Adverse Selection as a Policy Instrument: Unraveling Climate Change*

Steve Cicala
University of Chicago

David Hémous
University of Zurich

Morten Olsen
University of Copenhagen

PRELIMINARY AND INCOMPLETE.
PLEASE DO NOT QUOTE OR CITE. COMMENTS WELCOME.

Click here for the latest version.

March 1, 2020

Abstract

This paper applies principles of adverse selection to overcome obstacles that prevent the implementation of Pigouvian policies to internalize externalities. Focusing on negative externalities from production (such as pollution), we evaluate settings in which aggregate emissions are known, but individual contributions are unobserved by the government. The government provides firms with the option to pay a tax on their voluntarily and verifiably disclosed emissions, or an output tax based on the average of rate of emissions among the undisclosed firms. The certification of relatively clean firms raises the output-based tax, setting off a process of unraveling in favor of disclosure. We derive the conditions under which unraveling will yield an outcome close to the first best. We then implement our mechanism in an international setting with unilateral climate change policy as the motivation. We show how such a mechanism extends the reach of a carbon tax, and that the gains over a system of carbon tariffs depend on a small number of estimable parameters.

*We are grateful to Thom Covert, Meredith Fowlie, Michael Greenstone, James Sallee, Joseph Shapiro, Andrei Shleifer, Allison Stashko, Bob Topel and seminar participants at the Utah Winter Business Economics Conference, UC Santa Barbara, UC Berkeley, and the University of Chicago for helpful comments. Cicala gratefully acknowledges funding from the 1896 Fund at the University of Chicago, and along with Olsen thanks the University of Zurich for hospitality. All errors remain our own. e-mail: scicala@gmail.com, david.hemous@gmail.com, graugaardolsen@gmail.com

1 Introduction

Uninternalized externalities abound. In spite of the simplicity of economists' advice when the magnitude of the harm is known, the obstacles to correcting such market failures are myriad: political opposition, excessive implementation costs, the presence of havens induced by competing jurisdictions, among others. In this paper we show the extent to which such obstacles may be overcome in situations in which damage is caused by heterogenous agents. We apply results from the literature on mechanism design under asymmetric information as a policy lever to encourage program participation and voluntary revelation of harm.

We consider situations in which the aggregate level of harm (such as pollution or traffic congestion) is known by the government, but the exact contributions of specific agents is not. In such settings it is impossible to levy Pigouvian taxes due to the unobserved sources of harm. The optimal uniform fee (such as an output tax on producers rather than an emissions tax) falls short of the first best since the fee does not depend on one's contribution to the problem. It also fails to incentivize abatement to reduce damage. We propose creating the *option* to certify one's damage, upon which a Pigouvian tax will be levied, combined with an output-based fee that tracks the average level of damage among those choosing not to participate in the certification program.[1] This encourages those who inflict relatively little damage to certify, thus raising the output-based fee paid by non-participants. This sets off an unraveling in favor of program participation as increasingly damage-intensive agents seek to separate themselves from the tail of the distribution that becomes concentrated by adverse selection.

As a central application we consider the challenge of implementing a policy to mitigate climate change-inducing greenhouse gas emissions. Because greenhouse gases are global pollutants, it is natural that research on climate policy has focused on international environmental agreements between sovereign nations who regulate their respective producers.[2] Such agreements must overcome the unilateral incentive to shirk (Barrett (1994)), possibly by punishing countries outside of the agreement (Nordhaus (2015)). Dynamic considerations also come into play as costly investments in clean technology create hold-up problems in future negotiations due to their complimen-

[1]This can be calculated because the overall level of harm is observed, and subtracting the contribution of certified agents reveals the average contribution among those who remain uncertified.

[2]See Chan et al. (2018) for a review.

tarity with abatement (Beccherle and Tirole (2011); Harstad (2012); Battaglini and Harstad (2016)). Governments are the key decisionmakers in this paradigm, and only policies that are individually rational from each country's perspective are feasible.

The absence of strong, binding international agreements raises the question of how emissions might be reduced through unilateral policy. The ability of unilateral carbon taxation to reduce emissions is reduced (and possibly reversed) when production can profitably move to unregulated jurisdictions, a problem known as 'leakage' (Bohm (1993); Copeland and Taylor (1995); Aldy and Stavins (2012); Hémous (2016); Fowlie (2009); Fowlie et al. (2016)). This prospect strengthens the incentive to shirk on international committments by simultaneously reducing the effectiveness of the tax and increasing the benefits of becoming a haven. Trade policy in the form of a "Border Carbon Adjustment" (BCA) has been considered the primary instrument to mitigate the competitive disadvantage caused by taxing one's own emissions (Copeland (1996); Metcalf and Weisbach (2009); Elliott et al. (2010, 2013); Larch and Wanner (2017), see Condon and Ignaciuk (2013) for a literature review). BCAs levy tariffs based on the average carbon content of production in the country of origin so that foreign producers (on average) cannot undercut domestic firms. Under such a policy foreign producers remain effectively outside the reach of the government, as their tax burden is unrelated to firm-specific emissions. They face no individual incentives to abate their emissions, and any pollution reductions depend on price elasticities of demand and supply.[3]

The goal of our approach is to approximate the emissions reductions that might be achieved with a widely-adopted price on carbon, but without requiring the legally-binding international agreements that have proven elusive to date. To do so, we focus on the direct interactions between a government and firms whose disclosure of emissions is voluntary. International sovereignty may restrict what governments can mandate of foreign firms, but does not foreclose the possibility of creating incentives to shape their behavior. We do this by providing firms with the option to certify their emissions, and basing the default rate on the average emissions of uncertified firms. This recasts the problem of jurisdiction into one of screening, in which clean firms

[3]Markusen (1975) and Hoel (1996) derive the optimal tariff in the presence of transboundary pollution. As highlighted by Keen and Kotsogiannis (2014) and Balistreri, Kaffine and Yonezawa (2019), an optimal environmental tariff generally differs from the BCA formula (even if a BCA were able to distinguish between the carbon contents of different imports and even in the absence of terms of trade effects).

wish to separate themselves from more intensive polluters (Spence (1973); Stiglitz (1975)). This separation causes the uncertified mean to rise, setting off a process of unraveling that encourages further certification (Akerlof (1970)).

The combination of optional disclosure and a rolling default creates a policy that mimics the strategy that has been applied in private markets to ensure quality (Jovanovic (1982); Grossman (1981); Milgrom and Roberts (1986); Milgrom (2008). See Dranove and Jin (2010) for a review). In these settings firms voluntarily provide warranties or submit to audits in order to separate themselves from low-quality producers. Even relatively lower-quality firms become willing to make such disclosures to separate themselves from the absolute worst offenders when consumers update their beliefs regarding those who decline to disclose (Jin and Leslie (2003); Jin (2005); Lewis (2011)). It has also been used by firms to improve risk selection for credit (Einav et al. (2012)), improve safety (Viscusi (1978); Hubbard (2000); Jin and Vasserman (2019)), and has been suggested to encourage more efficient electricity consumption (Borenstein (2005, 2013)). To our knowledge this is the first paper to apply these principles to overcome obstacles to the implementation of Pigouvian policies.

The use of screening mechanisms in public policy has been successfully applied to improve the targeting of recipients of public benefits (Alatas et al. (2016); Finkelstein and Notowidigdo (2019); Deshpande and Li (2019)). In such settings the government creates hurdles so uptake is limited to those who value benefits more than the ordeal of enrollment. These policies typically do not entail unraveling as the government is is free to choose the magnitude of the enrollment ordeal so that the optimal point of separation is achieved immediately (Kleven and Kopczuk (2011); Besley and Coate (1992); Nichols and Zeckhauser (1982); Nichols et al. (1971)): The costs or benefits of non-participation are not programmatically adjusted with the extent of participation. In our setting this would be analogous to the government choosing its preferred carbon content of uncertified imports, which is likely to run afoul of strategic trade considerations.[4]

There is a long tradition of regulation under asymmetric information in the mech-

[4]At the extreme, the government could simply prohibit imports from firms whose emissions are uncertified. This is not without precedent—the U.S. Food and Drug Administration mandates access for inspectors at facilities abroad for any firm wishing to sell food or phamaceuticals in the US (Federal Food, Drug and Cosmetic Act, Section 807 (b) as amended by Section 306 of the FDA Food Safety Modernization Act). The basis of the jurisdiction problem we address is that such mandates are infeasible with respect to emissions.

anism design literature (Baron and Myerson (1982); Laffont and Tirole (1993)). In the pollution context, the regulator seeks to elicit information on abatement costs (Kwerel (1977); Roberts and Spence (1976); Dasgupta et al. (1980); Baron (1985); Laffont (1994)) and must design a policy schedule that elicits truthful revelation. In these settings, as in the context of non-point source pollution, the lack of verifiability is the key constraint on the regulator (Segerson (1988); Xepapadeas (1991); Laffont (1994); Xepapadeas (1995), among others. For a review see Xepapadeas (2011)). While emissions remain unobserved at uncertified firms, our focus on an optional, *verifiable* revelation of emissions converts the problem into a traditional point-source setting in which firms face incentives to abate. Recent work on voluntary environmental regulation notes the improved enforcement targeting for uncertified firms, but does not unravel non-participation with changing in audit probabilities (Foster and Gutierrez (2013, 2016)).

The paper proceeds as follows: We first develop a closed-economy model in which production is heterogenously associated with an externality and we derive the distance of an optimally-set output tax from the first-best Pigouvian policy. We then show how the option to reveal one's emissions yields welfare objects that are a linear combination of the outcomes under output and emissions taxes, depending on the share of the industry that certifies. These expressions depend in a simple manner on rate of certification and the variance of the emissions. In the third section, we show that, under certain conditions, the policy maker can achieve the same outcome by only knowing the mean of the emissions distribution. This is achieved through an algorithm that encourages the gradual unravelling of the emissions distribution converging to an equilibrium in which the policy maker has full information on the emissions distribution. We extend the analysis to allow firms to abate and show that there is a natural complementarity between the two; only through certification is it worthwhile for firms to abate. We then extend our model to an international setting, focusing on the constraints of unilateral climate policy and show the conditions under which the unravelling mechanism is preferable to a unilateral tax or a tax combined with a BCA.

2 Unraveling in the Domestic Case

We first consider a closed economy extend the model to an international context in Section 3 below. We start in a simplified setting in which firms differ in their emissions rates and the government knows the full distribution of emissions, though initially not those of an individual firm. We characterize the benefits to welfare from an optional certification program and write the benefits as an expression of the variance of emissions and supply elasticities. We then show when the same equilibrium can be achieved even when the government only knows the overall average emissions rate. Following this, we allow for abatement through which individual firms can reduce their emission rates and show that optional certification can further increase the benefits of abatement.

Throughout this section we focus on a simple partial equilibrium model with an externality, but the concept should be thought of as broader.

2.1 Baseline model

We consider a closed economy and focus on a specific industry which produces a homogenous good under perfect competition. A representative agent has preferences over this good, represented by the following quasi-linear utility function:

$$U = C_0 + u(C) - vG,$$

where C is total consumption of the homogenous good and C_0 is the consumption of an outside good. Central to the analysis that follows are the emissions resulting from the production of the good C, which we denote by G. The marginal social cost of emissions is v. The production of the outside good does not require pollution.

The polluting good is produced by an (exogenous) mass 1 of firms who operate under perfect competition. Each firm has the same strictly convex cost function $c(q)$, but differs in the extent to which they pollute. The emissions rate per unit produced is denoted e and follows the cdf $\Psi(e)$ on the domain $[0, \bar{e}]$ where $\bar{e} < \infty$. Though the overall distribution of emissions, Ψ, and the production of each firm is observable, the emissions of an individual firm are private information (unless the firm is certified as described below). Below, we discuss how a certification program can be implemented without the full information on $\Psi(e)$. Let $q(i)$ be the production of firm i and $e(i)$ be the emissions of firm i per unit produced. Then total production is given by:

$$Q = \int_0^1 q(i)di, \tag{1}$$

and total emissions are given by

$$G = \int_0^1 e(i)q(i)di \tag{2}$$

2.2 Equilibrium with an output tax or emission tax

In the following we introduce emissions and output taxes. Throughout, we keep the number of potential firms exogenous and assume that the least productive firm stays active.

It is possible to tax output, while the emissions of individual firms is unobserved. We label this tax t and solve for the decentralized equilibrium. Let the market price be given by p. Profit maximization by individual firms then results in:

$$p = c'(q) + t, \tag{3}$$

which defines a supply function $s(p - t)$. The resulting profit function follows as $\pi(p - t) = (p - t)s(p - t) - c(s(p - t))$. The supply curve is upward-sloping by the convexity of the cost function and the profit function is increasing in its argument by Hotelling's lemma.

Utility maximization gives:

$$u'(C) = p, \tag{4}$$

which together with equation (1) and $C = Q$ defines an equilibrium price, p, and quantity, C. All firms produce the same quantity, so resulting emissions are given by:

$$G = s(p - t) \int_0^{\bar{e}} e\Psi(e)de = s(p - t)E(e) \tag{5}$$

To facilitate the discussion of certification we also solve for a decentralized equilibrium in which emissions are observable and taxed at τ. Profit maximization then results in an individual supply function of $s(p - \tau e)$ such that aggregate supply and emissions are given by:

$$Q = \int_0^{\bar{e}} s(p - \tau e)\Psi(e)de = E\left[s(p - \tau e)\right], \quad G = E\left[es(p - \tau e)\right] \tag{6}$$

These equilibria can be compared with the social planner's problem when individual emissions are unobserved :

$$max_q u(q) - c(q) - vqE(e),$$

which has a first order condition of:

$$u'(q) = c'(q) - vE(e), \tag{7}$$

such that the social planner's solution can be implemented with a quantity tax of $t = vE(e)$. Analogously, a social planner who has information on the emissions of individual firms would implement the social optimum by imposing an emission tax of $\tau = v$.

In the following we solve for an equilibrium in which firms can choose to certify to be taxed at a level tailored to their individual emissions.

2.3 Equilibrium with certification

We now introduce the *voluntary* certification of emissions, in which a firm can choose between two tax schemes. If the firm chooses to (verifiably) reveal its level of emissions, e, it is taxed at τe (where we do not necessarily impose that $\tau = v$). If, on the other hand the firm chooses not to reveal its level of emissions it is taxed at the mean level of emissions of the firms who do not certify $t = \tau E(e|R)$, where R denotes the set of firms who have not certified. Presenting this choice to firms requires the government to be able to calculate R *ex ante* based on knowledge of the distribution of emissions, $\Psi(e)$. Subsection 2.5 shows the conditions under which this equilibrium can be achieved without complete information on $\Psi(e)$. The total cost to a firm of certification equals the technical cost of certification, in the form of a third-party expert, an objective monitoring system etc., $F > 0$ and a potential additional tax that the government might impose, $f \lesseqgtr 0$. In an equilibrium in which some firms certify and others do not, an indifferent firm with emissions level $\hat{e}$ is defined by:

$$\pi(p - \tau\hat{e}) - (F + f) = \pi(p - t). \tag{8}$$

Since the left hand side is decreasing in e, all firms with $e < \hat{e}$ certify and firms with $e > \hat{e}$ do not. The resulting tax rate per unit output for firms who do not certify is then:

$$t = \tau E[e|e > \widehat{e}].$$

To facilitate the discussion below, we introduce ε, which is equal to the emissions rate at which a firm is effectively taxed:

$$\varepsilon = \begin{cases} e & \text{if} \quad e < \hat{e} \\ E(e|e > \hat{e}) & \text{if} \quad e \geq \hat{e} \end{cases} \tag{9}$$

Production by firms who do not certify is then $s(p - \tau E(e|e > \hat{e}))$ and for those who do certify it is $s(p - e\tau)$ such that total production is:

$$Q = E(s(p - \tau\varepsilon)),$$

with corresponding emissions of:

$$G = E\left[\varepsilon s(p - \tau\varepsilon)\right]. \tag{10}$$

The equilibrium price then follows from market clearing ($C = Q$) and utility maximization $u'(C) = p$. In Appendix XXX, we derive that a sufficient condition for this equilibrium to be unique is that $E[e|e > \hat{e}] - \hat{e}$ is decreasing in $\hat{e}$ and i) $s(\cdot)$ is weakly convex or ii) τ is small. Labelling G^V total emissions under the scheme of "voluntary" certification (equation (10)) and G^U total emissions without certification, G^U (equation 5) we establish the following lemma

Lemma 1. *The difference between emissions under certification and without certification is given by:*

$$G^V - G^U = Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right] + E\left(e\right)\left\{E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E\left(e\right)\right)\right\},$$

where p^V and p^U are the equilibrium prices under certification and no certification, respectively. The effect of certification on emissions is generally ambiguous. However, emissions decline when s is weakly convex and $es(p^V - \tau e)$ is concave in e. This is satisfied for linear supply curves.

Lemma 1 is derived for a given tax on certification, f, not necessarily the optimal one. We address that issue in Section 2.4 below. Emissions do not necessarily decline since production potentially increases. This is essentially a "rebound" effect: the emissions from certified firms increases when they are taxed at a lower rate, and it is possible that this increase is greater than the corresponding decrease from higher taxes on uncertified firms.

Lemma 1 states that emissions necessarily decline for linear supply curves: Certification lowers the tax rate for some firms and raises it for others, keeping the average tax rate constant at $\tau E(\varepsilon) = \tau E(e)$. This being the case, total production and the market clearing price is unaffected by certification when supply curves are linear. With no change in aggregate production, but a reallocation towards less polluting firms, total emissions must decline.[5]

Analogously to Lemma 1 we label W^V welfare for the equilibrium with certification and derive the difference in welfare in the following proposition.

Proposition 1. *The difference between social welfare for voluntary certification, denoted "V", and without, denoted "U" is:*

$$\begin{aligned} W^V - W^U \;\; = \;\; & \underbrace{\int_{p^U}^{p^V} \left(s\left(p - \tau E\left(e\right)\right) - D\left(p\right)\right) dp}_{\text{price effect}} + \underbrace{E\left(\pi\left(p^V - \tau\varepsilon\right)\right) - \pi\left(p^V - \tau E\left(e\right)\right)}_{\text{reallocation effect}} \\ & \underbrace{- \left(v - \tau\right)\left(G^V - G^U\right)}_{\text{untaxed emissions effect}} - F\Psi\left(\widehat{e}\right), \end{aligned} \tag{11}$$

where p^V and p^U are the equilibrium prices under certification and no certification, respectively.

a) The "price effect" and the "reallocation effect" are always weakly positive.

b) If supply functions are linear and all firms remain in operation under certification, prices are the same for certification and no certification, $p^V = p^U$.

c) The "untaxed emissions effect" is generally of ambiguous sign, but i) if $v = \tau$ it is zero, ii) if s is weakly convex and $es(p^V - \tau e)$ concave it is positive for $v > \tau$.

Proof. Proof in Appendix □

Proposition 1 establishes the change in social welfare arising from a shift from an equilibrium without certification to an equilibrium with certification for given tax rate, τ. To see the intuition, first focus on the first two terms, the price and reallocation

[5] Alternatively, consider a convex supply function, which implies that for a given price total supply must increase with certification, so that the equilibrium price declines ($p^V < p^U$). As a result, $es(p^V - \tau e) < es(p^U - \tau e)$. In addition, when $es(p^V - \tau e)$ is concave in e an application of Jensen's inequality ensures that overall emissions decline.

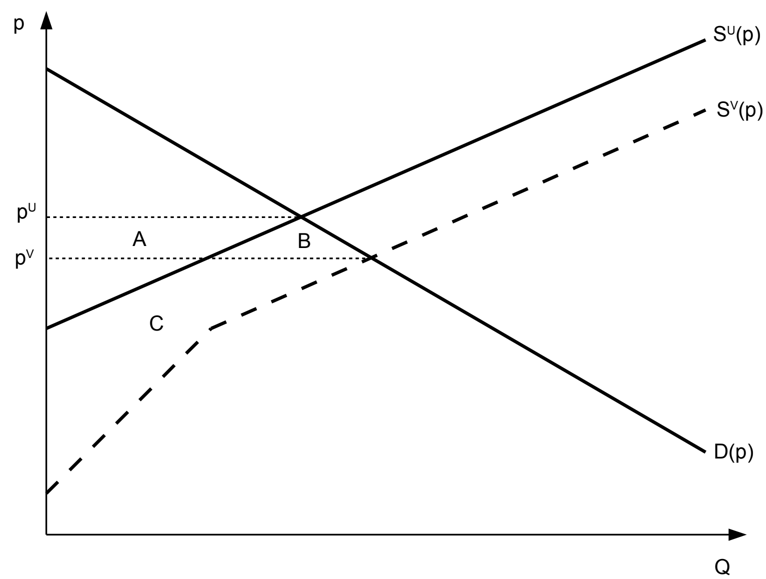
effect. A shift from an equilibrium without certification to one with certification implies a shift of the cost functions of individual firms as they face different taxes. This shifts individual supply functions and consequently the aggregate supply function, denoted by S. Figure 1 denotes $S^U(p)$ as the aggregate supply curve for the uncertified equilibrium and $S^V(p)$ for the certified. $S^V(p)$ always intersects the y-axis lower than the $S^U(p)$ curve because the certified firms with the lowest emission rate are taxed at a lower rate and are consequently willing to supply at a lower price. For sufficiently low prices, the number of firms producing is increasing along $S^V(p)$ and our assumption that all firms produce therefore requires an equilibrium price sufficiently high. $S^V(p)$ need not be strictly to the right of $S^U(p)$ at all points and hence the price need not decline. The figure illustrates the price and reallocation effect in the case in which the $S^V(p)$ is to the right of $S^U(p)$ and the price declines from p^U to p^V: a shift from p^U to p^V changes producer and consumer surplus, but always in a manner that leaves aggregate surplus (weakly) higher (in this case by a net increase of the area B).[6] This is the *price effect* which is consequently always weakly positive. In addition, the *reallocation effect* captures the increase in profits from a reallocation of production from firms with higher taxes to those with lower and is captured by the area C in the figure and the second term in equation (11). With ε being the effective taxed emission level, $E[\pi(p^V - \tau\varepsilon)]$ is the average profits for firms in the equilibrium with voluntary certification and $\pi(p^V - \tau E(e))$ are the profits when firms are all taxed at the same level, $E(e)$. This effect is always positive.

Next, consider the *untaxed emissions effect* which captures the welfare effects of changes to emissions. First, note changes to emissions have no welfare impact when emissions are taxed at the socially efficient level ($v = \tau$). If taxes are lower than this, then any increase in emissions have a negative impact. Consequently, the sign of this term inherits the properties of Lemma 1, and is positive if s is weakly convex and $es(p^V - \tau e)$ is concave in e. Finally, the last term of equation (11) captures the fraction $\Psi(\hat{e})$ of firms which each spend F to certify.

In anticipation of the analysis to come we consider the special case of linear supply curves (quadratic cost curves) and with little loss in generality we let the supply curve be $s(p) = \tilde{s}p$, $\tilde{s} > 0$. From the analysis above this implies that the price effect disappears and that the reallocation and untaxed emission effects are both positive.

[6]Appendix XXX illustrates the opposite case in which the price increases. Proposition 1 still holds but the allocation of welfare between producers and consumers is different.

Figure 1: Market Equilibria with and without Voluntary Certification



Do note that the entire benefit to certification accrues to the firms (through the reallocation effect), when the tax rate is Pigouvian. Corollary 1 gives the result:

Corollary 1. *For linear supply functions of the form $\tilde{s}p$, the expression $W^V - W^U$ in Proposition 1 reduces to:*

$$W^V - W^U = \left(v - \frac{\tau}{2}\right)\tilde{s}\tau Var\left(\varepsilon\right) - F\Psi\left(\widehat{e}\right),$$

where $W^V - W^U + F\Psi(\hat{e})$ is positive if $\tau < 2v$.

Recall that for linear supply curves the price doesn't change; that is the supply curves $S^V(p)$ and $S^U(p)$ intersect the demand curve at the same point. Consequently, the price effect is zero, whereas the reallocation effect remains positive. Further, with constant production, but a shift towards firms with fewer emissions total emissions are sure to decline. Even if emissions are already taxed efficiently, $\tau = v$, total welfare increases because production is reallocated towards less polluting firms. The size of this reallocation depends on the supply response, $\tilde{s}$, and the variance of the taxed emissions rate, $Var(\varepsilon)$.

In Corollary 2, we assume that taxes, τ, and the social costs of emissions, v, are sufficiently small relative to prices that we can conduct Taylor-expansions around $\tau = 0$, $v = 0$ (we label the price for a tax of $\tau = 0$ as p_0). Focusing on small taxes allows us to focus on the central welfare effects of Proposition 1. For small taxes, supply curves are approximately linear and building on Corollary 1 we obtain the following result:

Corollary 2. *To a first-order approximation, the expression $W^V - W^U$ in Proposition 1 can be written as:*

$$W^V - W^U = \left(v - \frac{\tau}{2}\right)s'\left(p_0\right)Var\left(\varepsilon\right)\tau - F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \qquad (12)$$

with a difference in emissions of:

$$G^V - G^U = -s'\left(p_0\right)\tau Var\left(\varepsilon\right) + o(\tau),$$

which is negative (to a first order).

As Corollary 2 makes clear, to a first order we replicate an expression as for linear supply curves and the intuition is analogous: there is no price effect and the primary driver of the welfare consequences of certification come from the shift in production from more to less polluting firms.

A natural question is how far welfare of the voluntary certification in Proposition 1 is to the welfare that would be obtained if firm emission rates were known (without necessarily imposing $\tau = v$). Labeling such an equilibrium with W^{FI} for "full information" we find (at first order)

$$W^V = \frac{Var\left(\varepsilon\right)}{Var\left(e\right)} W^{FI} + W^U \left(1 - \frac{Var\left(\varepsilon\right)}{Var\left(e\right)}\right) - F\Psi\left(\hat{e}\right) + o(\tau^2). \tag{13}$$

By construction $Var(\varepsilon) \leq Var(e)$ so welfare under voluntary certification, W^V, is a weighted average of welfare with no certification, W^U, and with full information, W^{FI}, where the weight depends on the relative variance of the effectively-taxed emission rate, ε, and the actual emission rates, e. Complete certification yields the first best allocation of production at the cost of implementation.

The expression in Proposition 1 and its two Corollaries is derived for any tax/subsidy to the certification scheme, under full knowledge of the emission distribution $\Psi(e)$ and without letting firms reduce their emissions through abatement. In the subsequent section we derive the optimal level of certification. Thereafter, we show that even if the policy maker and producers do not not know the full distribution of $\Psi(e)$ but only its first moment, the policy maker might still be able to implement the certification equilibrium through an "unravelling algorithm." Finally, we derive Proposition 2 where firms are allowed to invest in abatement to reduce their emissions.

2.4 Optimal certification

In the analysis above, individual firms choose whether to certify emissions given total costs $F + f$ of certification, where f is a potential tax on certification (see equation 8). A natural question to ask is whether the social planner ought to encourage or discourage the certification of individual firms. We answer this question by allowing the social planner to set τ and f optimally.[7]

$$max_{\tau,f} W^V$$

[7] This is also the solution to a social planner's problem when they can choose any allocation subject to the constraint that all uncertified firms are indistinguishable.

In an Appendix we show that this requires:

$$\tau = v,$$

$$\pi\left(p - v\widehat{e}\right) - \pi\left(p - vE\left(e|e > \widehat{e}\right)\right) - vE\left(e - \widehat{e}|e > \widehat{e}\right)s\left(p - vE\left(e|e > \widehat{e}\right)\right) - F = 0. \quad (14)$$

The condition $\tau = v$ recovers a standard Pigouvian result: emissions ought to be taxed at their social cost. Subtracting equation (8) from (14) returns:

$$f = vE\left(e - \widehat{e}|e > \widehat{e}\right)s\left(p - vE\left(e|e > \widehat{e}\right)\right) > 0, \quad (15)$$

that is: certification should be taxed when emissions are taxed at the Pigouvian level. The intuition for this comes from Proposition 1 where $\tau = v$ implies that there is no untaxed emissions effect. Optimal certification is defined as the level of certification that maximizes $W^V - W^U$. The subsidy itself is a transfer so welfare is affected only through the effect on price, p^V, the level of certification, $\hat{e}$, and the cost of certification, F. Consider first the effect of a change in price. For a given tax system of the government this essentially boils down to a transfer between producers and consumers and through an envelope-style argument (see the Appendix for details) the combined effect is zero. Second, consider the change in the level of certification. The price effect is not directly affected by a change in certifcation, and the effect on the reallocation effect - capturing profits of all producers is:

$$\left[\pi\left(p - v\widehat{e}\right) - \pi\left(p - vE\left(e|e > \widehat{e}\right)\right) - vE\left(e - \widehat{e}|e > \widehat{e}\right)s\left(p - vE\left(e|e > \widehat{e}\right)\right)\right]\psi(\hat{e})d\hat{e}.$$

The first two terms represent the change in profits for the mass of firms, $\psi(\hat{e})d\hat{e}$ who certify as a consequence of an increase in subsidy. The remaining term captures the fact that taxes go up for the firms who have yet to certify. Realizing that the certifying firms only consider the first two terms and taking into account the physical cost of certification returns equation (14). As a consequence, firms have a higher incentive to certify and will do so too much in an equilibrium without a tax on certification.

Having solved for the decentralized equilibrium and the social planner's allocation we take a step back and assess the informational requirements needed to implement such a policy. Whereas equation (12) gives an intuitive result of the welfare gains based on measures that are relatively easily obtained such as the variance of emission rates and supply elasticities, the optimal implementation requires complete information on the distribution of e, which is rarely available.

In the following, we show the conditions under which a given "algorithm" can achieve a comparable goal without complete information on the distribution of e.

2.5 An "Unravelling" Algorithm

We assume that neither firms nor the government know the distribution of emissions rates, but they do observe the average emissions rate (through aggregated accounts or changes in ambient polllution, for example). Initially, certification is not available and the government imposes an output tax $t_0 = \tau E(e)$. We assume that the government introduces certification which allows firms to pay the emission tax τ at some certification cost F. Since the government does not know the distribution $\Psi(e)$, it cannot predict the eventual threshold $\widehat{e}$ and therefore cannot implement the equilibrium described above by immediately announcing a new output tax $\tau E\left(e|e > \widehat{e}\right)$.

Instead, we show that an iterative process can asymptotically achieve the same equilibrium. We consider such a process where the government allows firms to certify at increasing levels of emissions rates (i.e. the government asks if firms with emissions rate e want to certify and only those firms are allowed to do so. If they do, the level of certification increases and the procedure starts again). The government constantly adjusts the output tax as the emissions distribution gets revealed. We show that we obtain a Nash equilibrium when firms decide on certification as if they were the last ones to certify with the information available at that point in time.

Assume that all firms with an emissions rate below $\widetilde{e}$ have certified and consider the situation of a firm with emissions rate $\widetilde{e}$. The firm will then choose to certify if:

$$g\left(\widetilde{e}\right) \equiv \pi\left(p\left(\widetilde{e}\right) - \tau\widetilde{e}\right) - \pi\left(p\left(\widetilde{e}\right) - \tau E\left(e|e > \widetilde{e}\right)\right) - F \geq 0,$$

where $p\left(\widetilde{e}\right)$ is the price that would prevail on markets should the certification stop here and the threshold be $\widehat{e} = \widetilde{e}$. Since the distribution up to $\widetilde{e}$ has been revealed publicly and since $E\left(e\right)$ is known, both the government and the firm can compute $E\left(e|e > \widetilde{e}\right)$. Assuming that $g\left(\underline{e}\right) > 0$ (where $\underline{e}$ is the lowest emissions rate), then at least some firms will certify. Furthermore, $g\left(\overline{e}\right) = -F$ so not all firms will certify as long as $F > 0$. As firms decide sequentially to certify, the process will continue up to the smallest emissions rate for which g switches sign, which we denote $\widehat{e}$ (which also corresponds to a laissez-faire equilibrium when the government knows the full distribution ψ).

This leads to a Nash equilibrium because for firms with a lower emissions rate, $\widetilde{e} < \widehat{e}$,

$$\pi\left(p\left(\widehat{e}\right) - \tau\widetilde{e}\right) - \pi\left(p\left(\widehat{e}\right) - \tau E\left(e|e > \widehat{e}\right)\right) - F > g\left(\widehat{e}\right) = 0,$$

so that none of these firms would benefit from deviating from certification at the equilibrium. Since g becomes negative just after $\widehat{e}$, firms with emission rate $\widehat{e} + de$, decide not to certify and would indeed be worse off otherwise. Firms with a higher emission rate $\widetilde{e} > \widehat{e}$, will not certify either. Consequently, even if the government does not know the distribution of $\Psi(e)$ it can implement an algorithm where the optimal certification decision of firms gradually reveals the shape of the distribution up until $\hat{e}$.

The analysis can be straightforwardly extended to the social optimum if the government constantly adjust the certification tax $f = \tau\left(E\left(e|e > \widetilde{e}\right) - \widetilde{e}\right) s\left(p(\widetilde{e}) - \tau E\left(e|e > \widetilde{e}\right)\right)$ with the information available.

2.6 Abatement

The analysis thus far has focused on the role of certification in reallocating production from more polluting to less polluting firms. In the following we analyze its role in increasing the incentive for individual firms to reduce emissions by allowing for abatement. We keep the same structure but allow firms to spend $b(a)$ per unit produced to reduce their emissions (per unit) by a. We require: $b'(a) > 0$ for $a > 0$ with $b(0) = b'(0) = 0$ and $b''(a) > 0$. Hence costs of abatement are proportional to production. The expression for emissions, equation (2), is then replaced by:

$$G = \int_0^1 \left(e(i) - a(i)\right) q(i) di.$$

We continue to let (pre-abatement) emissions be distributed according to $\Psi(e)$, but firms who are certified pay an emission tax on $e(i) - a(i)$ instead of $e(i)$. We continue to define ε as in equation (9), which is the pre-abatement emissions rate for certified firms, and the conditional mean of emissions for uncertified firms. Abatement investments are not observable and non-certified firms consequently have no economic incentive to abate. Hence, in an equilibrium without certification no abatement takes place. Certified firms, in contrast, do abate. They solve the problem:

$$max_{q,a} pq - c(q) - \tau(e - a)q - b(a)q$$

which leads to a common abatement level, a^*, amongst all firms that certify of:[8]

$$a^* = b'^{-1}(\tau)$$

Individual supply functions take the form:

$$q(i) = s\left(p - \tau(e(i) - a^*) - b(a^*)\right),$$

where for given price and tax, supply is higher under abatement ($\tau a^* > b(a^*)$). Using this we can derive the changes to emissions from certification in a setting in which abatement is possible. Lemma 2 gives the result:

Lemma 2. *When abatement is possible, the difference between emissions under voluntary certification and without certification is given by:*

$$G^V - G^U = Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right] + E\left(e\right)\left\{E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E\left(e\right)\right)\right\} \quad (16)$$

$$-a^*\Psi(\hat{e})E\left[s\left(p^V - \tau\left(e - a^*\right) - b\left(a^*\right)\right)|e < \hat{e}\right]$$
$$+\Psi(\hat{e})E\left\{e\left[s\left(p^V - \tau\left(e - a^*\right) - b\left(a^*\right)\right) - s\left(p^V - \tau e\right)\right]|e < \hat{e}\right\}$$

where p^V and p^U are the equilibrium prices under certification and no certification, respectively. The effect of certification on emissions is generally ambiguous, but if s is weakly convex and $es(p^V - \tau e)$ is increasing and weakly concave in e, then emissions must decrease following certification.

Lemma 2 mirrors Lemma 1 but adds two additional terms. The direct impact of $\Psi(\hat{e})$ firms certifying is that they abate their emissions by a^*. At the same time certification lowers their tax burden, which yields a supply response analogous to a rebound effect on the quantity produced. This second effect pulls in the direction of higher emissions.

Using Lemma 2 we derive Proposition 2 which establishes the change in social welfare from the uncertified equilibrium (where no abatement takes place) and an

[8] When $e(i) = 0$, this implies firms sequester emissions at rate a^*. In the absence of sequestration, one can think of $e(i) = 0$ as a normalization of the minimum emissions rate that is otherwise bounded away from zero even in the presence of abatement.

equilibrium with voluntary certification and the possibility to abate.

Proposition 2. *For the equilibrium with abatement as described above, the difference in social welfare between the equilibrium with voluntary certification and the one without is given by:*

$$\begin{aligned} W^V - W^U &= \underbrace{\int_{p^U}^{p^V} \left(s\left(p - \tau E\left(e\right)\right) - D\left(p\right)\right) dp}_{\textit{price effects}} \qquad (17) \\ &+ \underbrace{E\left(\pi\left(p^V - \tau\varepsilon\right)\right) - \pi\left(p^V - \tau E\left(e\right)\right)}_{\textit{reallocation gains}} \\ &+ \underbrace{\Psi\left(\widehat{e}\right)\left(E\left(\pi\left(p^V - \tau\left(e - a^*\right) - b\left(a^*\right)\right) - \pi\left(p^V - \tau e\right) \middle| e < \widehat{e}\right)\right)}_{\textit{abatement gains}} \\ & \underbrace{- \left(v - \tau\right)\left(G^V - G^U\right)}_{\textit{untaxed emissions}} - F\Psi\left(\widehat{e}\right), \end{aligned}$$

where

- a) The price effect, the reallocation gains, and the abatement gains are always positive

- b) The untaxed emissions effect is ambiguous. If s is convex and $es(p^V - \tau e)$ is increasing and weakly concave in e then emissions decline, and for $\tau < v$ the untaxed emissions effect is positive.

Proposition 2 carries over some of the results of Proposition 1 to the case of abatement. It includes a new term from *abatement gains*, which captures the increase in profits from firms who get an effective price increase of $ta^* - b(a^*) > 0$ from abatement. For given price p^V, profit maximization ensures that this term is positive. The untaxed emissions effect follows the logic of Lemma 1 and a decrease in emissions and $v > \tau$ is enough to ensure that the untaxed emissions effect is positive.

We complete the analysis of the setting with abatement by deriving two corollaries of Proposition 2. First, for the special case of linear supply curves the expression simplifies to:

Corollary 3. *For linear supply curves of the form $\tilde{s}p$, the expression $W^V - W^U$ in Proposition 2 simplifies to:*

$$W^V - W^U = \underbrace{\int_{p^U}^{p^V} \left(\tilde{s}\left(p - \tau E\left(e\right)\right) - D\left(p\right)\right) dp}_{\textit{price effect}} + \underbrace{\frac{\tilde{s}\tau^2}{2} Var\left(\varepsilon\right)}_{\textit{reallocation}}$$

$$\underbrace{+\Psi\left(\widehat{e}\right)\tilde{s}\left(\tau a^* - b\left(a^*\right)\right)\left(\left(p^V - \tau E\left(e|e<\widehat{e}\right)\right) + \frac{\left(\tau a^* - b\left(a^*\right)\right)}{2}\right)}_{\textit{abatement gains}}$$

$$\underbrace{-\left(v-\tau\right)\left(G^V - G^U\right)}_{\textit{untaxed emissions}} - F\Psi\left(\widehat{e}\right),$$

where the price effect, the reallocation effect and the abatement gains are all positive and a sufficient condition for $W^V - W^U + F\Psi(\hat{e})$ to be positive is that $p^V > 2\tau E(e|e < \hat{e})$.

Recall that for linear supply curves and without abatement the price remains constant and there is no price effect (Corollary 1). With abatement, firms reduce their taxes, produce more and consequently the equilibrium price declines and the price effect is positive as described in Corollary 2. The reallocation effect continues to have the same interpretation. The expression for the abatement gains, which captures savings to firms from abatement, are best seen by noting that savings per unit is $\tau a^* - b(a^*)$ and average number of units sold for certified firms before abatement is $\tilde{s}(p^V - \tau E(e|e < \hat{e}))$. The second term in the parenthesis captures the fact that with abatement supply increases. Finally, changes in emissions repeat the expression from Proposition 1, but gives an explicit sufficient condition for emissions to decline. If $p^V > 2\tau E(E|e < \hat{e})$, which is true for τ small, the *untaxed emission effect* is positive.

Finally, we continue our analysis at first order and derive Corollary 4:

Corollary 4. *To a first order (in (τ, v)) the welfare effect in equation (17) can be written as*

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right)\left(s'(p_0)Var(\varepsilon) + \frac{s(p_0)}{b''(0)}\Psi\left(\widehat{e}\right)\right) - F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right),$$

where $W^V - W^U + F\Psi(\hat{e})$ is positive.

Corollary 4 takes advantage of the fact that to a first order $\hat{e}$ remains unchanged with abatement and consequently we can add a single term to equation (12) from Corollary 2: $b''(0)$ captures the curvature in the abatement costs. By assumption it is always profitable to do some abatement and a low curvature of the abatement function, a low $b''(0)$, implies that to a first order the optimal abatement level will be higher. Abatement benefits are proportional to total production by the firms that abate: $s(p_0)\,\Psi(\hat{e})$. This effect requires taxes not to be higher than $2v$, since otherwise firms will abate "too much." Analogously to equation (13) we can write welfare under certification with abatement as a weighted average of the welfares without certification, W^U and with full information under abatement, W^{FIA}:

$$W^V = \frac{Var(\varepsilon) + \frac{s(p_0)}{s'(p_0)b''(0)}\Psi(\widehat{e})}{Var(e) + \frac{s(p_0)}{s'(p_0)b''(0)}} W^{FIA} + \left(1 - \frac{Var(\varepsilon) + \frac{s(p_0)}{s'(p_0)b''(0)}\Psi(\widehat{e})}{Var(e) + \frac{s(p_0)}{s'(p_0)b''(0)}}\right) W^U - F\Psi(\hat{e}) + o\left(\tau^2\right).$$

Welfare under certification continues to be a weighted average average of the welfare with complete information and with no information, but with the added term, $\frac{s(p_0)}{s'(p_0)b''(0)}\Psi(\hat{e})$ which captures (to a first order) the amount of abatement done with certification.

Finally, it is worth noting the possible complementarities between abatement and certification. At the level of the original firm, these two are complementary by assumption: A firm can only benefit from abatement if its emissions are observable and this is only possible if the firm has certified. However, in the special case in which the abatement function is quadratic: $b(a) = \frac{1}{2}\hat{b}a^2$ and supply and demand curves are linear, it further holds that cheaper abatement (lower $\hat{b}$) will raise certification rates. A similar result holds for small taxes τ.

3 Unraveling in the International Case

3.1 International Model

We extend the model to an international setting and introduce a Foreign country in addition to the Home country. A Home policy maker values welfare both in Home and Foreign but can affect policy only in Home. This assumption reflects the interest of relatively rich countries in the global social cost of carbon in spite of narrow self-interest. We keep the quasi-linear setting and let the individuals in Home and Foreign have potentially distinct utility functions of the form:

$$U_H = C_{0,H} + u_H(C_H) - vG,$$

$$U_F = C_{0,F} + u_F(C_F) - vG,$$

where u_H is the strictly concave utility of consumption of good C_H in Home, and u_F the corresponding function in Foreign. They result in demand functions of $D_H(p)$ and $D_F(p)$, respectively. Consumers in both countries experience the same negative disutility from global emissions, G. There is a mass of m consumers in Foreign making total marginal disutility from emissions $(m+1)v$. All trade is costless.

The outside good, C_0, is produced emissions-free and competitively in both countries using identical linear production technology with labor as the only input. The marginal productivity is 1 and we normalize the price to 1 in the outside good sector. Home and Foreign have labor supplies of L_H and L_F, respectively, and we choose these such that the outside sector is active in both countries. This equates the wage in both countries to 1. The identical wages play no role in what follows. Market clearing for the outside good requires:

$$C_{0,H} + C_{0,F} = Q_{0,H} + Q_{0,F},$$

where $Q_{0,H}$ and $Q_{0,F}$ are production of the outside good in Home and Foreign, respectively. As before the polluting good is produced competitively by a continuum of mass 1 of firms in Home, but we add a mass m of firms in Foreign. The emissions per unit produced in Home is $\Psi_H(e)$ and the corresponding distribution in Foreign is $\Psi_F(e)$. We continue to assume that $\Psi_H(e)$ and $\Psi_F(e)$ are known. Our focus is on the ability of the policy maker in Home to affect production of the polluting good in Foreign and consequently, we focus on an equilibrium in which Home exports the outside good and imports the polluting good and no tax scheme is large enough to overturn this comparative advantage. With a price of p of the polluting good, the trade balance of Home is:

$$Q_{0,H} - C_{0,H} = p\left(C_H - Q_H\right),$$

where Q_H is the production of the polluting good in Home. Like in the domestic model, individual emissions for the Foreign firms are not observable unless the firm is certified. To focus on the international aspect, we assume that all emissions in Home are observable, though this matters little for the analysis to follow. With complete observability, firms in Home face an individual tax of $\tau_H e$ and following equation (6) the production in Home of the polluting good is:

$$Q_H(p) = E_H\left[s_H(p - \tau_H e)\right],$$

where $s_H(p)$ is the supply function for an individual firm in home and E_H denotes the expectation over $\Psi_H(e)$. Foreign firms have a production structure with cost function of $c_F(q)$ and corresponding supply function of $s_F(p)$, such that without facing a tax, foreign production is:

$$Q_F(p) = ms_F(p).$$

Although Foreign production is not directly taxed, Home can impose a tax on imports from Foreign. Initially, consider the setting in which there is no certification amongst Foreign firms and they all face a tax of $\tau_F E_F(e)$ on exports to home. Let p_H be the consumer price in Home and p_F the consumer price in Foreign. With positive exports and consumption in Foreign the post-tax price must be equal for Foreign firms whether they produce domestically or export:

$$p_H - \tau_F E_F(e) = p_F, \tag{18}$$

with corresponding total production in Foreign of:

$$Q_F = ms_F(p_F).$$

Therefore a market equilibrium is defined by equation (18) together with a market clearing condition:

$$D_H(p_H) + mD_F(p_F) = Q_H(p_H) + Q_F(p_F).$$

This gives prices in Home and Foreign without certification: (p_H^U, p_F^U).

Now, consider a setting in which a Foreign firm with emissions rate e, can certify to get access to the Home market at a tariff rate of $\tau_F e$. At the same time non-certified firms will be taxed at the average emission rate of non-certified firms. We continue to use ε as the pre-abatement effective emissions rate as observed by the government (as in equation 9), though only referring to Foreign firms.

All firms are atomistic and cannot effect world prices. Therefore a Foreign firm that certifies and sells both to Home and Foreign could sell all of its production to Foreign, save the cost of certification and make higher profits. Consequently, all Foreign firms that certify will export their whole production. There are therefore three possibilities: either all Foreign firms only sell to Home, non-certified firms sell some of their production to Home, or only certified firms sell to Home. Below we show that if there is no demand in Foreign, the International model essentially boils down to the Domestic model. We therefore impose conditions such that there is positive demand in Foreign. We look for an equilibrium in which an indifferent firm with emissions rate $\hat{e}$ is defined as:

$$\pi(p_H - \tau_F \hat{e}) - (F + f) = \pi(p_F),$$

and all firms with higher emissions rates, $e > \hat{e}$, do not certify, whereas all with lower do. This is analogous to equation (8) in the Domestic setting. Equation (18) is replaced with:

$$p_F \geq p_H - \tau_F E_F\left(e | e > \widehat{e}\right), \tag{19}$$

where only certified firms will export if the inequality is strict and uncertified firms will continue to export if the condition is binding. In either case, the world market equilibrium for the polluting good is defined as

$$D_H(p_H) + mD_F(p_F) = E_H\left[s_H(p_H - \tau_H e)\right] + mE\left[s_F(p_H - \tau_F \varepsilon)\right], \tag{20}$$

which equates world demand to world production, consisting of the production of Home firms, Foreign certified firms and Foreign uncertified firms, respectively.[9]

Our focus is on small taxes for which only the equilibrium with exports from non-certified firms is relevant, i.e. equation (19) continues to bind.[10] Consequently, we replace the $\geq$ with $=$ in equation (19).

The corresponding emissions from production in Home and Foreign, respectively,

[9] When only certified firms export, the pricing equation (19) is not binding and is instead replaced with a condition that Foreign consumption equals production by uncertified Foreign firms: $mD_F(p_F) = m(1 - \Psi_F(\hat{e}))s_F(p_F)$.

[10] More formally, for a contradication, consider an equilibrium with small taxes (first order) where post-certification only certified firms continue to export. For small taxes, the price difference between Home and Foreign must be small both before and after certification. Consider a given level of certification $\hat{e}$: certified firms will switch their entire production to Home and non-certified firms will shift their entire production to Foreign. These changes are zeroth order and must consequently have zeroth order effects on the price difference between foreign and home. If Foreign firms only partially adjust their production towards home the price difference between Foreign and Home prices can remain first order.

are then:

$$G_H = E_H[es_H(p_H - \tau_H e)],$$

$$G_F = mE_F\left[\varepsilon s_F(p_H - \tau_F\varepsilon)\right].$$

With total emissions given by

$$G = G_H + G_F \tag{21}$$

We consider a policy maker who values the welfare of agents in both countries equally, but who only has tools available to her in Home (that is $W = U_H + mU_F$). This set-up is in line with the objectives of unilateral climate change policies, which often aim at reducing emissions (and thereby climate change damages) not only at home but also abroad. For instance, countries like Sweden or Denmark are pursuing unilateral climate policies (through high carbon taxes) even though the damages from climate change are likely to be limited in these countries relative to other parts of the world. In addition, such a set-up allows us to abstract from well-studied terms-of-trade consequences of policy changes.

Note that in the special case in which Foreign consumers do not consume, the setup mirrors the domestic setting of Section 2.1 (where all Home firms certify) and the analysis of that section carries through. The behavior of Foreign consumers is therefore central to the analysis.

Certification will affect prices in both countries. We continue to use the superscript "V" to denote the voluntary certification scheme and we use "U" to denote the equilibrium without certification. For the incidence formulas, we use ϵ^D and ϵ^S to denote world demand and supply elasticities, which are the sum of share-weighted local elasticities, so $\epsilon^D = \epsilon^D_F\theta^D_F + \epsilon^D_H\theta^D_H$, and similarly for supply. In the appendix we show that such a voluntary certification program will have the following effect on the price at Home and in Foreign:[11]

$$\Delta p_H \equiv p^V_H - p^U_H = \underbrace{\frac{-\epsilon^D_F\theta^D_F}{\epsilon^S - \epsilon^D}\tau_F\left(E_F\left(e|e>\hat{e}\right) - E_F\left(e\right)\right)}_{\Delta p_H} + o(\tau) \tag{22}$$

[11]Our first-order approximations require for τ_H, τ_F and v to be small. Instead of writing $o(\tau_H) + o(\tau_F) + o(v)$ we employ the shorthand $o(\tau)$ and use $o(\tau^2)$ for the second-order polynomial in these terms.

$$\Delta p_F \equiv p_F^V - p_F^U = \underbrace{\frac{\epsilon_H^D \theta_H^D - \epsilon^S}{\epsilon^S - \epsilon^D} \tau_F \left(E_F\left(e|e > \hat{e}\right) - E_F\left(e\right)\right)}_{\Delta p_F} + o(\tau) \tag{23}$$

where elasticities are evaluated at the pre-certification price, p_0, which is identical in Home and Foreign without taxes. It holds that $\Delta p_H \geq 0$ and $\Delta p_F \leq 0$.

To understand the intuition for equations (22) and (23) note that a binding equation (19) requires certification to increase the difference between prices in Home and the Foreign: Certification effectively drives up the output tax on the firms who do not certify. Since these are the ones who both export to Home and supply the Foreign market the price difference must increase. Consequently, $\Delta p_H + \Delta p_F = \tau_F \left(E_F(e|e > \hat{e}) - E_F(e)\right)$, which is the increase in tariffs for uncertified firms. The way this price difference is shared between prices at Home and in Foreign is determined by the slope of supply and demand curves as in standard incidence formulas. Consider first the case in which $\epsilon_F^D = 0$. Then the international setting in effect mirrors the setting in Corollary 2: to a first order approximation, supply curves are linear, and supply responses from a mean-preserving increase in the variance of effective taxes is zero. Hence, with $\epsilon_F^D = 0$ the price response in Home is zero and the entire tax change is borne by Foreign. However, when $\epsilon_F^D < 0$, the lower prices in Foreign drives up demand and with it prices at Home.

Using equations (22) and (23) we can derive the consequences for emissions from certification where $s'_H \equiv ds_H/dp(p_0)$ and $s'_F \equiv ds_F/dp(p_0)$. The results are given in Lemma 3

Lemma 3. *Consider total emissions given by equation (21) and let G^V denote emissions with voluntary certification and G^U emissions without. Then the change in emissions by Home firms is*

$$G_H^V - G_H^U = s'_H E_H\left(e\right) \Delta p_H + o\left(\tau\right), \tag{24}$$

and the change in emissions from production by firms in Foreign is:

$$G_F^V - G_F^U = m s'_F \left[E_F\left(e\right) \Delta p_H - \tau_F Var_F\left(\varepsilon\right)\right] + o(\tau), \tag{25}$$

where (to a first order) the effect of certification on

- Home emissions is positive.

- Foreign emissions is ambiguous.

- *Total emissions is ambiguous.*

Proof. See Appendix □

Consider first the changes to emissions from production in Home (equation 24). Firms producing in Home only sell to Home and their tax schedule does not change. Consequently when prices increase, $\Delta p_H > 0$, they increase production and with it emissions. The effect in Foreign is ambiguous. One term is analogous to the term just described for Home: Since all firms in Foreign produce, at least partly, for the market in Home, an increase in p_H increases production and with it emissions. This, however, is not the only effect. Analogous to the domestic setting, the change in import taxes for the exporters in Foreign reshuffles production from firms with higher emissions to firms with lower emissions. This is captured by the term with $Var(\varepsilon)$, which reduces emissions. In sum, these two effects are ambiguous and emissions in Foreign can potentially increase. This leaves the effect on total emissions ambiguous as well. Equation (25) reveals a central element of the model: If the elasticity of demand in Foreign is sufficiently low, $\epsilon_F^D = 0$, then $\Delta p_H = 0$ as well. In this case, the negative effect on emissions from a change in production away from high-polluters unambiguously lowers emissions. The importance of the slope of demand in Foreign, D'_F is clear in Proposition 3 as well, which analyzes the welfare consequences.

Proposition 3. *The difference between social welfare for certification, W^V and social welfare without W^U is:*

$$W^V - W^U = \tag{26}$$

$$\underbrace{ms'_F \frac{(\tau_F)^2}{2} Var_F(\varepsilon)}_{\textit{Reallocation effect}}$$

$$\underbrace{-\left(v\left(1+m\right)-\tau_H\right)\left(G_H^V - G_H^U\right)}_{\textit{Untaxed emissions - Home}} - \underbrace{\left(v\left(1+m\right)-\tau_F\right)\left(G_F^V - G_F^U\right)}_{\textit{Untaxed emissions - Foreign}}$$

$$\underbrace{-\frac{\left(E_F\left(e\right)+E_F\left(e|e>\hat{e}\right)\right)}{2}\tau_F m D'_F \Delta p_F}_{\textit{Untaxed Foreign Consumption}} - F\Psi_F\left(\hat{e}\right) + o(\tau^2),$$

where

- *The reallocation effect is positive,*
- *The untaxed emissions effect in Home is negative if $\tau_H < v(1+m)$*

- *The untaxed emission effect in Foreign is ambiguous.*
- *The untaxed foreign consumption effect is negative.*
- $W^V - W^U + F\Psi_F(\hat{e})$ *has an ambiguous sign.*

Proposition 3 mirrors Corollary 2 in the domestic setting. It is instructive to first consider the case in which $D'_F = 0$, that is when consumption in Foreign does not change. Then there is no change in the price in Home (equation 22) and consequently no change in emissions from production in Home. The *untaxed foreign consumption* effect also disappears and the reallocation and untaxed emissions in Foreign effects collapse to $(v(1+m)-\tau/2)ms'_F Var^F(\varepsilon)$. This is identical to equation (12), except for the fact that emissions now affect $(1+m)$ people instead of 1. This term captures the combined effect of less cost-efficient, but less polluting production and is unambiguously positive. With $D'_F < 0$ two new effects come into play. First, with an increase in demand in Foreign prices go up in both countries and with it production in Home (the effect from untaxed emissions in Home). The second term is more subtle. Home can only tax production from Foreign that reaches Home. The rest is consumption in Foreign which is left untaxed. In the equilibrium without certification, the import tariff introduces a distortion as consumers in Foreign face a price that is $E_F(e)\tau_F$ lower than in Home and will consequently have lower marginal utility. Certification increases this price difference to $\tau_F E_F(e|e > \hat{e})$ and (inefficiently) increases Foreign consumption by $mD'_F \Delta p_F$. Due to the Taylor approximation the cost of this increased consumption is measured at the average of the distortion before and after the change: $(E_F(e) + E_F(e|e > \hat{e}))/2$, leading to the expression in the Untaxed Foreign Consumption effect. [12] A similar intuition explains why in general a border tariff adjustment (even tailored to the exact emission rate of the exporter) is not the optimal environmental tariff (see Keen and Kotsogiannis, 2014, and Balistreri, Kaffine and Yonezawa, 2019).

Together Lemma 3 and Proposition 3 make it clear that the elasticity of demand in Foreign is crucial for the welfare and emissions effect of the certification scheme: When a certain class of more polluting Foreign firms see taxes on their exports to

[12] Alternatively, think of Foreign consumption as non-polluting negative production. To make this point the clearest, let supply in Foreign be exogenous such that $s'_F = 0$ in which case all marginal increases in exports from Foreign to Home must come through (non-polluting) reductions in consumption. In such a case marginal exports from Foreign are emissions-free, yet still taxed. An increase in the tax rate from $E_F(e)\tau_F$ to $E_F(e|e > \hat{e})\tau_F$ increases the distortion by an amount equal to the area of the trapezoid described in the Untaxed Foreign Consumption effect.

Home increase, whereas sales to Foreign are left untaxed, their incentives to serve the domestic market increases. If the domestic market can absorb this extra production, that is if demand is elastic, welfare might decrease.

Adding abatement to the international setting is straightforward: foreign firms which certify now have an incentive to undertake abatement leading to an increase in efficiency and possibly lower emissions (depending on whether the rebound effect overtakes the direct effect of a reduction in emissions or not). We can directly combine the results of Corollary 4 and Proposition 3 to get the welfare effects of certification with abatement for small taxes and environmental damage: the welfare increase from certification is given by equation (26) plus an additional term which corresponds to the additional gains from abatement: $\tau_F \left(v - \frac{\tau_F}{2}\right) \frac{s_F(p_0)}{(b^F)''(0)} \Psi_F(\hat{e})$.

4 Conclusion

Settings in which a relatively small set of agents disproportionately contribute to a global externality perfectly encapsulates the problem of concentrated benefits and diffuse costs as described by Olson (1965). It should be unsurprising that there is limited appetite for Pigouvian taxes to internalize such externalities. In this paper we study a mechanism that counterbalances and ultimately unravels this dynamic.

The counterbalance comes from offering a substantial reduction in the tax burden to those who contribute little to the externality. Low-emissions agents receive concentrated benefits when they voluntarily certify their emissions for direct taxation. Increased certification raises the output tax on uncertified firms, but this marginal increase is dispersed widely. Unraveling occurs when the default output tax for uncertified firms is updated to reflect the higher mean emissions of the uncertified group and the cost of certification is not too large. If unraveling is complete, such a voluntary program achieves the same outcome as the otherwise-infeasible mandate to tax emissions directly (minus certification costs). We show that the welfare gains of such a policy scale with the variance of emissions and the slope of supply. The welfare acheived by a voluntary program is a weighted average of the Pigouvian first-best and the output-based tax, with the weights reflecting the relative variance of effective emissions subject to taxation to the variance of emissions in the population.

In the international setting, our mechanism extends the incentive to abate emissions beyond the borders of the country adopting a carbon tax. Such a policy is therefore most attractive for countries whose carbon footprint is most heavily embodied in imports. We show that the elasticity of demand for carbon-intensive goods in non-adopting countries plays a key role in prospective emissions reductions, as demand responses to lower prices abroad offset reductions from certified firms' abatement efforts. We derive the condition that determines whether such a program would further reduce emissions beyond those achievable with border carbon adjustments.

References

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi, "Self-Targeting: Evidence from a Field Experiment in Indonesia," *The Journal of Political Economy*, 2016, *124* (2), 371–427.

Aldy, Joseph E. and Robert N. Stavins, "The Promise and Problems of Pricing Carbon : Theory and Experience," *Journal of Environment & Development*, 2012, *21* (2), 152–180.

Balistreri, Edward J, Daniel T Kaffine, and Hidemichi Yonezawa, *Optimal Environmental Border Adjustments Under the General Agreement on Tariffs and Trade*, Vol. 74, Springer Netherlands, 2019.

Baron, David P., "Regulation of Prices and Pollution Under Incomplete Information," *Journal of Public Economics*, 1985, *28*, 211–231.

_ **and Roger B. Myerson**, "Regulating a monopolist with unknown costs," *Econometrica*, 1982, *50* (4), 911–930.

Barrett, Scott, "Self-enforcing environmental international agreements," *Oxford Economic Papers*, 1994, *46*, 878–894.

Battaglini, Marco and B. Harstad, "Participation and Duration of Environmental Agreements," *Journal of Political Economy*, 2016, *124* (1), 160–204.

Beccherle, Julien and Jean Tirole, "Regional initiatives and the cost of delaying binding climate change agreements," *Journal of Public Economics*, 2011, *95* (11-12), 1339–1348.

Besley, Timothy and Stephen Coate, "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs," *The American Economic Review*, 1992, *82* (1), 249–261.

Bohm, Peter, "Incomplete International Cooperation to Reduce CO2 Emissions: Alternative Policies," *Journal of Environmental Economics and Management*, 1993, *24*, 258–271.

Borenstein, Severin, "Time-Varying Retail Electricity Prices: Theory and Practice," *Electricity deregulation: choices and challenges*, 2005.

_ , "Effective and Equitable Adoption of Opt-In Residential Dynamic Electricity Pricing," *Review of Industrial Organization*, 2013, *42* (April), 127–160.

Chan, Gabriel, Robert Stavins, and Zou Ji, "International Climate Change Policy," *Annual Review of Resource Economics*, 2018, *10*, 335–60.

Condon, Madison and Ada Ignaciuk, "Border Carbon Adjustment and International Trade: A Literature Review," 2013.

Copeland, Brian R, "Pollution content tariffs , environmental rent shifting , and the control of cross-border pollution," 1996, *40*, 459–476.

Copeland, Brian R. and M. Scott Taylor, "Trade and transboundary pollution," *American Economic Review*, 1995, *85* (4), 716–737.

Dasgupta, Partha, Peter Hammond, and Eric Maskin, "On Imperfect Information and Optimal Pollution Control," *The Review of Economic Studies*, 1980, *47* (5), 857–860.

Deshpande, Manasi and Yue Li, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 2019, *11* (4), 213–48.

Dranove, David and Ginger Zhe Jin, "Quality disclosure and certification: Theory and practice," 2010.

Einav, Liran, Mark Jenkins, and Jonathan Levin, "Contract Pricing in Consumer Credit Markets," *Econometrica*, 2012, *80* (4), 1387–1432.

Elliott, Joshua, Ian Foster, Samuel Kortum, G K Jush, Todd Munson, and David Weisbach, "Unilateral carbon taxes, border adjustments, and carbon leakage," *Theoretical Inquiries in Law*, 2013, *14* (June), 207–244.

_ , _ , _ , Todd Munson, Fernando Pérez Cervantes, and David Weisbach, "Trade and carbon taxes," *American Economic Review*, 2010, *100* (2), 465–469.

Finkelstein, Amy and Matthew J. Notowidigdo, "Take-up and targeting: experimental evidence from SNAP," *The Quarterly Journal of Economics*, 2019, (2), 1505–1556.

Foster, Andrew and Emilio Gutierrez, "The Informational Role of Voluntary Certification: Evidence from the Mexican Clean Industry Program," *The American Economic Review*, 2013, *103* (3), 303–308.

_ and _ , "Direct and Indirect Effects of Voluntary Certification: Evidence from the Mexican Clean Industry Program," 2016.

Fowlie, Meredith L., "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage," *American Economic Journal: Economic Policy*, 2009, *1* (2), 72–112.

_ , Mar Reguant, and Stephen P Ryan, "Market-Based Emissions Regulation and Industry Dynamics," *Journal of Political Economy*, 2016, *124* (1), 249–302.

Grossman, Sanford J., "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 1981, *24* (3), 461–483.

Harstad, B., "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations," *Review of Economic Studies*, 2012, *79* (March), 1527–1557.

Hémous, David, "The dynamic impact of unilateral environmental policies," *Journal of International Economics*, 2016, *103*, 80–95.

Hoel, Michael, "Should a carbon tax be differentiated across sectors?," *Journal of Public Economics*, 1996, *59*, 17–32.

Hubbard, Thomas N., "The Demand for Monitoring Technologies: The Case of Trucking," *The Quarterly Journal of Economics*, 2000, *115* (2), 533–560.

Jin, Ginger Zhe, "Competition and disclosure incentives: an empirical study of HMOs.," *The RAND Journal of Economics*, 2005, *36* (1), 93–112.

— **and Phillip Leslie**, "The Effect Of Information On Product Quality: Evidence From Restaurant Hygiene Grade Cards," *The Quarterly Journal of Economics*, 2003, *118* (2), 409–451.

Jin, Yizhou and Shoshana Vasserman, "Buying Data from Consumers," 2019.

Jovanovic, Boyan, "Truthful Disclosure of Information," *The Bell Journal of Economics*, 1982, *13* (1), 36–44.

Kleven, Henrik Jacobsen and Wojciech Kopczuk, "Transfer Program Complexity and the Take-Up of Social Benefits," *American Economic Journal: Applied Economics*, 2011, *3* (February), 54–90.

Kwerel, Evan, "To Tell the Truth: Imperfect Information and Optimal Pollution Control," *The Review of Economic Studies*, 1977, *44* (3), 595–601.

Laffont, Jean-Jacques, "Regulation of Pollution with Asymmetric Information," in "Nonpoint Source Pollution Regulation: Issues and Analysis," Netherlands: Springer, 1994, chapter 2, pp. 39–66.

— **and Jean Tirole**, *A Theory of Incentives in Procurement and Regulation*, MIT Press, 1993.

Larch, Mario and Joschka Wanner, "Carbon tariffs : An analysis of the trade, welfare, and emission effects," *Journal of International Economics*, 2017, *109* (2017), 195–213.

Lewis, Gregory, "Asymmetric information, adverse selection and online disclosure: The case of ebay motors," *American Economic Review*, 2011, *101* (4), 1535–1546.

Markusen, James R, "International Externalities and Optimal Tax Structures," *Journal of International Economics*, 1975, *5*, 15–29.

Metcalf, Gilbert E and David Weisbach, "The Design of a Carbon Tax," *Harvard Environmental Law Review*, 2009, *33* (2), 499–556.

Milgrom, Paul R, "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *The Journal of Economic Perspectives*, 2008, *22* (2), 115–131.

_ **and John Roberts**, “Relying on the Information of Interested Parties,” *The RAND Journal of Economics*, 1986, *17* (1), 18.

Nichols, Albert L and Richard J Zeckhauser, “Targeting Transfers through Restrictions on Recipients,” *The American Economic Review Papers and Proceedings*, 1982, *72* (2), 372–377.

Nichols, D, E Smolensky, and T N Tideman, “Discrimination by Waiting Time in Merit Goods,” *The American Economic Review*, 1971, *61* (3), 312–323.

Nordhaus, William D., “Climate Clubs: Overcoming Free-riding in International Climate Policy,” *American Economic Review*, 2015, *105* (4), 1339–1370.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, 1965.

Roberts, Marc J. and Michael Spence, “Effluent Charges and Licenses Under Uncertainty,” *Journal of Public Economics*, 1976, *5*, 193–208.

Segerson, Kathleen, “Uncertainty and Incentives for Nonpoint Pollution Control,” *Journal of Environmental Economics and Management*, 1988, *15*, 87–98.

Spence, A. Michael, “Job Market Signaling,” *The Quarterly Journal of Economics*, aug 1973, *87* (3), 355.

Stiglitz, Joseph E., “The Theory of ”Screening,” Education, and the Distribution of Income,” *The American Economic Review*, 1975, *65* (3), 283–300.

Viscusi, W. Kip, “A note on ” lemons ” markets with quality certification,” *The Bell Journal of Economics*, 1978, *9* (1), 277–279.

Xepapadeas, Anastasios, “Environmental policy under imperfect information: Incentives and moral hazard,” *Journal of Environmental Economics and Management*, 1991, *20* (2), 113–126.

_ , “Observability and choice of instrument mix in the control of externalities,” *Journal of Public Economics*, 1995, *56* (3), 485–498.

_ , “The Economics of Non-Point-Source Pollution,” *Annual Review of Resource Economics*, 2011, *3* (1), 355–373.