the cost differential $C(\beta)$ as follows. Consider $\beta_j^* < \beta_i^*$. When firms cut in both regimes ($\beta \leq \beta_j^*$), it becomes $C(\beta) = C_i(\beta) - C_j(\beta) = \beta_j^* - T_i^*$, that is, it is independent of their type β. When firms cut only when they stay ($\beta_j^* < \beta \leq \beta_i^*$), we have $C(\beta) = \beta - T_i^*$, and finally, when firms cut in neither location ($\beta_i^* < \beta$), we obtain $C(\beta) = \beta_i^* - T_i^*$, that is again independent of β. Consider $\beta_i^* < \beta_j^*$. When firms cut in both regimes ($\beta \leq \beta_i^*$), we have $C(\beta) = C_i(\beta) - C_j(\beta) = \beta_j^* - T_i^*$. When firms cut after moving ($\beta_i^* < \beta \leq \beta_j^*$), $C(\beta) = C_i(\beta) - C_j(\beta) = (\beta_i^* - T_i^*) - (\beta - \beta_j^*)$. When firms cut in neither location ($\beta_j^* < \beta$), we obtain again $C(\beta) = \beta_i^* - T_i^*$. The welfare objective is built from these expressions in various stages of the proof.

The optimal mechanism partitions the type space into four areas as shown graphically for the two steps in Figure A.2. Before going to the formal proof, we can use the graph to observe that an increase in global price β_j^* enlarges the area where firms "cut and relocate" in two ways: (i) It incentives leaving firms to cut externalities, and (ii) it also induces more relocation by low-cost firms who already cut in location i. The optimal global price optimally balances these two effects.

We prove Theorem 2 in two steps. Denote the benchmark for pure global mechanism as defined in Proposition 2: $\beta^* = \frac{D}{1+\lambda} - \frac{G(\beta^*)+\lambda}{1+\lambda}\frac{F(\beta^*)}{f(\beta^*)}$. To begin with, we consider the case that $\beta_j^* < \beta_i^*$ and show that this implies $\beta_i^* > \beta^*$ (Step 1). Second, we consider the case $\beta_j^* \geq \beta_i^*$ and show that this leads to a contradiction (Step 2).

Step 1. We begin by considering the case $\beta_j^* < \beta_i^*$. In that case, the maximization problem can be written as:

$$\max_{\beta_i^*, \beta_j^*, T_i^*} W_i = \int_{\underline{\beta}}^{\beta_j^*} \big(\gamma + D - \beta - \lambda T_i^*\big)\phi_i(\beta_j^* - T_i^*, \beta) + \big(D - (1+\lambda)\beta_j^*\big)\phi_j(\beta_j^* - T_i^*, \beta) d\beta +$$
$$\int_{\beta_j^*}^{\beta_i^*} \big(\gamma + D - \beta - \lambda T_i^*\big)\phi_i(\beta - T_i^*, \beta) +$$
$$\int_{\beta_i^*}^{\overline{\beta}} \big(\gamma + \lambda(\beta_i^* - T_i^*)\big)\phi_i(\beta_i^* - T_i^*, \beta) d\beta.$$

The first integral denotes firms under the global cap β_j^*, the second integral are the firms under only a local price β_i^* and the last integral the firms who do not cut in either location (see Figure A.2a for a graphical illustration).

To simplify notation, we use the definition of the net welfare effect of relocation, as defined in equation (2). For firms under the global cap ($\beta \leq \beta_j^*$), we have $-\Delta_G(\beta) = \gamma - \lambda T_i^* - \beta + (1+\lambda)\beta_j^*$, and, for those under the local cap ($\beta_j^* < \beta \leq \beta_i^*$), we have $-\Delta_L(\beta) = \gamma - \lambda T_i^* + D - \beta$.

We aim to show $\beta_j^* < \beta_i^* \Rightarrow \beta_i^* > \beta^*$ so that the local externality price is above the uniform-price benchmark. The proof is by contradiction: assume that $\beta_i^* \leq \beta^*$ which, by $\beta_j^* < \beta_i^*$, also implies that $\beta_j^* < \beta^*$. To show contradiction, we combine the first-order conditions for β_j^* (Step 1a), T_i^* (Step 1b) and β_i^* (Step 1c). The contradiction is shown at the end of Step 1c.

Step 1a. The first-order condition with respect to the global cap, β_j^*, is:

$$(D - (1+\lambda)\beta_j^*)\phi_j(C(\beta_j^*), \beta_j^*) + \int_{\underline{\beta}}^{\beta_j^*} \Big[\Delta_G(\beta)\phi_j'(C(\beta_j^*), \beta) - (1+\lambda)\phi_j(C(\beta_j^*), \beta)\Big] d\beta = 0, \quad \text{(A.13)}$$

where we used expressions (5) to write $\phi_j(C(\beta), \beta) = G(C(\beta))f(\beta)$, and $\phi_j'(C(\beta), \beta) = g(\beta)f(\beta)$ where, to save notation, we write $G(C(\beta)) = G(\beta)$ and $g(C(\beta)) = g(\beta)$. In the text, this condition is expressed in terms of the marginal surplus from firms under the global cap ($\beta^* \leq \beta_j^*$), $MS_j(C(\beta), \beta) = -\Delta_G(\beta)\phi_j'(C(\beta_j^*), \beta) + (1+\lambda)\phi_j(C(\beta_j^*), \beta)$. That is,

$$D - (1+\lambda)\beta_j^* = \frac{\mu_j(\underline{\beta}, \beta_j^*)}{\phi_j(C(\beta_j^*), \beta_j^*)} \quad \text{(A.14)}$$

with

$$\mu_j(\underline{\beta}, \beta_j^*) = \int_{\underline{\beta}}^{\beta_j^*} MS_j(C(\beta), \beta) d\beta.$$

In this proof we work with (A.13). It can be written as:

$$D-(1+\lambda)\beta_j^* + \int_{\underline{\beta}}^{\beta_j^*} \Big(\Delta_G(\beta)\frac{g(\beta_j^*)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)} - (1+\lambda)\frac{f(\beta)}{f(\beta_j^*)}\Big) d\beta = 0.$$

Now, we use the condition that $\beta_j^* < \beta_i^* < \beta^*$:

$$\begin{aligned}\int_{\underline{\beta}}^{\beta_j^*} \Delta_G(\beta)\frac{g(\beta_j^*)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta - (1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} &= -\big(D-(1+\lambda)\beta_j^*\big) < -\big(D-(1+\lambda)\beta^*\big) = \\ &-\big(G(\beta^*)+\lambda\big)\frac{F(\beta^*)}{f(\beta^*)} < -\big(G(\beta_i^*)+\lambda\big)\frac{F(\beta_j^*)}{f(\beta_j^*)}\end{aligned} \tag{A.15}$$

Where the first inequality follows from the assumption $\beta_j^* < \beta^*$, the second inequality follows from the definition of β^* and the last inequality follows from the fact that both $G(\beta)$ and $F(\beta)/f(\beta)$ are increasing in β (Assumption 1). The inequality (A.15) above can be rewritten as:

$$\int_{\underline{\beta}}^{\beta_j^*} \Delta_G(\beta)\frac{g(\beta_j^*)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta < \big(1-G(\beta_i^*)\big)\frac{F(\beta_j^*)}{f(\beta_j^*)}$$

$$\Rightarrow$$

$$\int_{\underline{\beta}}^{\beta_j^*} -\Delta_G(\beta)\frac{g(\beta_j^*)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta > -G(\beta_j^*)\frac{F(\beta_j^*)}{f(\beta_i^*)} > -G(\beta_i^*)\frac{F(\beta_i^*)}{f(\beta_i^*)} \tag{A.16}$$

where for the last step we used the assumption that $G(\beta)$ and $F(\beta)/f(\beta)$ are increasing (Assumption 1) and $\beta_j^* < \beta_i^*$.

Step 1b. The first-order condition for the optimal β_i^* is

$$\big(D-(1+\lambda)\beta_i^*\big)\phi_i(\beta_i^*-T_i^*,\beta_i^*) + \underbrace{\int_{\beta_i^*}^{\overline{\beta}} \big(\gamma+\lambda(\beta_i^*-T_i^*)\big)\phi_i'(\beta_i^*-T_i^*,\beta) + \lambda\phi_i(\beta_i^*-T_i^*,\beta)d\beta}_{-\mu_i(\beta_i^*,\overline{\beta})} = 0$$

Use expressions $\phi_i(C(\beta),\beta)) = (1-G(\beta))f(\beta)$, and $\phi_i'(C(\beta),\beta) = -g(\beta)f(\beta)$ and solve the integral to write:

$$\big(D-(1+\lambda)\beta_i^*\big)(1-G(\beta_i^*)) = \Big((\gamma+\lambda(\beta_i^*-T_i^*))g(\beta_i^*) - \lambda(1-G(\beta_i^*))\Big)\frac{1-F(\beta_i^*)}{f(\beta_i^*)}.$$

Divide both sides by $1-G(\beta_i^*)$ to get:

$$\Big([\gamma+\lambda(\beta_i^*-T_i^*)]\frac{g(\beta_i^*)}{1-G(\beta_i^*)} - \lambda\Big)\frac{1-F(\beta_i^*)}{f(\beta_i^*)} = D-(1+\lambda)\beta_i^* > 0$$

where the inequality follows from $\beta_i^* < \beta^* < \frac{D}{1+\lambda}$. It immediately follows that:

$$\gamma + \lambda(\beta_i^* - T_i^*) > 0 \tag{A.17}$$

Write the definition of $\Delta_L(\beta)$,

$$-\Delta_L(\beta) = \gamma + \lambda(\beta_i^* - T_i^*) + D - \beta - \lambda\beta_i^* > D - (1+\lambda)\beta_i^* \geq (G(\beta^*) + \lambda)\frac{F(\beta^*)}{f(\beta^*)} > \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)}, \tag{A.18}$$

where: for the first inequality we have used (A.17) and the fact that $\beta < \beta_i^*$; for the second inequality we have used $\beta_i^* < \beta^*$; and for the last inequality we have used $G(\beta^*) > 0$, the fact that $F(\beta)/f(\beta)$ is increasing in β and $\beta_i^* < \beta^*$.

In the text, the first-order condition for the optimal β_i^* is expressed in terms of the marginal surplus from firms that may move and not cut ($\beta_i^* \leq \beta$), $MS_i(C(\beta), \beta) = \Delta(\beta)\phi_i'(C(\beta), \beta) - \lambda\phi_i(C(\beta), \beta)$. That is,

$$D - (1+\lambda)\beta_i^* = \frac{\mu_i(\beta_i^*, \overline{\beta})}{\phi_i(C(\beta), \beta)} \tag{A.19}$$

with

$$\mu_i(\beta_i^*, \overline{\beta}) = \int_{\underline{\beta}}^{\beta_i^*} MS_i(C(\beta), \beta) d\beta.$$

Step 1c. Next, after some scrutiny, we write the first-order condition with respect to T_i^* by using also the first-order condition for the optimal β_i^* as follows:

$$D - (1+\lambda)\beta_i^* + \int_{\underline{\beta}}^{\beta_j^*} \Delta_G(\beta)\frac{\phi_i'(\beta_j^* - T_i^*, \beta)}{\phi_i(\beta_i^* - T_i^*, \beta_i^*)} - \lambda\frac{\phi_i(\beta_j^* - T_i^*, \beta)}{\phi_i(\beta_i^* - T_i^*, \beta_i^*)} d\beta + \int_{\beta_j^*}^{\beta_i^*} \Delta_L(\beta)\frac{\phi_i'(\beta - T_i^*, \beta)}{\phi_i(\beta_i^* - T_i^*, \beta_i^*)} - \lambda\frac{\phi_i(\beta - T_i^*, \beta)}{\phi_i(\beta_i^* - T_i^*, \beta_i^*)} d\beta = 0$$

Use expressions $\phi_i(C(\beta), \beta)) = (1 - G(\beta))f(\beta)$, and $\phi_i'(C(\beta), \beta) = -g(\beta)f(\beta)$ to write:

$$D - (1+\lambda)\beta_i^* + \int_{\underline{\beta}}^{\beta_j^*} -\Delta_G(\beta)\frac{g(\beta_j^*)}{1 - G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)} - \lambda\frac{1 - G(\beta_j^*)}{1 - G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)} d\beta + \int_{\beta_j^*}^{\beta_i^*} -\Delta_L(\beta)\frac{g(\beta)}{1 - G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)} - \lambda\frac{1 - G(\beta)}{1 - G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)} d\beta = 0 \tag{A.20}$$

Using the assumption $\beta_i^* \leq \beta^*$, implying: $D - (1+\lambda)\beta_i^* \geq (G(\beta^*) + \lambda)\frac{F(\beta^*)}{f(\beta^*)}$, and simplifying, we get:

$$\int_{\underline{\beta}}^{\beta_j^*} -\Delta_G(\beta)\frac{g(\beta_j^*)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta - \lambda\frac{1-G(\beta_j^*)}{1-G(\beta_i^*)}\frac{F(\beta_j^*)}{f(\beta_i^*)}+$$
$$\int_{\beta_j^*}^{\beta_i^*} -\Delta_L(\beta)\frac{g(\beta)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)} \underbrace{-\lambda\frac{1-G(\beta)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta}_{\equiv H} \leq -\big(G(\beta^*)+\lambda\big)\frac{F(\beta^*)}{f(\beta^*)} \tag{A.21}$$

Integrate term H in equation (A.21) by parts (Note, that $dG(C(\beta)) = C'(\beta)g(C(\beta)) = g(C(\beta))$, because $C'(\beta) = 1$):

$$H = -\int_{\beta_j^*}^{\beta_i^*} \lambda\frac{1-G(\beta)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta = -\lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} + \lambda\frac{1-G(\beta_j^*)}{1-G(\beta_i^*)}\frac{F(\beta_j^*)}{f(\beta_i^*)} - \int_{\underline{\beta}}^{\beta_i^*} \lambda\frac{g(\beta)}{1-G(\beta_i^*)}\frac{F(\beta)}{f(\beta_i^*)}d\beta$$

Use this, together with (A.16), to write the left-hand side of equation (A.21) as:

$$-\Big(G(\beta_i^*)+\lambda\Big)\frac{F(\beta_i^*)}{f(\beta_i^*)} + \int_{\beta_j^*}^{\beta_i^*} \frac{g(\beta)}{f(\beta_i^*)}\frac{1}{1-G(\beta_i^*)}\Big(\underbrace{-\Delta_L f(\beta) - \lambda F(\beta)}_{\equiv I}\Big)d\beta > -\big(G(\beta^*)+\lambda\big)\frac{F(\beta^*)}{f(\beta^*)} \tag{A.22}$$

The inequality follows from, firstly, the fact that $G(\beta)$ and $F(\beta)/f(\beta)$ are increasing in β and secondly, because the second term is positive since, as can be seen by using the definition for $\Delta_L(\beta)$ in (A.18):

$$I = -\Delta_L(\beta)f(\beta) - \lambda F(\beta) > \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)}f(\beta) - \lambda F(\beta) \geq 0, \tag{A.23}$$

However, there is contradiction between (A.21) and (A.22). This proves that if $\beta_i^* > \beta_j^*$, we must have $\beta_i^* > \beta^*$.

Step 2. Let us next assume $\beta_i^* \leq \beta_j^*$. This is shown to lead to a contradiction both when $\gamma + \lambda(\beta_i^* - T_i^*) > 0$ (Step 2a) and $\gamma + \lambda(\beta_i^* - T_i^*) \leq 0$ (Step 2b). With $\beta_i^* \leq \beta_j^*$ the maximization problem is written as:

$$\max_{\beta_i^*,\beta_j^*,T_i^*} W_i = \int_{\underline{\beta}}^{\beta_i^*} \big(\gamma + D - \beta - \lambda T_i^*\big)\phi_i(\beta_j^* - T_i^*, \beta) + \big(D - (1+\lambda)\beta_j^*\big)\phi_j(\beta_j^* - T_i^*, \beta)d\beta+$$
$$\int_{\beta_i^*}^{\beta_j^*} \big(\gamma + \lambda(\beta_i^* - T_i^*\big)\phi_i\big((\beta_i^* - T_i^*) - (\beta - \beta_j^*), \beta\big) + \big(D - (1+\lambda)\beta_j^*\big)\phi_j\big((\beta_i^* - T_i^*) - (\beta - \beta_j^*), \beta\big)d\beta+$$
$$\int_{\beta_j^*}^{\overline{\beta}} \big(\gamma + \lambda(\beta_i^* - T_i^*)\big)\phi_i(\beta_i^* - T_i^*, \beta)d\beta.$$

In this case, firms cut only in j if $\beta_i^* \leq \beta \leq \beta_j^*$. In this cost range, the welfare effects of the relocation under the foreign cap is defined as $-\Delta_F(\beta) = \gamma + \lambda(\beta_i^* - T_i^*) - D + (1+\lambda)\beta_j^*$, and the cost difference is $C(\beta) = C_i(\beta) - C_j(\beta) = (\beta_i^* - T_i^*) - (\beta - \beta_j^*)$, implying $C'(\beta) = -1$.

Step 2a. Assume first that $\gamma + \lambda(\beta_i^* - T_i^*) > 0$. In that case, we can write:

$$-\Delta_G(\beta) = \gamma+\lambda(\beta_i^*-T_i^*)+(1+\lambda)\beta_j^*-\beta \geq -\Delta_G(\beta_i^*) = \gamma+\lambda(\beta_i^*-T_i^*)+(1+\lambda)\beta_j^*-\beta_i^* > 0 \quad \text{(A.24)}$$

where the first inequality follows from the fact that β_i^* is the upper integral bound ($\beta \leq \beta_i^*$) and the last inequality follows from $\gamma + \lambda(\beta_i^* - T_i^*) > 0$ and $\beta_j^* \geq \beta_i^*$. The first-order condition for β_i^* can merged with the first-order condition for T_i^* to obtain:

$$-D + (1+\lambda)\beta_i^* = \int_{\underline{\beta}}^{\beta_i^*} -\Delta_G(\beta)\frac{g(\beta_i^*)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta - \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} > -\lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} \quad \text{(A.25)}$$

where the last inequality follows from $-\Delta_G(\beta) > 0$, as stated in (A.24). Using this, we can write the definition for $\Delta_F(\beta)$ as:

$$-\Delta_F(\beta) = \gamma + \lambda(\beta_i^* - T_i^*) - D + (1+\lambda)\beta_j^* \geq -D + (1+\lambda)\beta_i^* > -\lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} \quad \text{(A.26)}$$

where the first inequality follows from using (A.24), and the for the second inequality we have used (A.25). The first-order condition for the foreign price, β_j^*, can be written as:

$$\begin{aligned} D - (1+\lambda)\beta_j^* + \int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta)\frac{g(\beta_i^*)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta - (1+\lambda)\frac{G(\beta_i^*)}{G(\beta_j^*)}\frac{F(\beta_i^*)}{f(\beta_j^*)} + \\ \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta)\frac{g(\beta)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)} - (1+\lambda)\frac{G(\beta)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta = 0. \end{aligned} \quad \text{(A.27)}$$

From the assumption $\beta_j^* \geq \beta_i^*$ it follows that we must have:

$$-\big(D - (1+\lambda)\beta_j^*\big) \geq -\big(D - (1+\lambda)\beta_i^*\big) \quad \text{(A.28)}$$

Plug the first-order condition for β_i^* in (A.25) and β_j^* in equation (A.27) into (A.28), to write:

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta)\frac{g(\beta_i^*)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta - (1+\lambda)\frac{G(\beta_i^*)}{G(\beta_j^*)}\frac{F(\beta_i^*)}{f(\beta_j^*)} + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta)\frac{g(\beta)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)} \underbrace{-(1+\lambda)\frac{G(\beta)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}}_{\equiv V} d\beta \geq$$

$$\int_{\underline{\beta}}^{\beta_i^*} -\Delta_G(\beta)\frac{g(\beta_i^*)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta - \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} \tag{A.29}$$

Integrate term V by parts, noting that $C'(\beta) = -1$, to write:

$$V = \int_{\beta_i^*}^{\beta_j^*} -(1+\lambda)\frac{G(\beta)}{G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta = -(1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} + (1+\lambda)\frac{G(\beta_i^*)}{G(\beta_j^*)}\frac{F(\beta_i^*)}{f(\beta_j^*)} - \int_{\beta_i^*}^{\beta_j^*}(1+\lambda)\frac{g(\beta)}{G(\beta_j^*)}\frac{F(\beta)}{f(\beta_j^*)}d\beta$$

Plugging term V into equation (A.29) it can be written as:

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta)g(\beta_j^*)f(\beta)\Big(\frac{1}{G(\beta_j^*)}\frac{1}{f(\beta_j^*)} + \frac{1}{1-G(\beta_i^*)}\frac{1}{f(\beta_i^*)}\Big)d\beta$$

$$+\int_{\beta_i^*}^{\beta_j^*} \frac{g(\beta)}{G(\beta_j^*)}\frac{1}{f(\beta_j^*)}\Big(\underbrace{\Delta_F(\beta)f(\beta) - (1+\lambda)F(\beta)}_{\equiv J}\Big)d\beta - \Big((1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} - \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)}\Big) \geq 0 \tag{A.30}$$

From (A.24) we know that the first term is non-positive. We can also see that term J is strictly negative as:

$$J = \Delta_F(\beta)f(\beta) - (1+\lambda)F(\beta) < \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)}f(\beta) - (1+\lambda)F(\beta) \leq 0$$

where the first inequality follows from (A.26) and the second inequality follows from the fact that β_i^* is the lower integral bound and $F(\beta)/f(\beta)$ is increasing in its argument by Assumption 1. Last, we can see that this and the fact that $\beta_i^* > \beta_j^*$ imply that the last term in equation (A.30) is strictly negative. The left-hand side of (A.30) is therefore strictly negative. However, this contradicts the inequality in (A.30). Therefore, $\gamma + \lambda(\beta_i^* - T_i^*) > 0$ implies that $\beta_i^* > \beta_j^*$.

Step 2b. To complete the proof, we are left to show that the case $\gamma + \lambda(\beta_i^* - T_i^*) \leq 0$ also leads to a contradiction. We move to first-order condition for the baseline compensation, T_i^*, and write it as follows:

$$\Big([\gamma + \lambda(\beta_i^* - T_i^*)]\frac{g(\beta_j^*)}{1-G(\beta_j^*)} - \lambda\Big)\frac{1-F(\beta_j^*)}{f(\beta_j^*)} =$$

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta)\frac{g(\beta_i^*)}{1-G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta + \lambda\frac{1-G(\beta_i^*)}{1-G(\beta_j^*)}\frac{F(\beta_i^*)}{f(\beta_j^*)} + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta)\frac{g(\beta)}{1-G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)} + \lambda\frac{1-G(\beta)}{1-G(\beta_j^*)}\frac{f(\beta)}{f(\beta_j^*)}d\beta$$

We can notice that $\gamma + \lambda(\beta_i^* - T_i^*) \leq 0$ implies that the right-hand side must be non-positive:

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) \frac{g(\beta_i^*)}{1-G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta + \lambda \frac{1-G(\beta_i^*)}{1-G(\beta_j^*)} \frac{F(\beta_i^*)}{f(\beta_j^*)} + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) \frac{g(\beta)}{1-G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} + \lambda \frac{1-G(\beta)}{1-G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta \leq 0. \tag{A.31}$$

First, note that the sum of the second and the fourth term in (A.31) is always strictly positive, and a necessary condition for the inequality to hold is therefore:

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) g(\beta_i^*) f(\beta) d\beta + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) g(\beta) f(\beta) d\beta < 0 \tag{A.32}$$

By reorganizing the terms, the inequality (A.31) can be written as:

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) g(\beta_i^*) f(\beta) d\beta + \lambda(1-G(\beta_i^*))F(\beta_i^*) + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) g(\beta) f(\beta) + \lambda(1-G(\beta)) f(\beta) d\beta < 0$$

$$\Rightarrow$$

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) g(\beta_i^*) f(\beta) d\beta - \lambda G(\beta_i^*) F(\beta_i^*) + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) g(\beta) f(\beta) - \lambda G(\beta) f(\beta) d\beta < -\lambda F(\beta_i^*) - \int_{\beta_i^*}^{\beta_j^*} \lambda f(\beta) d\beta$$

$$\Rightarrow$$

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) g(\beta_i^*) f(\beta) d\beta - \lambda G(\beta_i^*) F(\beta_i^*) + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) g(\beta) f(\beta) - \lambda G(\beta) f(\beta) d\beta < -\lambda F(\beta_j^*) \tag{A.33}$$

Write the first-order condition for β_j^*, given in (A.27), as:

$$-\big(D - (1+\lambda)\beta_j^*\big) =$$

$$\int_{\underline{\beta}}^{\beta_i^*} \Delta_G(\beta) \frac{g(\beta_i^*)}{G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta - (1+\lambda) \frac{G(\beta_i^*)}{G(\beta_j^*)} \frac{F(\beta_i^*)}{f(\beta_j^*)} + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta) \frac{g(\beta)}{G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} - (1+\lambda) \frac{G(\beta)}{G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta <$$

$$- \frac{G(\beta_i^*)}{G(\beta_j^*)} \frac{F(\beta_i^*)}{f(\beta_j^*)} \underbrace{- \int_{\beta_i^*}^{\beta_j^*} \frac{G(\beta)}{G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta}_{\equiv K} - \lambda \frac{1}{G(\beta_j^*)} \frac{F(\beta_j^*)}{f(\beta_j^*)} \tag{A.34}$$

where for the inequality we have used (A.33). Integrate the term in the K by parts (note: $C'(\beta) = -1$ when $\beta_i^* \leq \beta \leq \beta_j^*$).

$$K = -\int_{\beta_i^*}^{\beta_j^*} \frac{G(\beta)}{G(\beta_j^*)} \frac{f(\beta)}{f(\beta_j^*)} d\beta = -\frac{F(\beta_j^*)}{f(\beta_j^*)} + \frac{G(\beta_i^*)}{G(\beta_j^*)} \frac{F(\beta_i^*)}{f(\beta_j^*)} - \int_{\beta_i^*}^{\beta_j^*} \frac{g(\beta)}{G(\beta_j^*)} \frac{F(\beta)}{f(\beta_j^*)} d\beta$$

Plug this in to write the inequality in (A.34) as:

$$-\big(D-(1+\lambda)\beta_j^*\big) < -\frac{F(\beta_j^*)}{f(\beta_j^*)} - \lambda\frac{1}{G(\beta_j^*)}\frac{F(\beta_j^*)}{f(\beta_j^*)} - \int_{\beta_i^*}^{\beta_j^*}\frac{g(\beta)}{G(\beta_j^*)}\frac{F(\beta)}{f(\beta_j^*)}d\beta < -(1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} \quad \text{(A.35)}$$

where the last inequality follows from $G(\beta_j^*) \in (0,1]$. From the assumption $\beta_i^* \leq \beta_j^*$, it follows, using equations (A.28) and (A.35), that:

$$-\big(D-(1+\lambda)\beta_i^*\big) \leq -\big(D-(1+\lambda)\beta_j^*\big) < -(1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} < 0 \quad \text{(A.36)}$$

$$\Rightarrow$$

$$\int_{\underline{\beta}}^{\beta_i^*} -\Delta_G(\beta)\frac{g(\beta_i^*)}{1-G(\beta_i^*)}\frac{f(\beta)}{f(\beta_i^*)}d\beta \leq \lambda\frac{F(\beta_i^*)}{f(\beta_i^*)} - (1+\lambda)\frac{F(\beta_j^*)}{f(\beta_j^*)} < 0 \quad \text{(A.37)}$$

where, for the last inequality, we have used the fact that $F(\beta)/f(\beta)$ is increasing and $\beta_i^* \leq \beta_j^*$. Use the definition of $\Delta_F(\beta)$:

$$-\Delta_F(\beta) = \gamma + \lambda(\beta_i^* - T_i^*) - D + (1+\lambda)\beta_j^* < 0 \quad \text{(A.38)}$$

Where the first inequality follows from the assumption $\gamma + \lambda(\beta_i^* - T_i^*) \leq 0$ and using (A.35). Equations (A.37) and (A.38) together imply a contradiction to equation (A.32) when $\gamma + \lambda(\beta_i^* - T_i^*) \leq 0$: equation (A.37) shows that the first term in (A.32) must be positive, and equation (A.38) shows that the second term in (A.32) must be positive, implying a contradiction. This completes the proof: Steps 2a and 2b together prove that $\beta_i^* > \beta_j^*$. Q.E.D.

Proof of Proposition 3

Best responses in the local-mechanism game. Regimes $n = \{i, j\}$ non-cooperatively choose a two-part tariff; externality prices β_i^* and β_j^* and base compensations T_i^* and T_j^*. We find first the best response by i to policies in j, and then analyze the implications for a symmetric equilibrium.

Consider the case where $\beta_i^* \geq \beta_j^*$. Taking (T_j^*, β_j^*) by regime j as given, regime i finds the optimal two part tariff (T_i^*, β_i^*) by solving (or identically for j if $\beta_i^* < \beta_j^*$):

$$\max_{\beta_i^*, T_i^*} W_i = \int_{\underline{\beta}}^{\beta_j^*} \left(\gamma + D - \beta - \lambda T_i^*\right)\phi_i(T_j^* - T_i^*, \beta) + D\phi_j(T_j^* - T_i^*, \beta)d\beta +$$
$$\int_{\beta_j^*}^{\beta_i^*} \left(\gamma + D - \beta - \lambda T_i^*\right)\phi_i(\beta - \beta_j^* - T_i^* + T_j^*, \beta)d\beta +$$
$$\int_{\beta_i^*}^{\overline{\beta}} \left(\gamma + \lambda(\beta_i^* - T_i^*)\right)\phi_i(\beta_i^* - \beta_j^* + T_j^* - T_i^*, \beta)d\beta.$$

The first integral denotes firms under the global cap β_j^* ($\beta \leq \beta_j^*$). In this case the firm cuts in both countries, firms' β-type becomes irrelevant for the location and the cost differential becomes $C(\beta_j^*) = T_j^* - T_i^*$. To simplify notation, denote the welfare effect of relocation in this case $-\Delta_G(\beta) = \gamma - \lambda T_i^* - \beta$. The second integral are the firms that cut only if they stay ($\beta_j^* < \beta \leq \beta_i^*$). For them, the cost differential depends on their type, $C(\beta) = \beta - \beta_j^* - T_i^* + T_j^*$, and the welfare effect of relocation is $-\Delta_L = \gamma + D - \beta - \lambda T_i^*$. The last integral the firms who do not cut in either location ($\beta > \beta_i^*$). For that part, cost differential $C(\beta_i^*) = \beta_i^* - \beta_j^* + T_j^* - T_i^*$ and welfare effects of relocation $-\Delta_H(\beta) = \gamma + \lambda(\beta_i^* - T_i^*)$.

To find country i's best response to an artificial policy in j we first use Leibniz' rule and maximize country i's welfare with respect to β_i^*:

$$(D - (1 + \lambda)\beta_i^*)\phi_i(C(\beta_i^*), \beta) = \int_{\beta_i^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_i^*), \beta) - \lambda\phi_i(C(\beta_i^*), \beta)d\beta = \mu_i(\beta_i^*, \overline{\beta}) \quad \text{(A.39)}$$

Taking β_j^* and T_j^* as given, the first order condition for the optimal transfer in i, T_i^*, becomes:

$$\int_{\underline{\beta}}^{\beta_j^*} \Delta_G(\beta)\phi_i'(C(\beta_j^*), \beta) - \lambda\phi_i(C(\beta_j^*), \beta)d\beta + \int_{\beta_j^*}^{\beta_i^*} \Delta_L(\beta)\phi_i'(C(\beta), \beta) - \lambda\phi_i(C(\beta), \beta)d\beta + \quad \text{(A.40)}$$
$$\int_{\beta_i^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_i^*), \beta) - \lambda\phi_i(C(\beta_i^*), \beta)d\beta = \mu_i(\underline{\beta}, \overline{\beta}) = 0 \quad \text{(A.41)}$$

Combine equations (A.39) and (A.40) to write:

$$(D-(1+\lambda)\beta_i^*)\phi_i(\beta_i^*) = -\mu_i(\underline{\beta},\beta_j^*) - \mu_i(\beta_j^*,\beta_i^*) \tag{A.42}$$

Consider the case where $\beta_i^* < \beta_j^*$. Again, take (T_j^*, β_j^*) by regime j as given, and find the optimal two part tariff (T_i^*, β_i^*) for regime i (or similarly for j if $\beta_i^* \geq \beta_j^*$):

$$\begin{aligned}\max_{\beta_i^*,T_i^*} W_i = &\int_{\underline{\beta}}^{\beta_i^*} \big(\gamma + D - \beta - \lambda T_i^*\big)\phi_i(T_j^* - T_i^*, \beta) + D\phi_j(T_j^* - T_i^*, \beta)d\beta + \\ &\int_{\beta_i^*}^{\beta_j^*} \big(\gamma + \lambda(\beta_i^* - T_i^*)\big)\phi_i(\beta_i^* - \beta - T_i^* + T_j^*, \beta) + D\phi_j(\beta_i^* - \beta - T_i^* + T_j^*, \beta)d\beta \\ &\int_{\beta_j^*}^{\overline{\beta}} \big(\gamma + \lambda(\beta_i^* - T_i^*)\big)\phi_i(\beta_i^* - \beta_j^* + T_j^* - T_i^*, \beta)d\beta.\end{aligned}$$

The first integral denotes firms under the global cap ($\beta \leq \beta_i^*$) whose cost differential is $C(\beta_i^*) = T_j^* - T_i^*$ and the welfare effect of relocation in this case $-\Delta_G(\beta) = \gamma - \lambda T_i^* - \beta$. The second integral now denotes the firms that cut only if they move ($\beta_i^* < \beta \leq \beta_j^*$). For them, the cost differential depends on type, $C(\beta) = \beta_i^* - \beta - T_i^* + T_j^*$, and the welfare effect of relocation is $-\Delta_F = \gamma - D - \beta + \lambda(\beta_i^* - T_i^*)$. The last integral the firms who do not cut in either location ($\beta > \beta_j^*$). For that part, cost differential $C(\beta_j^*) = \beta_i^* - \beta_j^* + T_j^* - T_i^*$ and welfare effects of relocation $-\Delta_H(\beta) = \gamma + \lambda(\beta_i^* - T_i^*)$. Taking β_j^* and T_j^* as given we find the optimal externality price:

$$\begin{aligned}&(D-(1+\lambda)\beta_i^*)\phi_i(C(\beta_i^*),\beta_i^*) \\ = &\int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta)\phi_i'(C(\beta),\beta) - \lambda\phi_i(C(\beta),\beta)d\beta + \int_{\beta_j^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_j^*),\beta) - \lambda\phi_i(C(\beta_j^*),\beta)d\beta \\ &= \mu_i(\beta_i^*,\beta_j^*) + \mu_i(\beta_j^*,\overline{\beta})\end{aligned}$$

The first order condition for the optimal transfer in i, T_i^*, becomes:

$$\begin{aligned}&\int_{\underline{\beta}}^{\beta_i^*} \Delta_L(\beta)\phi_i'(C(\beta_i^*),\beta) - \lambda\phi_i(C(\beta_i^*),\beta)d\beta + \int_{\beta_i^*}^{\beta_j^*} \Delta_F(\beta)\phi_i'(C(\beta),\beta) - \lambda\phi_i(C(\beta),\beta)d\beta + \\ &\int_{\beta_j^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_j^*),\beta) - \lambda\phi_i(C(\beta_j^*),\beta)d\beta = \mu_i(\underline{\beta},\beta_i^*) + \mu_i(\beta_i^*,\beta_j^*) + \mu_i(\beta_j^*,\overline{\beta}) = 0\end{aligned}$$

Combine the two:

$$(D-(1+\lambda)\beta_i^*)\phi_i(C(\beta_i^*),\beta_i^*) = -\mu_i(\underline{\beta},\beta_i^*) \tag{A.43}$$

Characterization of the symmetric equilibrium. Impose symmetry: $\beta_i^* = \beta_j^*$ and $T_i^* = T_j^*$, to write (A.42) and (A.43) as

$$(D - (1+\lambda)\beta_i^*)\phi_i(C(\beta_i^*), \beta_i^*) = -\mu_i(\underline{\beta}, \beta_i^*) \tag{A.44}$$

Symmetry also lets us write the condition for optimal transfer T_i^* as:

$$\mu_i(\underline{\beta}, \beta_i^*) + \mu_i(\beta_i^*, \overline{\beta}) = 0 \tag{A.45}$$

We prove that in any symmetric equilibrium of the game, the emission price is distorted below the first-best levels: $\beta_i^* = \beta_j^* < \beta_P = D/(1+\lambda)$. Proof is by contradiction. Assume $\beta_i^* \geq \beta_P$ which, by equation (A.44) implies $\mu_i(\underline{\beta}, \beta_i^*) \geq 0$. This allows us to write equation (A.45), using $G(\beta)$ as shorthand for $G(C(\beta))$, as:

$$\int_{\beta_i^*}^{\overline{\beta}} \big[\gamma + \lambda(\beta_i^* - T_i^*)\big] g(\beta_i^*) f(\beta) - \lambda\big[1 - G(\beta_i^*)\big] f(\beta) d\beta \leq 0$$

$$\Rightarrow$$

$$\big[\gamma + \lambda(\beta_i^* - T_i^*)\big] g(\beta_i^*) - \lambda\big[1 - G(\beta_i^*)\big] \leq 0 \tag{A.46}$$

We can also write:

$$\mu_i(\underline{\beta}, \beta_i^*) = \int_{\underline{\beta}}^{\beta_i^*} \big[\gamma - \beta - \lambda T_i^*\big] g(\beta_i^*) f(\beta) - \lambda\big[1 - G(\beta_i^*)\big] f(\beta) d\beta \leq \int_{\underline{\beta}}^{\beta_i^*} \big[-\beta - \lambda\beta_i^*\big] g(\beta_i^*) f(\beta) < 0 \tag{A.47}$$

where the first inequality follows from applying equation (A.46). However, by equation (A.44), equation (A.47) must imply $\beta_i^* < \beta_P$; a contradiction. This proves Proposition 3. Q.E.D

Best responses in the global-mechanism game. Both countries set a two-part tariff, (T_i^*, β_i^*) and (T_j^*, β_j^*) at home. In addition, they set one-part tariffs abroad: Country i's foreign price in j is $\beta_{ij}^* \geq 0$ and, likewise, j's policy in i is $\beta_{ji}^* \geq 0$. This creates total emission price $\beta_i^* + \beta_{ji}^*$ in country i and $\beta_j^* + \beta_{ij}^*$ in country j: A firm reduces emissions in i if the total compensation is higher than its abatement cost, $\beta < \beta_i^* + \beta_{ij}^*$ and likewise for j.

Assume first $\beta_i^* + \beta_{ji}^* \geq \beta_j^* + \beta_{ij}^*$. To find best response for country i, we take $(\beta_j^*, T_j^*, \beta_{ji}^*)$ as given and solve (and similarly for j if $\beta_i^* + \beta_{ji}^* < \beta_j^* + \beta_{ij}^*$):

$$\max_{\beta_i^*, T_i^*, \beta_{ij}^*} W_i =$$

$$\int_{\underline{\beta}}^{\beta_j^*+\beta_{ij}^*} \big(\gamma + D - \beta + \beta_{ji}^* - \lambda T_i^*\big)\phi_i(\beta_{ij}^* - \beta_{ji}^* + T_j^* - T_i^*, \beta) + \big(D - (1+\lambda)\beta_{ij}^*\big)\phi_j(\beta_{ij}^* - \beta_{ji}^* + T_j^* - T_i^*, \beta)d\beta +$$

$$\int_{\beta_j^*+\beta_{ij}^*}^{\beta_i^*+\beta_{ji}^*} \big(\gamma + D - \beta + \beta_{ji}^* - \lambda T_i^*\big)\phi_i(\beta - \beta_j^* - T_i^* + T_j^* - \beta_{ji}^*, \beta)d\beta$$

$$\int_{\beta_i^*+\beta_{ji}^*}^{\overline{\beta}} \big(\gamma + \lambda(\beta_i^* - T_i^*)\big)\phi_i(\beta_i^* - \beta_j^* + T_j^* - T_i^*, \beta)d\beta.$$

The first integral denotes firms under the global cap; emission price in j plus the top-up by i ($\beta \le \beta_j^* + \beta_{ij}^*$). In this case the firm cuts in both countries, firms' β-type becomes irrelevant for the location and the cost differential becomes $C(\beta_j^* + \beta_{ij}^*) = T_j^* - T_i^* + \beta_{ij}^* - \beta_{ji}^*$. To simplify notation, denote the welfare effect of relocation in this case $-\Delta_G(\beta) = \gamma - \lambda T_i^* - \beta + \beta_{ji}^* + (1+\lambda)\beta_{ij}^*$. Note that country j's payment is a cross-border transfer and it shows up in country i's welfare. The second integral are the firms that cut only if they stay ($\beta_j^* + \beta_{ij}^* < \beta \le \beta_i^* + \beta_{ji}^*$). For them, the cost differential depends on their type, $C(\beta) = \beta - \beta_j^* - T_i^* + T_j^* - \beta_{ji}^*$, and the welfare effect of relocation is $-\Delta_L = \gamma + D - \beta + \beta_{ji}^* - \lambda T_i^*$. The last integral the firms who do not cut in either location ($\beta > \beta_i^* + \beta_{ji}^*$). For that part, cost differential $C(\beta_i^* + \beta_{ji}^*) = \beta_i^* - \beta_j^* + T_j^* - T_i^*$ and welfare effects of relocation $-\Delta_H(\beta) = \gamma + \lambda(\beta_i^* - T_i^*)$.

To find a best response to an artificial policy by country j (β_j^* and β_{ji}^*), we maximize with respect to β_i^*:

$$\big(D - (1+\lambda)\beta_i^*\big)\phi_i(C(\beta_i^* + \beta_{ji}^*), \beta_i^* + \beta_{ji}^*) = \underbrace{\int_{\beta_i^*+\beta_{ji}^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_i^* + \beta_{ji}^*), \beta) - \lambda\phi_i(C(\beta_i^* + \beta_{ji}^*), \beta)d\beta}_{\equiv \mu_i(\beta_i^* + \beta_{ji}^*, \overline{\beta})} \tag{A.48}$$

Maximize with respect to β_{ij}^*:

$$\big(D - (1+\lambda)\beta_{ij}^*\big)\phi_j(C(\beta_j^* + \beta_{ij}^*), \beta_j^* + \beta_{ij}^*) \le \underbrace{\int_{\underline{\beta}}^{\beta_j^*+\beta_{ij}^*} -\Delta_G(\beta)\phi_j'(\beta_i^* + \beta_{ji}^*) + (1+\lambda)\phi_j(\beta_i^* + \beta_{ji}^*)d\beta}_{\equiv -\mu_{ij}(\underline{\beta}, \beta_j^* + \beta_{ij}^*)} \tag{A.49}$$

which holds with strict inequality if $\beta_{ij}^* = 0$ and strict equality otherwise. Note that the right-hand side does not vanish when β_{ij}^* as long as the local price, β_j^*, is positive. Therefore it may not be optimal to set the foreign price in a strategic setting between countries, in contrast to the unilateral mechanism analyzed in Section 4. In addition, T_i^* is optimally set at $\mu_i(\underline{\beta}, \beta_j^* + \beta_{ij}^*) + \mu_i(\beta_j^* + \beta_{ij}^*, \beta_i^* + \beta_{ji}^*) + \mu_i(\beta_i^* + \beta_{ji}^*, \overline{\beta}) = 0$.

Assume $\beta_j^* + \beta_{ij}^* > \beta_i^* + \beta_{ji}^*$. Benefits for country i are (and similarly for j if $\beta_j^* + \beta_{ij}^* \leq \beta_i^* + \beta_{ji}^*$):

$$\max_{\beta_i^*, T_i^*, \beta_{ij}^*} W_i =$$
$$\int_{\underline{\beta}}^{\beta_i^*+\beta_{ji}^*} \left(\gamma + D - \beta + \beta_{ji}^* - \lambda T_i^*\right)\phi_i(T_j^* - T_i^* + \beta_{ij}^* - \beta_{ji}^*, \beta) + \left(D - (1+\lambda)\beta_{ij}^*\right)\phi_j(T_j^* - T_i^* + \beta_{ij}^* - \beta_{ji}^*, \beta) d\beta +$$
$$\int_{\beta_i^*+\beta_{ji}^*}^{\beta_j^*+\beta_{ij}^*} \left(\gamma + \lambda(\beta_i^* - T_i^*)\right)\phi_i(\beta_i^* - \beta + \beta_{ij}^* + T_j^* - T_i^*, \beta) d\beta + \left(D - (1+\lambda)\beta_{ij}^*\right)\phi_j(\beta_i^* - \beta + \beta_{ij}^* + T_j^* - T_i^*, \beta) d\beta$$
$$\int_{\beta_j^*+\beta_{ij}^*}^{\overline{\beta}} \left(\gamma + \lambda(\beta_i^* - T_i^*)\right)\phi_i(\beta_i^* - \beta_j^* + T_j^* - T_i^*, \beta) d\beta.$$

In this case, firms cut only in j if $\beta_i^* + \beta_{ji}^* \leq \beta \leq \beta_j^* + \beta_{ij}^*$. In this cost range, the welfare effects of the relocation under the foreign cap is defined as $-\Delta_F(\beta) = \gamma + \lambda(\beta_i^* - T_i^*) - D + (1+\lambda)\beta_{ij}^*$, and the cost difference is $C(\beta) = C_i(\beta) - C_j(\beta) = \beta_i^* - \beta + \beta_{ij}^* + T_j^* - T_i^*$. To find a best response to a policy set by country j (β_j^* and β_{ji}^*), we maximize with respect to β_i^*:

$$\begin{aligned}\left(D - (1+\lambda)\beta_i^*\right)\phi_i(C(\beta_i^* + \beta_{ij}^*), \beta_i^* + \beta_{ji}^*) = &\underbrace{\int_{\beta_i^*+\beta_{ji}^*}^{\beta_j^*+\beta_{ij}^*} \Delta_F(\beta)\phi_i'(C(\beta), \beta) - \lambda\phi_i(C(\beta), \beta) d\beta}_{\equiv \mu_i(\beta_i^*+\beta_{ij}^*, \beta_j^*+\beta_{ji}^*)} + \\ &\underbrace{\int_{\beta_j^*+\beta_{ij}^*}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta_j^* + \beta_{ij}^*), \beta) - \lambda\phi_i(C(\beta_j^* + \beta_{ij}^*), \beta) d\beta}_{\equiv \mu_i(\beta_j^*+\beta_{ij}^*, \overline{\beta})}\end{aligned} \tag{A.50}$$

Maximize with respect to β_{ij}^*:

$$\begin{aligned}&\left(D - (1+\lambda)\beta_{ij}^*\right)\phi_j(C(\beta_j^* + \beta_{ij}^*), \beta_j^* + \beta_{ij}^*) \leq \\ &\underbrace{\int_{\underline{\beta}}^{\beta_i^*+\beta_{ji}^*} -\Delta_G(\beta)\phi_j'(C(\beta_i^* + \beta_{ji}^*), \beta) + (1+\lambda)\phi_j(C(\beta_i^* + \beta_{ji}^*), \beta) d\beta}_{\equiv \mu_{ij}(\underline{\beta}, \beta_i^*+\beta_{ji}^*)} + \\ &\underbrace{\int_{\beta_i^*+\beta_{ji}^*}^{\beta_j^*+\beta_{ij}^*} -\Delta_F(\beta)\phi_j'(C(\beta), \beta) + (1+\lambda)\phi_j(C(\beta), \beta) d\beta}_{\equiv \mu_{ij}(\beta_i^*+\beta_{ji}^*, \beta_j^*+\beta_{ij}^*)}\end{aligned} \tag{A.51}$$

which holds with strict inequality if $\beta^*_{ij} = 0$ and strict equality otherwise. The transfer, T^*_i, is optimally set at $\mu_i(\underline{\beta}, \beta^*_i + \beta^*_{ji}) + \mu_i(\beta^*_i + \beta^*_{ji}, \beta^*_j + \beta^*_{ij}) + \mu_i(\beta^*_j + \beta^*_{ij}, \overline{\beta}) = 0$. Note again that the right-hand side does not vanish when β^*_{ij} (due to β^*_j) and it may not be optimal to set the foreign price in a strategic setting between countries.

Characterization of the symmetric equilibrium. Using symmetry we can write $\beta^*_i = \beta^*_j$, $\beta^*_{ij} = \beta^*_{ji}$ and $T^*_i = T^*_j$. Then can write both (A.48) and (A.50) as

$$\begin{aligned}\big(D-(1+\lambda)\beta^*_i\big)\phi_i(C(\beta^*_i+\beta^*_{ij}), \beta^*_i+\beta^*_{ij}) &= \int_{\beta^*_i+\beta^*_{ij}}^{\overline{\beta}} \Delta_H(\beta)\phi_i'(C(\beta^*_i+\beta^*_{ij}),\beta) - \lambda\phi_i(C(\beta^*_i+\beta^*_{ij}),\beta)d\beta \\ &= -\int_{\underline{\beta}}^{\beta^*_i+\beta^*_{ij}} \Delta_G(\beta)\phi_i'(C(\beta^*_i+\beta^*_{ij}),\beta) - \lambda\phi_i(C(\beta^*_i+\beta^*_{ij}),\beta)d\beta\end{aligned} \tag{A.52}$$

Where the last step follows from using the optimal condition for T^*_i: $\mu_i(\underline{\beta}, \beta^*_i+\beta^*_{ij})+\mu_i(\beta^*_i+\beta^*_{ij}, \overline{\beta}) = 0$. Then consider the first-order conditions (A.49) and (A.51) under symmetry:

$$\big(D-(1+\lambda)\beta^*_{ij}\big)\phi_i(C(\beta^*_i+\beta^*_{ij}), \beta^*_i+\beta^*_{ij}) \leq \int_{\underline{\beta}}^{\beta^*_i+\beta^*_{ij}} \Delta_G(\beta)\phi_i'(C(\beta^*_i+\beta^*_{ij}),\beta)+(1+\lambda)\phi_i(C(\beta^*_i+\beta^*_{ij}),\beta)d\beta \tag{A.53}$$

Consider a very small $\beta^*_{ij} \approx 0$. Then the right-hand side becomes $D - \phi_i$ and the left-hand side stays positive; without further assumptions it is not possible to say whether equation (A.53) holds with equality or inequality; that is, whether $\beta^*_{ij} = 0$ or $\beta^*_{ij} > 0$.

Derivation of equations (15) and (16).

Unlike in the main section, the optimal solution no longer takes a bang-bang form (Lemmas A.3 and A.5 no longer hold) when the abatement cost function is not linear. The objective function writes as:

$$\max_{\substack{X_i(\beta),C_i(\beta),\\ X_j(\beta),C_j(\beta)}} W_i = \int_{\underline{\beta}}^{\overline{\beta}} \Big(\gamma + DX_i(\beta) - (1+\lambda)A(X_i(\beta),\beta) + \lambda C_i(\beta)\Big)\phi_i(C(\beta),\beta) + \Big(DX_j(\beta) - (1+\lambda)A(X_j(\beta),\beta) + (1+\lambda)C_j(\beta)\Big)\phi_j(C(\beta),\beta)d\beta$$

so that the voluntary participation constraint $C_j(\beta) \leq 0$ holds, as well as the incentive compatibility, which can be written as $C_k'(\beta) = A_\beta(X_k(\beta),\beta)$, $k = i,j$ (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 neither $X_i(\beta)$ nor $X_j(\beta)$ binds. Hamiltonian for the problem is:

$$\mathcal{H} = \Big(\gamma + DX_i(\beta) - (1+\lambda)A(X_i(\beta),\beta) + \lambda C_i(\beta)\Big)\phi_i(C(\beta),\beta) + \Big(DX_j(\beta) - (1+\lambda)A(X_j(\beta),\beta) + (1+\lambda)C_j(\beta)\Big)\phi_j(C(\beta),\beta) + \nu_i(\beta)A_\beta(X_i(\beta),\beta) + \nu_j(\beta)A_\beta(X_j(\beta),\beta) \tag{A.54}$$

where $\nu_k(\beta)$, $k = i,j$ denotes the co-state variables of the two incentive compatibility constraints. We assume that $X_i(\beta)$ and $X_j(\beta)$ are differentiable. Using Pontryagin's principle, the necessary conditions for the optimum are:

$$\Big(D - (1+\lambda)A_x(X_i(\beta),\beta)\Big)\phi_i(C(\beta),\beta) + \nu_i(\beta)A_{x\beta}(X_i(\beta),\beta) = 0 \tag{A.55}$$

$$\Big(D - (1+\lambda)A_x(X_j(\beta),\beta)\Big)\phi_j(C(\beta),\beta) + \nu_j(\beta)A_{x\beta}(X_j(\beta),\beta) = 0 \tag{A.56}$$

$$-\nu_i'(\beta) = -\Delta(\beta)\phi_i'(C(\beta),\beta) + \lambda\phi_i(C(\beta),\beta) \tag{A.57}$$

$$-\nu_j'(\beta) = -\Delta(\beta)\phi_j'(C(\beta),\beta) + (1+\lambda)\phi_j(C(\beta),\beta) \tag{A.58}$$

$$\nu_i(\underline{\beta}) = 0, \qquad \nu_j(\underline{\beta}) = 0 \tag{A.59}$$

Here $-\Delta(\beta) = \gamma + D(X_i(\beta) - X_j(\beta)) - C_i(\beta) - (1+\lambda)(T_i(\beta) - T_j(\beta))$ denotes the net losses from relocation. A uniform change in costs $C(\beta)$ for all β must balance the welfare effects: $\int_{\underline{\beta}}^{\overline{\beta}} \Delta(\tilde{\beta})\phi_i'(C(\tilde{\beta}),\tilde{\beta}) - \lambda\phi_i(C(\tilde{\beta}),\tilde{\beta})d\tilde{\beta} = 0$. Integrating over (A.57) and (A.58), and fixing the lower bound by using the transversality conditions (A.59), we get:

$$\nu_i(\beta) = \int_{\underline{\beta}}^{\beta} \Big(\Delta(\tilde{\beta})\phi_i'(C(\tilde{\beta}),\tilde{\beta}) - \lambda\phi_i(C(\tilde{\beta}),\tilde{\beta})\Big)d\tilde{\beta} = -\int_{\beta}^{\overline{\beta}} \Big(\Delta(\tilde{\beta})\phi_i'(C(\tilde{\beta}),\tilde{\beta}) - \lambda\phi_i(C(\tilde{\beta}),\tilde{\beta})\Big)d\tilde{\beta} = -\mu_i(\beta,\overline{\beta}) \tag{A.60}$$

$$\nu_j(\beta) = \int_{\underline{\beta}}^{\beta} \Big(\Delta(\tilde{\beta})\phi_j'(C(\tilde{\beta}),\tilde{\beta}) - (1+\lambda)\phi_j(C(\tilde{\beta}),\tilde{\beta})\Big)d\tilde{\beta} \equiv -\mu_j(\underline{\beta},\beta) \tag{A.61}$$

These definitions allow us to write (A.55) and (A.56) as in (15) and (16).

Proof of Proposition 5

Consider partly local damage, that is, a firm that causes damage D at home, but αD if it moves (with $\alpha = [0,1)$). The welfare function in equation (1) changes to:

$$W_i = \int_{\underline{\beta}}^{\overline{\beta}} \Big(\gamma + DX_i(\beta) - C_i(\beta)\Big)\phi_i(C(\beta),\beta) + \Big((1-\alpha)D + \alpha DX_j(\beta)\Big)\phi_j(C(\beta),\beta) - (1+\lambda)T(\beta)d\beta,$$

In contrast to the basic model in Section 3 (represented by $\alpha = 1$), country i gains $(1-\alpha)D$ when a polluting firm moves and keeps polluting abroad ($X_i = 0, X_j = 0$). When a clean firm moves and pollutes abroad ($X_i = 1, X_j = 0$), the gain is: $(1-\alpha)D - D = -\alpha D$. Consider the optimal local mechanism, $M_i(\beta) = \{T_i(\beta), X_i(\beta)\}$. Lemmas A.1-A.4 continue hold, and the optimal policy takes the two-part tariff form as in Lemma 1. The optimal policy finds β_i^* and T_i^* to maximize:

$$\begin{aligned} W_i &= \int_{\underline{\beta}}^{\beta_i^*} \Big(\gamma + \alpha D - \beta - \lambda T_i^*\Big)\phi_i(\beta - T_i^*,\beta)d\beta + \int_{\beta_i^*}^{\overline{\beta}} \Big(\gamma - (1-\alpha)D + \lambda(\beta_i^* - T_i^*)\Big)\phi_i(\beta_i^* - T_i^*,\beta)d\beta \\ &= \int_{\underline{\beta}}^{\beta_i^*} \Big(\gamma' + D - \beta - \lambda T_i^*\Big)\phi_i(\beta - T_i^*,\beta)d\beta + \int_{\beta_i^*}^{\overline{\beta}} \Big(\gamma' + \lambda(\beta_i^* - T_i^*)\Big)\phi_i(\beta_i^* - T_i^*,\beta)d\beta \end{aligned}$$

where we define $\gamma' = \gamma - (1-\alpha)D$. The proof of the proposition now follows directly from equation (A.1) and the steps thereafter.

Contact.

MIT CEEPR Working Paper Series
is published by the MIT Center for Energy and Environmental Policy Research from submissions by affiliated researchers.
For inquiries and/or for permission to reproduce material in this working paper, please contact:

General inquiries: ceepr@mit.edu
Media inquiries: ceepr-media@mit.edu

MIT Center for Energy and
Environmental Policy Research

MIT Center for Energy and
Environmental Policy Research
Massachusetts Institute of Technology
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139-4307
USA

ceepr.mit.edu

MASSACHUSETTS INSTITUTE OF TECHNOLOGY