Adverse Selection as a Policy Instrument: Unraveling Climate Change*

Steve Cicala
Tufts University and NBER

David Hémous
University of Zurich

Morten Olsen
University of Copenhagen

Click here for the latest version.

3rd August, 2022

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 consider settings in which aggregate emissions are known, but individual contributions are unobserved by the government. We evaluate a policy that gives firms the option to pay a tax on their voluntarily and verifiably disclosed emissions, or pay an output tax based on the average 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 sufficient statistics formulas to calculate the welfare of such a program relative to mandatory output or emissions taxes. We find that the voluntary certification mechanism would deliver significant gains over output-based taxation in two empirical applications: methane emissions from oil and gas fields, and carbon emissions from imported steel.

JEL Codes: D82, H2, Q54, L51, H87, K32.

*We are grateful to Thom Covert, Carolyn Fischer, Meredith Fowlie, Michael Greenstone, Suzi Kerr, Gib Metcalf, Frikk Nesje, Mark Omara, Mar Reguant, James Sallee, Joseph Shapiro, Andrei Shleifer, Jim Stock, Allison Stashko, Bob Topel, Reed Walker, and seminar participants at the Public Economics meetings of the 2022 NBER Summer Institute, EARE 2022, University of St. Gallen, IMF, World Bank, ETH Zurich, HEC Lausanne, Tufts, Universität Bern, the Environmental Defense Fund, UC San Diego, the Midwest EnergyFest, Yale, Harvard, the Utah Winter Business Economics Conference, UC Santa Barbara, UC Berkeley, and the University of Chicago for helpful comments. Iván Higuera-Mendieta provided excellent research assistance. Cicala gratefully acknowledges funding from the 1896 Energy and Climate 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: jurisdictional limitations, excessive implementation costs, and political opposition, among others. In this paper we show the extent to which such obstacles may be overcome when damage is caused by heterogenous agents. We apply results from the literature on mechanism design under asymmetric information, using adverse selection as a policy lever to encourage the voluntary revelation of harm and participation in Pigouvian taxation.

We consider situations in which the aggregate level of harm (such as pollution) is known by the government, but the exact contributions of specific agents are not. In such settings it is impossible to levy Pigouvian taxes due to the unobserved sources of pollution. The optimal uniform fee on goods whose production is associated with pollution (i.e. an output 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 (Cropper and Oates (1992); Schmutzler and Goulder (1997); Fullerton et al. (2000); Böhringer et al. (2017); Farrokhi and Lashkaripour (2021)).

We devise a tax mechanism that offers the *option* to certify one's damage, upon which a Pigouvian tax will be levied, combined with an output tax that tracks the average rate 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 tax paid by non-participants. Such a design 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 (Akerlof (1970)).

Structuring a voluntary certification program in this way has the potential to yield benefits in settings where it is otherwise impossible or undesirable to simply tax externalities directly. When enforcement is costly, for example, the regulator must balance the harm of the externality with that of ensuring compliance (Becker and Stigler (1974); Glaeser and Shleifer (2001)). We show how adverse selection can be employed to optimally separate high- and low-intensity polluters in a manner that economizes on enforcement costs. Voluntary emissions taxes also create large concentrated benefits for clean firms relative to output taxes (such as royalty fees). This can help yield an equilibrium outcome that is otherwise infeasible due to political economy opposition (Dal Bó et al. (2018)). When international jurisdiction

[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.

prevents direct externality taxation, we show how favorable treatment of low-pollution foreign firms allows a local government to better target the externality occurring outside of its borders.

We first develop a closed-economy model in which production is heterogenously associated with an externality and derive a sufficient statistics formula that approximates the difference between the first-best Pigouvian policy and an optimal output tax. This difference depends on marginal damages, the slope of the supply curve, variance of emissions, and monitoring costs. We 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, with weights equal to the relative variance of emissions under each policy.

We then devise a certification algorithm that policymakers can use to encourage certification when only the mean of the distribution is known. Under certain conditions this algorithm converges to an equilibrium in which the policymaker has full information. 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.

As an empirical application of the closed-economy model, we use this mechanism to internalize the cost of methane emissions from oil and gas production in the Permian basin in Texas and New Mexico. Methane is a potent greenhouse gas that leaks from the supply chain or is intentionally vented into the atmosphere in an unmonitored fashion. The Permian is the source of 30% of U.S. oil production and 10% of natural gas (Energy Information Administration (2019)), and recent estimates put the cost of methane emissions in the Permian basin alone at around \$4B per year (Zhang et al. (2020)). This setting is particularly well-suited for such a mechanism because output taxes are already in place (in the form of mineral royalties), and there is enormous heterogeneity in emissions rates across wells (Robertson et al. (2020)).

We find that while a royalty adder levied on production would reduce emissions by about 4%, a tax levied directly on pollution would reduce emissions by 80%. This striking difference is driven by a combination of incentives to abate emissions and a reallocation of production away from the dirtiest sites. In total, we estimate that welfare would be about \$1.2B higher per vintage (i.e. wells drilled in a particular year) under an emissions tax than an output tax. We find the marginal benefits from certification are increasing in the share of sites opting into an emissions tax in this setting. This implies that the voluntary mechanism is likely to unravel the distribution of emissions completely, and the cost efficacy of the program can be judged based on the enforcement cost-inclusive benefits of a mandatory emissions tax versus an output tax.

We then extend the model to an international setting as a mechanism of unilateral climate policy. Rather than examining the feasibility of international agreements subject to shirking incentives (Barrett (1994); Harstad (2012); Nordhaus (2015); Chan et al. (2018)), our approach focuses on the direct interactions between a government and foreign firms whose disclosures of emissions are 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. A voluntary certification mechanism does this by providing exporting firms outside the Home jurisdiction with the option to pay a carbon tax based on their certified emissions, or a tariff equal to the average emissions of uncertified firms.

Such a mechanism is therefore a twist on the tariff-based "Border Carbon Adjustment" (BCA) policies widely 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); Fowlie et al. (2021), see Condon and Ignaciuk (2013) for a literature review).[2] Our approach recasts the problem of jurisdiction into one of screening, in which clean foreign firms wish to separate themselves from more intensive polluters (Spence (1973); Stiglitz (1975)). In doing so, they adopt incentives to abate emissions that would otherwise be lacking under a BCA regime.

The key complication in the international setting is that firms serving foreign consumption remain untaxed, and their behavior responds to equilibrium prices induced by the tax mechanism. Unraveling leads to the expansion of dirtier, uncertified firms to serve the foreign market, or 'backfilling', which erodes the program's benefits. We characterize the conditions under which an unraveling mechanism is preferable to domestic carbon tax combined with a BCA on imports.

To demonstrate these forces at work, we consider the case of international trade in steel, an energy-intensive, trade-exposed sector that is central to environmental trade policy (Miller and Boak (2021)). We consider trade policy between the OECD and Brazil, a major steel exporter. Using the sufficient statistics formulas we develop, we estimate that an optimally-implemented certification program would achieve nearly three-quarters of the welfare gains of a global carbon tax. However, we also find that backfilling is a significant problem, which can be managed by restricting access to the emissions tax opt-in. At the extreme, an opt-in

[2]This literature generally assumes that BCAs are output-based. Böhringer et al. (2017) show in a CGE model that incentivizing abatement with an emissions tax would deliver large benefits over taxing output. Moreover, Cosbey et al. (2012) argue that allowing foreign firms to reveal their emissions would make the implementation of a BCA more compatible with WTO rules.

program without participation restrictions is slightly inferior to a standard border carbon adjustment. This example highlights the countervailing forces that limit a government's ability to reduce externalities outside of its jurisdiction, while also presenting a mechanism to productively expand its reach.

The combination of optional disclosure and a rolling default creates a policy that mimics the strategies 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)).[3] We 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 (Nichols et al. (1971); Nichols and Zeckhauser (1982); Besley and Coate (1992); Kleven and Kopczuk (2011)). A key distinction with a voluntary certification mechanism is that the treatment of agents outside of the opt-in is endogenously determined by the extent of program participation (González (2011)).

There is a long tradition of regulation under asymmetric information in the mechanism 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.[4] 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

[3]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)) and fisheries management (Holzer (2015)).

[4]See Segerson (1988); Xepapadeas (1991); Laffont (1994); Xepapadeas (1995), among others. Xepapadeas (2011) provides a review.

improved enforcement targeting for uncertified firms (Foster and Gutierrez (2013, 2016)), but does not considering changing audit probabilities as an instrument to encourage certification.

Though we focus on environmental externalities in our exposition and applications, the mechanism we describe may be applied by policymakers in a wide variety of settings. Additional examples include voluntary odometer certification to pay vehicle miles traveled taxes (shifting the tax burden to driving-intensive household's income tax bills), audits and criminal investigations (where scarce enforcement resources can be redirected toward higher risks), as well as programs where states retain primary jurisdiction over the federal government. The common thread running through prospective applications is that voluntary participation endogenously determines treatment under the default rate, and selection properties determining opt-in drive additional participation.

A note on feasibility is in order. We take as our point of departure that a policymaker cannot simply (and costlessly) implement the first-best and tax the externality directly. Would a voluntary emissions tax be feasible when a mandatory one is not? Since the initial circulation of this paper in 2019, policy developments regarding the empirical applications we study indicate real-world demand for this kind of regulatory design. A voluntary methane certification program-plus-royalty-adder was included in the Biden Administration's "Building Back Better" package and remains one of the key climate change mitigation programs under consideration (Whitehouse (2021)). Internationally, the European Union has made significant progress adopting a "Carbon Border Adjustment Mechanism" (CBAM) for voluntary certification of foreign firms and high default rate otherwise (Council of the European Union (2022)).[5] Given the potential challenges of certifying emissions in foreign jurisdictions through the supply chain, the concrete progress made by the EU suggests that such practical issues are not insurmountable.

Organizationally, the paper separates the analysis between domestic and international settings. In each setting we develop a theoretical model, and then apply the results from the model to an empirical application. The final section concludes.

2 Unraveling in the Domestic Case

To focus on the central issue of disclosure, we first consider a simplified setting of a closed economy in which firms differ only in their emission rates. Throughout this section we focus on a simplified model economy with an externality, though our approach would also apply

[5] In particular, draft language states that default values shall equal mean emissions by country-good plus a mark-up "building on the most up-to-date and reliable information, including on the basis of information gathered during the transition period."

to a broader class of models. We take as given that mandatory certification is not possible and characterize the welfare benefits of an optional certification program.

We derive "sufficient statistics" in the sense of approximations to changes in welfare that can be expressed through simple objects such as emissions variances and supply elasticities. We begin by solving a benchmark model for any level of certification, and then derive the optimal level of certification. We then weaken the information available to the regulator and show the program may be implemented knowing only the first moment of the emissions distribution. We conclude the section with two main extensions to the benchmark model: allowing for abatement and adjustments that only occur through entry and exit. These extensions yield expressions that share a common fundamental structure with the benchmark model. We provide further extensions, accounting for heterogeneous productivity and free entry in Appendix A.

2.1 Baseline model

A representative agent has preferences represented by the following quasi-linear utility function:

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

where C is total consumption of the polluting good and C_0 is the consumption of an outside good with a price of normalized to 1. G denotes emissions from the production of good C. The marginal social cost of emissions is v, and the outside good does not pollute.

The polluting good is produced by an (exogenous) mass 1 of firms who operate under perfect competition. Firms have the same strictly convex cost function $c(q)$ but vary in the extent to which they pollute. The emissions rate per unit produced is denoted by e and follows the cdf $\Psi(e)$, with full support and pdf of $\psi(e) > 0$, on the domain $[\underline{e}, \bar{e}]$ where $\underline{e} \geq 0$ and $\bar{e}$ may be infinite. 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).

2.2 Equilibrium with an output tax or emissions tax

In the following, we distinguish between a tax on emissions (if they were observable) and one on output. First, consider an output tax, t, which can be implemented even when individual emissions are not observed. Let the market price be p and solve the firm's problem in a decentralized equilibrium to get:

$$p = c'(q) + t. \tag{1}$$

This defines a supply function $q = s(p-t)$. With a mass 1 of firms this is also total production, Q. 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 $p-t$. Utility maximization gives: $u'(C) = p$ which together with $Q(p-t) = s(p-t)$ and $C = Q$ defines an equilibrium price, p, and quantity, Q. Emissions are given by:

$$G = \int_{\underline{e}}^{\bar{e}} s(p-t)e\psi(e)de = s(p-t)E(e) \tag{2}$$

Next, and in anticipation of the discussion of certification below, we solve for a setup in which emissions are observable and taxed at τ. This gives an individual supply function of $s(p-\tau e)$ and aggregate supply and emissions of:

$$Q = \int_{\underline{e}}^{\bar{e}} s(p-\tau e)\psi(e)de = E\left[s(p-\tau e)\right] \;\; and \;\; G = E\left[es(p-\tau e)\right] \tag{3}$$

The social planner would optimally set $t = vE(e)$ when taxing output and $\tau = v$ when emissions are taxable.

2.3 Equilibrium with certification

We now introduce *voluntary* certification of emissions, in which a firm can choose between two tax policies. If the firm chooses to (verifiably) reveal its level of emissions, e, it is taxed at τe (where we do not necessarily impose that τ equals the social cost of carbon, v). If 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. We will keep this relationship between τ and t as an assumption throughout most of the paper. We consider it a natural starting point since our interest lies with the *reallocation* of taxes based on better information of underlying emissions and not with changing the overall tax rates. Second, when the price of certification is set optimally, as in Section 2.4, $t = \tau E(e|R)$ is in fact optimal.

When presenting this choice to firms, the policy maker must calculate R *ex ante* based on knowledge of the distribution of emissions, $\Psi(e)$. Here we assume that she can do so and consider an alternative setting in Section 2.5. 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/subsidy 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:

$$\pi(p - \tau\hat{e}) - (F + f) = \pi(p - t). \tag{4}$$

Since the left-hand side is decreasing in e, all firms with $e < \hat{e}$ certify and firms with $e > \hat{e}$ do not. Consequently f is a tool to determine $\hat{e}$. The resulting tax rate on output for firms who do not certify is:

$$t = \tau E[e|e > \widehat{e}], \tag{5}$$

and where importantly, the rate at which uncertified firms are taxed is increasing in the mass of certified firms, $\partial t/\partial\hat{e} > 0$.

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

$$\varepsilon = \begin{cases} e & \text{if} \quad e \leq \hat{e} \\ E(e|e > \hat{e}) & \text{if} \quad e > \hat{e}, \end{cases} \tag{6}$$

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

$$Q = \int_{\underline{e}}^{\hat{e}} s(p - \tau e)\psi(e)de + (1 - \Psi(\hat{e}))s(p - \tau E(e|e > \hat{e})) = E(s(p - \tau\varepsilon)), \tag{7}$$

with corresponding emissions of:

$$G = E\left[\varepsilon s(p - \tau\varepsilon)\right]. \tag{8}$$

The equilibrium price follows from utility maximization $u'(C) = p$, which leads to a demand function $C = D(p)$, and market clearing $C = Q$. We consider sufficient conditions for this equilibrium to be unique in Appendix A. These include i) $E[e|e > \hat{e}] - \hat{e}$ is decreasing in $\hat{e}$ and ii) $s(\cdot)$ is weakly convex or τ is small. The condition on $E[e|e > \hat{e}]$ is satisfied for most frequently used distributions - such as the normal and the log-normal - and below we will rely on first order approximations in which case ii) is satisfied.

For any variable x we let x^V denote its value with certification and x^U its value under the output tax without certification. Comparing equations (8) and (2) gives the difference in emissions between the two tax systems:

Lemma 1. *The difference between emissions under voluntary certification, G^V, and the output tax, G^U, is given by:*

$$G^V - G^U = Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right] + E\left(e\right)\left\{E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E\left(e\right)\right)\right\}, \qquad (9)$$

where p^V and p^U are the equilibrium prices under certification and the output tax, respectively. The effect of certification on emissions is generally ambiguous. However, emissions decline when i) s is weakly convex and $es(p^V - \tau e)$ is concave in e; or ii) τ is small.

Proof. See Appendix A.1.2. □

The first term in equation (9) represents the reduction in emissions from output reallocation as low-emissions firms expand and high emissions firms contract. The second term combines two effects: i) a classical rebound effect as certified firms are taxed at a lower rate and increase production and with it emissions and ii) a price effect in that possible increases in the equilibrium price could further increase production. The conditions in the lemma rule out these severe cases of rebound. They are automatically satisfied for linear supply curves, as total production and hence the market-clearing price is unaffected by certification ($E(\varepsilon) = E(e)$ when all firms remain in operation). For a small τ, supply is close to linear and this remains true. With a reallocation towards less polluting firms, but no change in aggregate production, total emissions must decline.[6]

The following proposition gives the difference in welfare between the two settings for a given tax rate, τ:

Proposition 2. *The difference between social welfare with certification and the output tax is given by:*

$$\begin{aligned} W^V - W^U \quad = \quad & \underbrace{E\left(\pi\left(p^V - \tau\varepsilon\right)\right) - \pi\left(p^V - \tau E\left(e\right)\right)}_{\text{Output Reallocation Effect}} + \underbrace{\int_{p^U}^{p^V}\left(s\left(p - \tau E\left(e\right)\right) - D\left(p\right)\right)dp}_{\text{Price Effect}} \\ & \underbrace{-\left(v - \tau\right)\left(G^V - G^U\right)}_{\text{Untaxed Emissions Effect}} - F\Psi\left(\widehat{e}\right). \qquad (10) \end{aligned}$$

Where

[6] 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.

a) The "Output Reallocation Effect" is always positive and the "Price Effect" weakly positive.

b) The "Untaxed Emissions Effect" is zero for $v = \tau$, otherwise its sign is positive if falling (rising) emissions are under (over) taxed relative to Pigouvian levels.

Proof. Proof in Appendix 2. □

The first term in equation (10) is an *Output Reallocation Effect* and captures the increase in aggregate profits from the reallocation of production from firms with higher taxes to those with lower taxes. ε is the effective level of emission taxation under certification such that $E[\pi(p^V - \tau\varepsilon)] - \pi(p^V - \tau E(e))$ is the average gain for firms. This effect is always positive.

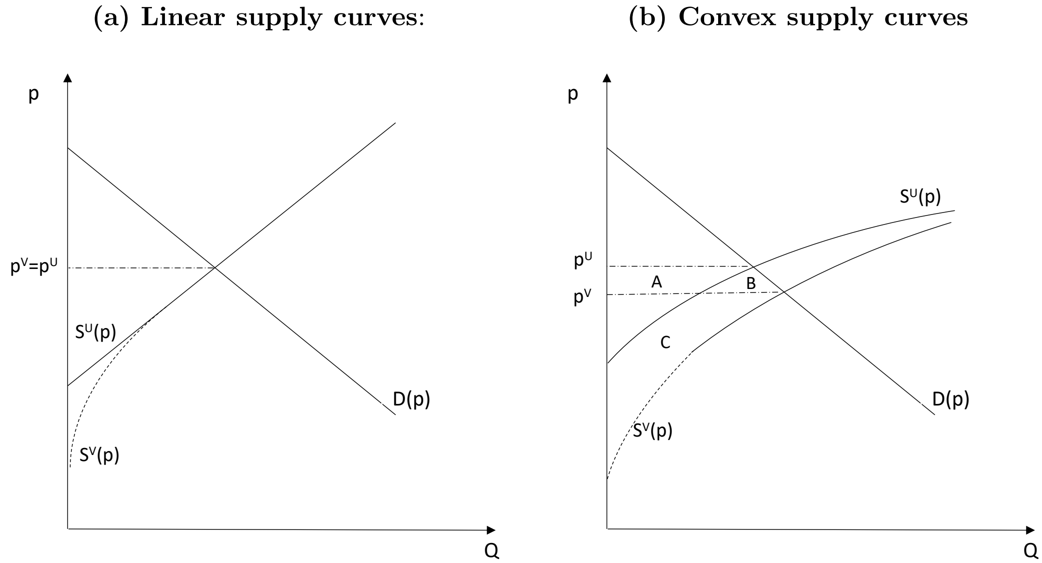
The *untaxed emissions effect* captures the welfare effects of changing emissions. These are zero when emissions are taxed at the Pigouvian level, $\tau = v$, and they are positive if emissions are undertaxed ($\tau < v$) and emissions decrease–which occurs depending on the conditions of Lemma 1. $F\Psi(\hat{e})$ captures the fraction $\Psi(\hat{e})$ of firms and the F resources required to certify them.

Figure 1 illustrates how output reallocation and price effects depend on the concavity of the supply function. $S^V(p)$ always intersects the y-axis lower than the $S^U(p)$ curve because firms with low emissions face lower taxes. For sufficiently low prices, the number of firms producing is increasing along $S^V(p)$. When the price is high enough for all firms to produce, the two curves overlap when supply curves are linear, as in Panel (a). In this region, the positive supply effect for firms with lower taxes is exactly matched by the negative effect for firms facing a higher tax. Total quantities are unchanged and there is no price effect. The output reallocation effect is measured by the increase in producer surplus, represented by the area between $S^V(p)$ and $S^U(p)$.

When supply curves are not linear, the price need not remain constant. Panel (b) considers the case of convex supply curves.[7] Convexity implies that the firms that face lower taxes will increase their production by more than the firms who face higher taxes will reduce their production. Consequently, the supply curve will be to the right. Again, the output reallocation effect is captured by the area between the curves (C), whereas the price effect is captured by B. The area A is a transfer from producers to consumers and does not feature in aggregate welfare changes.

[7] Section A.1.4 in the Appendix considers the opposite case in which supply is concave and the price increases. The price effect is still positive, but the allocation of welfare between producers and consumers is different.

Figure 1: Market Equilibria with and without Voluntary Certification



Note: p^V and p^U are equilibrium prices with and without voluntary certification, respectively. $D(p)$ denotes aggregate demand. $S^U(p)$ is the aggregate supply curve when firms are not certified, and they all face a tax of $t = \tau E(e)$. $S^V(p)$ denotes the case where some firms are certified, face different taxes, and consequently, different supply curves.

As can be intuited from Panel (b), the price effect is of a higher order than the other terms. Our goal is to establish "sufficient statistics" for changes in emissions and welfare that can be easily evaluated using readily-available data. We consider first-order approximations in τ, v in much of the analysis to come, as external costs are generally a small share of social marginal cost. This implies that the price effect is zero and the output reallocation effects and untaxed emission effects are both positive (for $\tau \leq v$). We obtain:

Corollary 3. *The expression $W^V - W^U$ in Proposition 2 can be written as:*

$$W^V - W^U = \left(v - \frac{\tau}{2}\right) s'\left(p_0\right) Var\left(\varepsilon\right)\tau - F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \tag{11}$$

with a difference in emissions of:

$$G^V - G^U = -s'\left(p_0\right)\tau Var\left(\varepsilon\right) + o(\tau), \tag{12}$$

where p_0 is the price when $\tau = 0$. It holds that:

a) The emission change $G^V - G^U$ is negative (to a first order) and its magnitude increases *with $\hat{e}$.*

b) Welfare gains gross of certification costs, $W^V - W^U + F\Psi(\hat{e})$ are positive and growing in $\hat{e}$ if $\tau < 2v$ (to a second order).

c) These expressions hold exactly for linear supply curves.

Proof. Appendix A.2. □

Corollary 3 demonstrates that the primary driver of the welfare consequences of certification come from the shift in production from more to less polluting firms (price effects are of a higher order). With constant aggregate production but a shift towards firms with fewer emissions, total emissions are sure to decline. Even if average emissions were already taxed at $\tau = v$, total welfare increases because production is reallocated towards less polluting firms. The size of this reallocation depends on the supply response, $s'(p_0)$, and the variance of the taxed emissions rate, $Var(\varepsilon)$, which, naturally increases as more firms certify. There are additional welfare gains from emissions reductions when $\tau < v$. When $\tau = v$ the entire welfare benefit from the certification program accrues to firms (at first order).[8]

How far is welfare under the voluntary certification in Proposition 2 from what would be achievable if firm emission rates were freely known and taxable (Jacobsen et al., 2020)? Labeling this latter equilibrium with W^{FI} for "full information" we find (at second order):

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

By construction $Var(\varepsilon) \leq Var(e)$ so welfare under voluntary certification (gross of certification costs) is a weighted average of welfare with no certification, W^U, and with full information, W^{FI}. The weight reflects the relative variance of the effectively-taxed emission rate, ε, and the actual emission rates, e. This means that greater benefits of voluntary certification accrue as a larger share of the variance of emissions certify. Having additional firms certify is only beneficial insofar as they represent a higher share of the emissions variance.

[8] A full political economy model is beyond the scope of this paper. However, the result that welfare benefits accrue to firms, coupled with the fact that in most settings the median firm pollutes less than the average, implies that the median firm would benefit from moving from an output tax to a certification system and would be willing to switch to such a system. Whether firms would be willing to accept an output tax compared with no tax depends on the allocation of property rights. If the output tax were implemented with a quota, the free allocation of quotas to firms could ensure that firms benefit from the certification scheme over and above no taxes. The result that the welfare benefits accrue to firms does depend on the assumption of an exogenous mass of firms (see Appendix A.5.3).

2.4 Optimal policy

We now show conditions under which the certification program decentralizes the optimal allocation. Assume that the social planner can offer any contract $(\pi^V(e), q^V(e))$ to any firm that reveals its emission rate where $\pi^V(e)$ denotes the payment to that firm and $q^V(e)$ its output, but is forced to offer the same contract to all non-certified firms (π^U, q^U). Revealing an emission rate involves a social cost F. We further assume that firms can imitate a higher emission rate.[9] Following the revelation principle, this leads to the IC constraint $\pi^V(e) \geq \pi^V(e')$ for $e \leq e'$. In addition, revelation is voluntary which implies the IC constraint $\pi^V(e) \geq \pi^U$ for all firms which choose to reveal. Combining these two constraints, one gets that the set of revealed firms must be an interval $[\underline{e}, \hat{e}]$. The problem of the social planner then becomes

$$\begin{aligned}\max_{\hat{e}, q^V(e), q^U} W = \;& u\left(\int_{\underline{e}}^{\hat{e}} q^V(e)\,\psi(e)\,de + (1-\Psi(\hat{e}))\,q^U\right) - \int_{\underline{e}}^{\hat{e}} \left(veq^V(e) + c\left(q^V(e)\right)\right)\psi(e)\,de \\ & - (1-\Psi(\hat{e}))\left(vE\left(e|e>\widehat{e}\right)q^U + c\left(q^U\right)\right) - F\Psi\left(\widehat{e}\right)\end{aligned} \tag{14}$$

with payments $\pi^V(e)$ and π^U that can be adjusted independently.[10]

Appendix A.3 shows that the optimal allocation can be decentralized using an emissions tax on certified firms, $\tau = v$, an output tax on uncertified firms, $t = vE(e|e > \hat{e})$, and a subsidy/tax on certification, f given by

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

The conditions $\tau = v$ and $t = vE(e|e > \hat{e})$ recover the standard Pigouvian result that emissions ought to be taxed at their (expected) social cost. Equation (15) shows that certification is excessive when τ and t are set optimally, and should be *taxed*. To see why, we employ the first order condition with respect to $\hat{e}$ which can be written as:

$$\pi\left(p - v\widehat{e}\right) - \pi\left(p - vE\left(e|e > \widehat{e}\right)\right) - vE\left(e - \widehat{e}|e > \widehat{e}\right)s\left(p - vE\left(e|e > \widehat{e}\right)\right) - F = 0. \tag{16}$$

Increased certification implies higher profits for the marginal firm, $\hat{e}$, captured by the first two terms in the expression. In addition, the quantity tax on uncertified firms goes

[9] If instead firms were unable to imitate a higher emissions rate, the optimum may involve revealing a set which does not take the form $[\underline{e}, \hat{e}]$.

[10] An auditing mechanism may lead to further welfare gains, but this is beyond the scope of this paper.

up, which has a negative effect on profits for the uncertified firms. The third term captures this pecuniary externality, whereas the fourth term captures the social cost of certification, F. As $\tau = v$, emissions are taxed correctly, and there is no marginal gain from changes to emissions. Comparing equation (16) with the decentralized economy indifference condition on certification (equation 4), we get equation (15). Since certifying firms do not internalize the costs they impose on uncertified firms, the optimal policy requires a tax.[11]

2.5 An "Unraveling" Algorithm

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 (11) gives an intuitive result of the welfare gains based on statistics that are relatively easily obtained — such as the variance of emission rates and supply elasticities — the implementation requires complete information on the distribution of e, which may not be available. We show the conditions under which a given "algorithm" can achieve a comparable outcome without complete information on the distribution of e.

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 pollution, 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 Ψ, it cannot predict the eventual threshold $\widehat{e}$ and therefore cannot implement the equilibrium described above by immediately announcing a new output tax $\tau E\left(e|e > \widehat{e}\right)$ (and potentially the optimal certification tax f).

Instead we consider an iterative process where the government progressively adjusts the output tax t_n, leading to a series of revelation thresholds $\widehat{e}_n$. As the government can observe the distribution of emissions below the threshold of certification in the previous period $\widehat{e}_{n-1}$, it can compute the average emission rate above certification from observables as $E\left(e|e > \widehat{e}_{n-1}\right)$. In period n, the government updates the output tax $t_n = \tau E\left(e|e > \widehat{e}_{n-1}\right)$ for non-certified firms. In period n, firms' decision to certify or not depends on comparing profits under certification $\pi\left(p - \tau e\right) - F$ with profits under no certification $\pi\left(p - t_n\right)$. Technically, this requires firms to form expectations about prices, so for simplicity here, we assume that the

[11] These considerations relate to the size of F and f, which are not specified in Corollary 3. Consider $f = 0$ (no tax or subsidy on certification). The private gains from certification are first order, so if F is first order, some firms will certify. Since social welfare gains are second order, equation (11) then delivers negative welfare benefits. If F is second order, almost all firms certify. Neither of these is efficient, as just discussed. An $f > 0$ of first order according to equation (15) ensures the optimal certification.

price is fixed. Therefore for each period we get the following stages:

1. Government chooses $t_n = \tau E\left(e|e > \widehat{e}_{n-1}\right)$ (with $e_0 = 0$)

2. Firms certify if $\pi\left(p - \tau e\right) - F > \pi\left(p - t_n\right)$, leading to a certification threshold $\widehat{e}_n$.

3. Certification is public, and the market equilibrium arises.

The logic can be straightforwardly extended to include an iteratively adjusted certification tax given by $f_n = \tau\left(E\left(e|e > \widehat{e}_{n-1}\right) - \widehat{e}_{n-1}\right) s\left(p - \tau E\left(e|e > \widehat{e}_{n-1}\right)\right)$. Proposition 4 shows conditions under which this process converges toward the market equilibrium with full information with and without the potential certification tax (see proof in Appendix A.4).

Proposition 4. *Assume that prices are exogenous and that $E[e|e > x] - x$ is decreasing in x. Then the procedure without a certification tax converges monotonically toward the unique equilibrium level of certification $\widehat{e}$. If in addition s is weakly convex or τ is small, then the procedure with a certification tax converges toward the social optimum.*

The certification equilibrium and the social optimum can therefore be implemented even if the government has no information on the distribution of emission rates. When the price p is endogenous, the evolution of the process depends on how price expectations are formed. Our results generalize if price variations are small—which is the case for instance when τ is small and firms forecast prices in period n assuming that no additional firms will certify in the current period.[12]

2.6 Extensions

The following subsection adds two dimensions of flexibility to the baseline model: abatement and adjustments that only occur along the extensive margin of entry/exit decisions. In Appendix A we additionally present results that allow for free entry and heterogenous productivity. The extensions fit easily in the framework presented thus far, and we present them with a focus on the added terms to the emissions and welfare expressions. Full derivations and proofs are in Appendix A.5.

[12] Alternatively, one could consider a "continuous" algorithm where the government allows firms to certify at increasing levels of emissions rates. That is, 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 continues. The government continuously adjusts the output tax and potentially the certification tax as the emissions distribution is revealed. We then 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.

2.6.1 Abatement

We keep the same structure as above but allow firms to spend $b(a)$ per unit produced to reduce their per-unit emissions by a. We alternatively consider proportional abatement below. We require: $b'(a) > 0$ and $b''(a) > 0$ for $a > 0$. For simplicity, we further add $b'(0) = b(0) = 0$. Pre-abatement emissions are still distributed according to $\Psi(e)$, and a certified firm i pays an emission tax on $e(i) - a(i)$ instead of $e(i)$. We continue to define ε in equation (6) as the pre-abatement emissions rate for certified firms and the conditional mean of emissions for uncertified firms.

Certification creates rewards for abatement that do not exist otherwise. When certified, firms solve the problem:

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

which leads to a common abatement level, a^*, among all firms that certify of $a^* = b'^{-1}(\tau)$ (if $\underline{e} < a$, some firms sequester emissions).[13]

For a small τ, we can write the abatement level as $a^* = \tau/b''(0) + o(\tau)$. Intuitively, $b''(0)$ captures the curvature in abatement costs. Since it is always profitable to do some abatement ($b'(0) = 0$), a low curvature of the abatement function directly implies that, to a first order, the optimal abatement level will be higher. Using Taylor expansions we then derive the analog of Corollary 3 with abatement (proof and details in Appendix A.5.1):

Corollary 5. *For small τ, the change in emission between an equilibrium with voluntary certification and one with an output tax is given by:*

$$G^V - G^U = -s'(p_0)\,\tau Var(\varepsilon) - \Psi(\widehat{e})\frac{\tau}{b''(0)}s(p_0) + o(\tau), \tag{17}$$

which is negative. The change in welfare is given by:

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right)\left(s'(p_0)Var(\varepsilon) + \frac{s(p_0)\Psi(\widehat{e})}{b''(0)}\right) - F\Psi(\widehat{e}) + o\left(\tau^2\right), \tag{18}$$

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

Corollary 5 takes advantage of the fact that to a first order $\hat{e}$ remains unchanged with abatement and consequently we can simply add a single term to the corresponding expres-

[13] The level of abatement chosen by firms is optimal, so the analysis of section 2.4 straightforwardly extends to this case.

sions (equations (12) and (11)) in Corollary 3. Abatement allows certifying firms to reduce their emissions. These firms also expand their production but for a small tax rate τ, this scale effect is dominated and the emissions of certifying firms decrease. The aggregate emission reduction is proportional to the aggregate initial production of certifying firms $s(p_0)\,\Psi(\hat{e})$. This explains the new term in equation (17).

In addition, abatement allows certifying firms to save $\tau a^*(\tau) - b(a^*(\tau)) = \tau^2/\left(2b''(0)\right) + o\left(\tau^2\right) > 0$ per unit of production, raising the profits of certifying firms. Taking into account the effect of untaxed emissions on aggregate welfare, we then obtain the new term in equation (18). For the same reasons as in the case without abatement, welfare gains gross of certifications costs are positive as long as $\tau < 2v$.[14]

Alternatively, one could consider that abatement allows to reduce the emission rate of firms by a share a. In that case one gets that the optimal abatement level is $a^*(e) = b'^{-1}(\tau e)$. At first order, a firm with emission rate e, will then abate a total amount $a^*(e)\,e = \tau e^2/b''(0) + o(\tau)$ per unit. The expressions (17) and (18) can then be directly generalized if one replaces $1/b''(0)$ with $E\left(e^2|e \leq \hat{e}\right)/b''(0)$ to reflect the average amount of abatement undertaken by certifying firms.

Overall this section illustrates a second important benefit of certification: by taxing emissions for a subset of firms, certification and the resulting individual taxation encourage abatement. Interestingly, for small environmental costs, under a Pigouvian level of taxation ($\tau = \nu$) and without free-entry, all the welfare gains from certification still accrue to (some) producers.

2.6.2 Adjustments along the extensive margin

The model presented above considered reallocation on the intensive margin. In this subsection we show how our main results carry through when all reallocation occurs on the extensive margin, such as the decision of whether or not to drill oil and gas wells (Anderson et al. (2018)). To focus on this alternative form of reallocation (and anticipating the context of our application to oil and gas wells) we consider an exogenous price p such that consumer welfare (excluding emissions) is constant.

[14] Analogously to equation (13), we can write welfare under certification with abatement (gross of certification costs) as a weighted average of welfare without certification, W^U, and welfare with full information under abatement, W^{FIA} as:

$$W^V = W^{FIA}\omega + (1-\omega)W^U - F\Psi(\hat{e}) + o(\tau^2),$$

where $\omega = \left(Var\,(\varepsilon) + \frac{s(p_0)}{s'(p_0)b''(0)}\Psi\left(\widehat{e}\right)\right)/\left(Var\,(e) + \frac{s(p_0)}{s'(p_0)b''(0)}\right)$.

Consider a mass 1 of potential firms each characterized by potential production, q, entry costs c and emissions, e. Firms have no additional production cost, and may not exceed their potential production, so the only decision is whether or not to enter. In anticipation of our empirical investigation, we specify the following relationship between cost and potential production: $c = uq$, where u is independent of e and q. One could think of costs scaling with the number of horizontal wells (which raise production) in the oil and gas drilling context, for example. Methane emissions per unit of production might depend on management practices (ensuring flares are lit) or distance to gas collection infrastructure. It is possible to solve the model without the assumption of independence of u from (e, q), but it allows us to more easily identify key moments of the (unobserved) distribution of potential entrants from the (observed) distribution of actual producers.

The triple (u, q, e) is distributed according to $\Psi(u, q, e)$ with a corresponding pdf of ψ. The domain is $[\underline{u}, \bar{u}] \times [\underline{q}, \bar{q}] \times [\underline{e}, \bar{e}]$ with weakly positive lower bounds and finite upper bounds and Ψ has full support. It will further be convenient to define the unconditional distribution of u, $\psi_u(u)$ and the joint distribution of (q, e), $\psi_{q,e}(q, e)$ such that $\psi(u, q, e) = \psi_u(u)\psi_{q,e}(q, e)$. In the following we present only results and keep derivations in Appendix B.

Equilibrium under an output tax. Firms face a uniform price of p. In laissez-faire, a firm i produces if its individual draw satisfies $pq_i \geq c_i \Leftrightarrow p \geq u_i$. When a uniform quantity tax is imposed, this condition becomes $p - t \geq u_i$. Total production and emissions are then given by:

$$S(p-t) = \int q\mathbf{1}_{(p-t)\geq u} d\Psi(u, q, e) \text{ and } G(p-t) = \int eq\mathbf{1}_{(p-t)\geq u} d\Psi(u, q, e), \tag{19}$$

where $\mathbf{1}_{(p-t)\geq u}$ is the indicator function and it is understood that when not otherwise specified integrals are over the full support of (u, q, e).

The slope of the aggregate supply curve is:

$$S'(p-t) = \psi_u(p-t) \int_e \int_q q\psi_{q,e}(q, e) dq de. \tag{20}$$

Price or tax changes only affect production by firms on the margin of entry so that ψ_u is only evaluated where $u = p - t$. Analogously the change in emissions from price changes are then:

$$G'(p-t) = \psi_u(p-t) \int_e \int_q eq\psi_{q,e}(q, e). \tag{21}$$

One can demonstrate that (to a first order) the optimal uniform tax rate is given by:

$$t^* = v\frac{G'(p)}{S'(p)} = \frac{\left(\int_e \int_q eq\psi_{q,e}(q,e)dqde\right)/\Psi_u(p-t)}{\left(\int_e \int_q q\psi_{q,e}(q,e)dqde\right)/\Psi_u(p-t)}. \tag{22}$$

The first equality holds under any $\Psi(u, q, e)$ but the second equality relies on the assumption of independence. The expression on the RHS is the average emissions divided by average production by operating firms which is observed in the data. In our empirical application we will rely on sample analogues of these expressions.

Equilibrium with certification. We set up a certification system along the same lines as for the intensive-margin case. We assume that the certification cost, F, is second-order in τ. Firms with $e \leq \hat{e}$ certify whereas those with $e > \hat{e}$ do not.[15] A firm that certifies will pay τe in emission tax and those that do not will pay according to the average emissions of active non-certified firms:

$$t = \tau\frac{E\left(eq|e \geq \hat{e}, (p-t)q \geq c\right)}{E\left(q|e \geq \hat{e}, (p-t)q \geq c\right)} = \tau\frac{\int eq\mathbf{1}_{e>\hat{e}}\mathbf{1}_{(p-t)\geq u}d\Psi(u,q,e)}{\int q\mathbf{1}_{e>\hat{e}}\mathbf{1}_{(p-t)\geq u}d\Psi(u,q,e)}. \tag{23}$$

With independence of u this is also the optimal tax rate on non-certified firms. This implies that total supply is given by:

$$S(p,\tau,t) = \int q\left(\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{(p-\tau e)\geq u+F/q} + \mathbf{1}_{e>\hat{e}}\mathbf{1}_{(p-t)\geq u}\right)d\Psi(u,q,e), \tag{24}$$

where expressions for total costs $C(p, \tau, t)$ and emissions $G(p, \tau, t)$ follow the structure of $S(p, \tau, t)$ but replace q with c and eq, respectively. The total mass of firms who certify is given by $M = \int \mathbf{1}_{e\leq\hat{e}}1_{(p-\tau e)\geq u+F/q}d\Psi(u,q,e)$. We proceed along the same lines as for Corrollary 3 to arrive at the following result:

Proposition 6. *Consider $\tau = v$. The difference in welfare and emissions between certifica-*

[15] In the intensive margin case, it was immaterial whether we chose an exogenous $\hat{e}$ or a tax f to incentivize the same level of certification. A tax here, however, will affect both the margins of certification and of entry and these two setups will not be equivalent. Therefore, we define $\tilde{e}$ as the maximum emission rate that the government permits to certify. With F being second order in τ, the constraint binds as long as $\tilde{e}$ is not too close to the average emission rate above that threshold $\frac{E(eq|e\geq\tilde{e})}{E(q|e\geq\tilde{e})}$. If that is the case, however, nearly all firms certify whether the constraint binds or not and our following analysis based on Taylor expansions still applies with the mass of certifying firms close to 1.

tion and the output tax is given by:

$$W^V - W^U = \frac{v^2}{2} S'(p) \left[E\left(\varepsilon^2 \frac{q}{E(q)}\right) - E\left(\varepsilon \frac{q}{E(q)}\right)^2 \right] - FM + o(\tau^2), \tag{25}$$

with ε *given by* $\varepsilon = e$ *if* $e \leq \hat{e}$ *and* $\varepsilon = E\left(e\frac{q}{E(q)} | e > \widehat{e}\right)$ *otherwise.* M *is the mass of firms that certify.*

The change in emissions is given by:

$$G^V - G^U = -\tau S'(p) \left[\left[E\left(\varepsilon^2 \frac{q}{E(q)}\right) - E\left(\varepsilon \frac{q}{E(q)}\right)^2 \right] \right] + o(\tau). \tag{26}$$

Proof. See Appendix B. □

The variance term is weighted by relative size of firms and is over the emissions rate at which firms are effectively tax, ε. The results mirror those of the intensive margin case and the intuition is analogous.[16]

2.6.3 Other extensions

Appendix A.5.2 studies the case where firms are also heterogeneous in productivity. We find that our baseline expressions for the change in emissions and welfare from Corollary 3 still apply provided that i) variance and expectations are taken over the (laissez-faire) quantity-weighted distribution of emission rates and that ii) the supply function is (close to) isoelastic.[17] Next, Appendix A.5.3 allows for free-entry. The expressions from Corollary 3 still apply so that the welfare gains remain identical, however, profits are zero in that case and the entire welfare benefit accrues to consumers.

2.7 Domestic Empirical Application: Methane Emissions in the Permian Basin

To demonstrate the potential power of a voluntary certification program to reduce pollution, we consider the case of methane leaks from oil and gas production. Methane (CH_4) is a powerful greenhouse gas, with an estimated social cost of \$1500/metric ton (Interagency Working Group on Social Cost of Greenhouse Gases (2021)), compared to \$51/metric ton

[16] The results of Proposition 6 do depend on the assumption of independence of u. In a more general setting, only the first equality of equation (22) holds. The optimal tax on uncertified firms is not in general equal to the average emissions of existing wells. In this case, the expression of welfare in Proposition 6 would feature an additional term for the welfare benefits of setting t optimally.

[17] Otherwise the optimal output tax is no longer proportional to the quantity weighted emissions rate, which complicates the analysis substantially.

for carbon dioxide. It is the primary compound in natural gas, and is released into the atmosphere during production due to faulty equipment, from safety valves to relieve pressure, or as a means of disposing waste as a byproduct when producing oil.[18] We focus our analysis on the Permian Basin of west Texas and southeast New Mexico, whose oil-rich shales yield nearly one-third of U.S. oil and 10% of U.S. natural gas production (Energy Information Administration (2019)). Zhang et al. (2020) calculate methane emissions from the Permian to have roughly similar global warming potential as CO_2 emissions from the entire U.S. residential sector, for a social cost of about $4B/year. As of 2022, there are no federal regulations restricting methane emissions from oil and gas production.

The oil & gas sector is amenable to a certification program along the lines we describe because output taxes are already ubiquitous in the form of royalty payments. On federal lands, for example, U.S. law specifies a minimum 12.5% royalty rate (30 USC 226), while increases may be made administratively without new legislation. This has catalyzed a push for 'royalty adders' to tax the carbon content of extracted energy (Gillingham et al. (2016); Gerarden et al. (2020); Prest and Stock (2021)). Following our results above, such an output tax would closely mirror the effects of a carbon tax if there were relatively little variance in carbon content for each fuel taxed.

The propensity to leak methane, however, is far more heterogenous. A ground-based random sample of oil and gas wells in the Permian found that 70% of emissions came from 15% of the measured sites, while nearly one third of measurements were below detectible levels (Robertson et al. (2020)). Using variation in natural gas prices to infer the marginal cost of abatement (methane leaked is methane not sold), Marks (2022) estimates that pricing methane pollution at its social cost would induce abatement rates of over 50%. Following Corollary 5, the combination of high emissions rate variance and plentiful abatement opportunities indicates the importance of taxing methane emissions directly rather than simply increasing the royalty rate to reflect average marginal damages. Passing new legislation to tax methane has proven politically challenging to date. The tax mechanism we describe, however, may be implemented as an opt-in tax rebate built on top of a royalty adder that is adjusted administratively. This suggests a fruitful path forward on methane emissions (Dal Bó et al. (2018)).

[18] Best practices for disposing of waste methane entail flaring the in situ, which converts the gas into less-harmful carbon dioxide. In North Dakota's oil-rich Bakken Shale, for example, approximately one third of natural gas was flared in the mid 2010's, making the sparsely-populated oil fields prominently visible at night from space (Cicala (2015)). The state adopted regulations to reduce flaring, and recent work has found a drop of about 20% through 2016 (Lade and Rudik (2020)).

We estimate the efficacy of an opt-in emissions taxation program based on our sufficient statistics results. We apply the formulas for extensive margin reallocation in subsection 2.6.2 because drilling is the key decision firms make—well production tends to decay deterministically over time regardless of prices (Anderson et al. (2018)). These results correspond to a long-run equilibrium in which firms make drilling and abatement decisions based on the expected lifetime production of sites. To account for details of this particular setting we expand our framework to account for output price heterogeneity and proportional abatement, detailed in Appendix B.2.

We calculate the impact of output and emissions taxes on the decision to drill and abate emissions from wells in the 2019 vintage. We combine the complete monthly production data from New Mexico and Texas with emissions measurements from a stratified random sample of wells (Robertson et al. (2020)) as well as collection and gathering (C&G) sites (Zimmerle et al. (2020)) via bootstrap sampling. We draw from these emissions samples to produce estimates of emissions rates for all wells drilled in 2019. Historical production decline curves are used to estimate the lifetime (i.e. 8 year) production from each well.

With the variance of emissions in hand, we require only the social cost of pollution ($1500/t), an estimate of the slope of the marginal abatement curve, and the slope of the supply curve to approximate the emissions and welfare changes due to voluntary certification according to Corollaries 3 and 5. For this last number we use an elasticity of supply of 1.26, which is the mid-point between the main estimates from Newell et al. (2019) (gas-oriented) and Newell and Prest (2019) (oil-oriented).[19] Output prices per BOE are constructed by weighing spot market prices for fuels from the Energy Information Administration by the site-level share of production from oil, gas, or natural gas liquids, respectively. Further details on data construction, the bootstrap procedure, and summary statistics are provided in Appendix D.1. Exact formulas for each calculation are provided in Appendix B.2.

We present the results of this analysis in Table 1. The first column summarizes expected eight-year production and emissions for wells drilled in 2019.[20] Working across columns we calculate changes relative to observed outcomes (laissez faire) for a universal output tax and a mandatory emissions tax without certification costs. We discuss certification costs at the

[19] These papers estimate the drilling elasticity based on Anderson et al. (2018), not the short-term change in production from existing wells, whose marginal costs are essentially ignorable. This is the relevant figure for us because we are interested in the long-term supply response to the voluntary tax mechanism, and in steady state, the production elasticity is equal to the drilling elasticity (Hausman and Kellogg (2015)).

[20] Our approximations for the impact of tax instruments are invariant to any underlying cost distribution, so to begin our analysis we need not take a stand on prevailing producer surplus. We assume throughout that world prices are unchanged, and therefore do not calculate impacts on consumer surplus.

end of the section.

We estimate that wells drilled in 2019 will produce about 3B barrels of oil-equivalent (BOE) over the subsequent eight years, and release methane emissions causing about $3B in external costs. For comparison, annual production from all vintages in 2019 was about 2.5B BOE in the Permian basin.

The second column of Table 1 presents estimates for approximate changes under an output tax that reflects the average social cost of methane per BOE, which is just over $1. This amounts to a 2.8% tax on oil and a 7.6% tax on natural gas at 2019 prices. The impact on production quantities is quite small, reflecting the modest size of the tax and approximate unit elasticity. The poor emissions targeting of an output tax in this setting and lack of abatement incentives means that it is not particularly effective at reducing pollution. Instead, the main effect of an output tax is to raise revenue, which it does at the cost of producer surplus. The value of emissions reduction is only slightly larger than the production distortion.

In contrast, we estimate taxing emissions directly would reduce methane releases by about 80% through a combination of reallocation and abatement. These results are in the third column of Table 1. Using our approximation formulas, we find that an emissions tax reduces producer surplus by about one third less than an output tax. There are substantial benefits to net welfare, worth over $1B for the 2019 vintage of wells.

The results thus far—and the core of our analyses—are based on approximations that do not depend on functional form assumptions. In the final column of Table 1 we provide a point of comparison with exact calculations for a specific case. For the distribution of costs, we assume a uniform distribution for u calibrated to observed prices, quantities, and our elasticity of supply of 1.26. The uniform distribution allows for an analytical solution, but also aligns well with published estimates of drilling cost distributions from Energy Information Administration (2016), presented in Appendix Figure B.3. For abatement, we assume a hyperbolic marginal abatement cost function that we parameterize to Marks (2022). In these calculations we also use the empirical distribution of emissions in each bootstrap sample, rather than simply the variance required for the sufficient statistics approach. All details of the model and formulas for each outcome are presented in Appendix B.2.4.

We find a smaller supply response and relatively less abatement when solving for exact solutions. This is consistent with a skewed emissions distribution.[21] The greater production,

[21] In particular, for wells with very high emissions rates, the cost of emissions may be greater than the value of output, making the well uneconomical to drill for any realization of costs. Our approximation assumes

but higher tax burden offset so that the aggregate change in producer surplus is quite close to that of the approximation formula. Greater emissions under the exact model yield a smaller reduction in external costs and greater tax revenues. The relatively lower emissions predicted under the approximation makes tax revenues for the exact model appear much larger by comparison. It is in fact just the difference in emissions times its social cost. Overall, both the approximation and exact calculation for a uniform cost distribution predict similar net social benefits of about \$1.2B per vintage.

We estimate gains along the unraveling path in Figure 2. The blue line represents the welfare gains (relative to laissez faire and excluding certification costs) when the share $\Psi(\hat{e})$ of wells certify and pay τe, while uncertified wells pay an output tax equal to $\tau E(e|e > \hat{e})$. When no firms certify, all production faces an output tax equal to $t_0 = vE(e)$. This corresponds to the intercept in the figure, and is equal to the approximate welfare change under an output tax in Table 1. At the other end, a mandatory emissions tax is represented by $\Psi(\hat{e}) = 1$, and the welfare gain is equal to that of the emissions tax approximation in Table 1.

To calculate welfare between these points, for each bootstrap sample we apply our approximation formula supposing each well were the marginal certifier. The figure plots the mean of these values across bootstrap samples against the emissions distribution.

We further illustrate how unraveling would develop when implemented by an information-constrained regulator using the algorithm in Section 2.5. There is a subtle difference in welfare achieved under the algorithm that merits discussion. Under the algorithm, the sequence of events is such that in the first round, the share $\Psi(\hat{e}_1)$ of wells certify, but uncertified wells pay an output tax of $\tau E(e)$, rather than $\tau E(e|e > \hat{e}_1)$. This under-taxation of uncertified wells yields a welfare gap represented by the difference between the blue line and the markers. We find this difference is relatively small but noticeable in the first round, but quickly converges to the upper line as the increments in conditional expectations become smaller.

In the first round, we estimate that about 65% of sites (denoted $\Psi(\hat{e}_1)$) would certify to have their emissions taxed versus pay an output tax based on average emissions rates, reflecting the skewness of the emissions distribution. In the first round there is minimal gain relative to a simple output tax. This is because the first wells to certify have relatively low emissions, and therefore deliver small abatement gains. At the same time, the uncertified wells would face the same tax rate as a simple output tax, for no additional welfare benefits.

that the cost of emissions is small relative to output and therefore misses this possibility: in effect it assumes that a negative amount of wells will be drawn for high emissions rate, thereby overstating the reduction in supply.

The first update of the default output tax to reflect the mean emissions of uncertified firms would yield about $300M in gains, with an additional 20% of sites opting into the methane tax. With each update the default output tax rate would grow until only the very dirtiest of wells remain uncertified. At that point, the output tax reflects emissions, but does not reward abatement. The producer surplus created from the lower tax burden under abatement would induce the final sites to certify.

The figure is plotted against the site-level emissions distribution, and the marginal welfare gains from certification in this setting are generally rising as additional sites' methane is taxed. If the costs of certification are relatively fixed per site, this implies that it is unlikely that there would be a point of diminishing marginal benefit from certification that equals the marginal cost. If the program fails to deliver benefits net of certification costs for complete certification, then it does not yield gains for any level of certification. It is clear, for example, that a program that does not update the default output tax would be particularly unattractive: it would require the costs of certifying 65% of wells, but do little otherwise. With about 3,000 sites in the 2019 vintage, we estimate pre-certification gains of about $425,000 per site. Exact certification costs are unknown, but the EPA has recently estimated the cost of a site visit and audit to be about $600 (U.S. Environmental Protection Agency (2020)). Even supposing monthly visits over eight years (and no returns to scale for such an extensive program) plus another $40,000 in equipment (to monitor whether flares are lit, for example), this would bring costs to only $100,000 for each site. The estimated benefits of certification would remain far in excess of certification costs. Combined with rising marginal benefits, this indicates that a voluntary certification program for methane emissions would ultimately deliver the welfare benefits of a mandatory emissions tax.

3 Unraveling in the International Case

We extend the model to an international setting. As discussed in the introduction, this is a particularly relevant setting for a voluntary tax mechanism. A country without jurisdiction abroad, but aiming to tax emissions embedded in its imports would need the agreement of the foreign exporter to reveal the actual carbon content of the goods. At the same time, it is generally possible to know the average carbon content of imports and therefore impose an output tariff based on this information. In this section we develop a model of emissions and welfare to measure the impact of a voluntary certification program that operates within

Table 1: Methane Emissions in the Permian Basin: Effects of Output and Emissions Taxes on the 2019 Vintage of Wells

		Predicted Change		
	Observed	Output Tax (Approximation)	Emissions Tax (Approximation)	Emissions Tax (Exact)
Quantities				
Production (Billions BOE)	2.92	-0.10	-0.10	-0.06
		[-0.19,-0.05]	[-0.19,-0.05]	[-0.12, -0.04]
Methane Emissions (Tg)	2.11	-0.08	-1.70	-1.14
	[1.12, 4.21]	[-0.28, -0.02]	[-3.74, -0.76]	[-2.40, -0.59]
Welfare				
Producer Surplus (Billion USD)		-3.10	-1.89	-1.96
		[-6.10, -1.67]	[-3.56, -1.07]	[-3.86, -1.09]
Tax Revenue (Billion USD)		3.04	0.61	1.45
		[1.66, 5.89]	[0.24, 0.99]	[0.81, 2.73]
External Cost (Billion USD)	3.16	-0.12	-2.55	-1.71
	[1.69, 6.31]	[-0.42, -0.03]	[-5.61, -1.14]	[-3.60, -0.88]
Total (Billion USD)		0.06	1.27	1.20
		[0.01, 0.21]	[0.57, 2.80]	[0.61, 2.46]

Note: Bootstrapped 95% confidence intervals in brackets. World prices are assumed to be invariant to policy, so consumer surplus is not calculated. All calculations are for the 2019 vintage of wells, based on estimated lifetime production and emissions (8 years). Formulas for each outcome are provided in Appendix B.2.3 and B.2.4.

these constraints.[22]

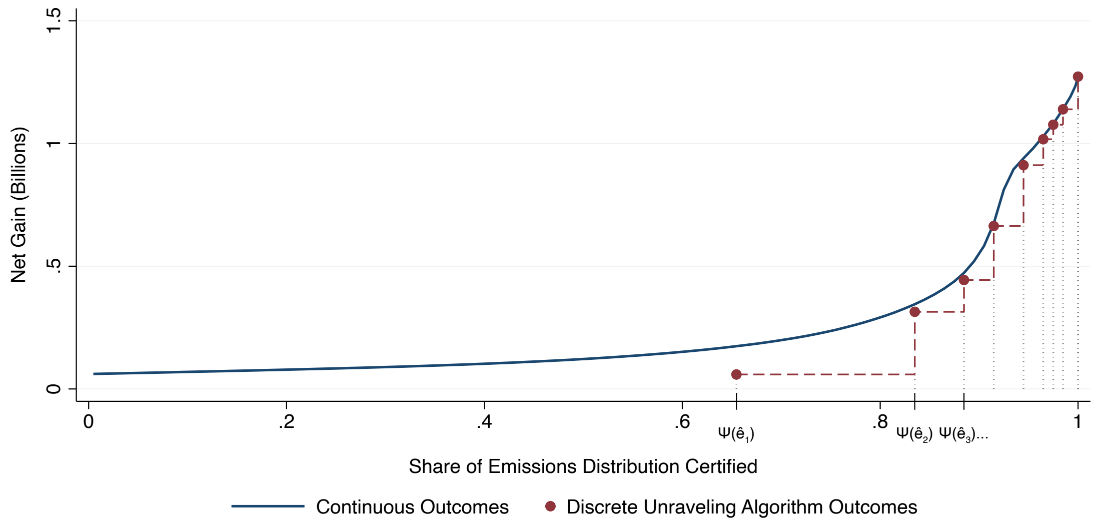
Specifically, we consider a Home policy maker who values welfare both in Home and Foreign but can implement policies only in Home. There is no Foreign policy maker. This assumption avoids various well-understood results regarding terms-of-trade manipulation and reflects the interest of relatively rich countries in the global social cost of carbon.[23] We consider a policy design analogous to that of the domestic model in Section 2: a uniform import tax on Foreign exporters is replaced with a voluntary certification program while uncertified firms pay a unit tax of $t = \tau E(e|e > \hat{e})$.

In the purely domestic model, certification lowers the tax on the less-polluting firms, raises it on those that pollute more, and, to a first order, keeps total production constant

[22]We note again that since the initial circulation of this paper, the EU has made substantial progress toward adopting a Carbon Border Adjustment Mechanism whose design closely follows the one we describe. These developments speak to the feasibility of implementing such a policy.

[23]As a matter of fact, the social cost of carbon used by the U.S. government is supposed to reflect global damages.

Figure 2: Methane Emissions in the Permian Basin:
Welfare Gains from Certification for the 2019 Vintage of Wells



Note: Gains are calculated relative to laissez-faire. Zero certification corresponds to an output tax based on the average emissions rate times the social cost of methane. The x-axis reflects $\Psi(\hat{e})$, the cumulative distribution of certified emissions rates. Outcomes for each point in the distribution are calculated for 1,000 bootstrap samples, and averaged. For continuous outcomes, wells with emissions rates in $[\underline{e}, e]$ certify, and wells in $(e, \bar{e}]$ face an output tax equal to the mean social cost of uncertified wells, $vE(e|e > \hat{e})$. For the discrete unraveling algorithm, the marker $\Psi(\hat{e}_1)$ indicates the share of wells that certify when faced with an output tax equal to $t_0 = vE(e)$. The markers $\Psi(\hat{e}_j)$ indicate certification thresholds as output taxes escalate to reflect average social cost of previously uncertified wells, i.e. $\Psi(\hat{e}_2)$ is determined by the output tax $t_2 = vE(e|e > \hat{e}_1)$, etc. The height of each marker indicates the associated welfare benefit.

and with it consumer prices. The international setting also features a reallocation of tax rates amongst Foreign producers, but two additional forces are present, both arising from (untaxed) Foreign consumption. First, whereas in a domestic setting the most polluting firms are forced to sell in a market where they face higher taxes, in an international setting, they might focus their production entirely on their domestic (Foreign) market, thereby avoiding the tax entirely. Second, if prices decline in Foreign, consumption there increases. Since foreign consumption is untaxed and consequently inefficiently high, this increase lowers overall welfare. We formally derive these two effects below.

We build on the structure of the model in Section 2. Individuals in Home and Foreign have potentially distinct utility functions of the form:

$$U_H = C_{0,H} + u_H(C_H) - \alpha v G \tag{27}$$

$$U_F = C_{0,F} + u_F(C_F) - (1 - \alpha)\, v G, \tag{28}$$

where $\alpha \in (0,1)$ denotes the population weight of H.

These result in demand functions for Home, $D_H(p)$, and foreign, $D_F(p)$. Consumers in both countries experience the same negative disutility, v, from global emissions, G. It costs κ units of the outside good to transport the polluting good between Home and Foreign. It is free to ship the outside good. The outside good, C_0, is produced emissions-free, competitively and one-for-one with labor. Labor is the only factor of production, and we choose the stock in both Home and Foreign such that the outside sector is active in both countries. We normalize the price to 1 in the outside good sector such that wages in both countries equal 1. The identical wages play no role in what follows.

The polluting good is produced competitively by a continuum of mass 1 of firms in Home and a mass of 1 in Foreign. The emissions per unit produced in Home are distributed according to $\Psi_H(e)$ with $\Psi_F(e)$ describing the Foreign distribution. Within each country, firms differ only in their emissions and share a common cost function. To focus on the international aspect and with little impact on the analysis to follow, we assume that all emissions in Home are observable and emissions are taxed at τ_H. Firms can abate as described in Section 2.6 and a Home firm will have a supply curve of $s_H(p - \tau_H e + A_H(\tau_H))$, where $A_H(\tau_H) \equiv \tau_H a_H(\tau_H) - b(a_H(\tau_H))$ is the net gain in price from abatement and $a_H(\tau_H) = b'^{-1}(\tau_H)$ is the level of abatement.

To streamline the presentation, the main text will be devoted to comparing a tax policy analogous to that of the domestic setting. Foreign firms can choose to certify and be taxed on their exports according to emissions at rate τ^F. Alternatively, non-certified Foreign firms pay an output tariff according to the average emissions of uncertified firms $t_F = \tau_F E_F(e|e > \hat{e})$, where E_F is the expectation operator over Ψ_F. Other comparisons exist, in particular allowing the output tax t_F to be set optimally. We will discuss alternatives towards the end of this section and explore the optimal output tax in Appendix C.3.

3.1 Equilibrium

As in the domestic case, there exists a cutoff $\hat{e}$ such that all Foreign firms with $e \leq \hat{e}$ certify while all other Foreign firms pay a common output tariff on their (potential) exports. There are two subclasses of equilibria: in a *pooled equilibrium*, uncertified firms are indifferent between selling domestically and exporting to Home. In a *separating equilibrium* uncertified firms only sell in the Foreign market. In either case, their production can be evaluated at the foreign price denoted p_F and be written as $s_F(p_F)$. Should an uncertified firm export, it would face an effective price $p_H - t_F - \kappa$. We can then define the difference in net price between selling in Foreign and Home for an uncertified firm, ρ, as:

$$\rho \equiv p_F - (p_H - \tau_F E_F(e|e > \hat{e}) - \kappa). \tag{29}$$

In a pooled equilibrium, $\rho = 0$ while in a separating equilibrium, $\rho \geq 0$.

The production of a certified Foreign firm can be written as $s_F(p_H - \tau_F e + A_F - \kappa)$. This expression reflects that a certified Foreign exporter pays the emission tax τ_F, abates at rate $a_F = b'^{-1}(\tau_F)$, leading to a net gain in price A_F, and faces the trade cost κ. The threshold $\hat{e}$ is then determined by an indifference condition similar to equation (4):

$$\pi_F(p_H - \tau_F \hat{e} + A_F - \kappa) - (F + f) = \pi_F(p_F), \tag{30}$$

where we take into account that in both types of equilibria, only uncertified Foreign firms serve the Foreign market. We can take $\hat{e}$ as a policy parameter determined by the certification tax f, or that $\hat{e}$ is set exogenously by the government only permitting firms below some predetermined level to certify.

The world market clearing condition for the polluting good is:

$$\begin{aligned} D_H(p_H) + D_F(p_F) \quad &= \quad E_H[s_H(p_H - \tau_H e + A_H)] \\ &\quad\; +\Psi_F(\hat{e})E_F\left[s_F(p_H - \tau_F e - \kappa + A_F)|e < \hat{e}\right] + (1 - \Psi_F(\hat{e}))s_F(p_F). \end{aligned} \tag{31}$$

The left-hand side constitutes world demand as a function of the two consumer prices p_H and p_F. The first two supply terms on the right reflect production subject to emissions taxes in Home and Foreign, respectively. The final supply term represents the output of uncertified firms in Foreign, some of which may be exported. The equilibrium is characterized by the two endogenous variables (p_H, p_F). Equation (31) contributes one equilibrium equation.

In the pooled equilibrium equation (29) binds which adds the second equation of the equilibrium. In that case, equation (29) directly gives that the difference in consumer prices, $p_H - p_F$, is increasing in $\hat{e}$: the average emissions of uncertified firms rises with certification, and the wedge between p_H and p_F grows to keep uncertified firms indifferent between serving Home and Foreign markets. Following a classic incidence formula the Foreign price decreases and the Home price increases in $\hat{e}$ when τ is small (see Appendix C.1).

The separating equilibrium features $\rho \geq 0$ and the second identifying equation is instead provided by market clearing in the Foreign market:

$$D_F(p_F) = s_F(p_F)(1 - \Psi_F(\hat{e})). \tag{32}$$

This equation alone pins down p_F as a function of the certification cutoff $\hat{e}$. The function is monotone positive: Higher certification implies fewer firms producing for the Foreign market and consequently a higher Foreign price. The Home price p_H is then given by equation (31) and depends negatively on $\hat{e}$. In the separating equilibrium, an increase in certification lowers prices in Home and raises them in Foreign, so that the price difference $p_H - p_F$ is decreasing in $\hat{e}$.

Whether greater certification raises or lowers the gap between p_H and p_F therefore depends on which equilibrium is active. This is determined by whether or not certified supply at p_H is sufficient to meet Home demand. If it is not, uncertified firms must be induced to export to the Home market at a price that makes them indifferent between selling locally and exporting. This activates the pooling equilibrium, $\rho = 0$ in equation (29). With greater certification in the pooled equilibrium p_H rises, and certified supply becomes sufficient to match $D_H(p_H)$. This then activates the separating equilibrium as uncertified exports cease and the price gap $p_H - p_F$ becomes unmoored from output taxes (which no one pays).

3.2 Welfare

We slightly abuse notation and let W denote world welfare. Under certification it obeys:

$$W = CS_H + CS_F + PS_H + PS_F - [(v - \tau_H)G_H + (v - \tau_F)(G_F - G_{F,F}) + vG_{F,F}] - F\Psi_F(\hat{e}),$$

where CS_H and PS_H refer to consumer surplus and producer surplus in Home with corresponding expressions for Foreign. These expressions are standard and the details are delegated to Section C in the appendix. The novel term is $-vG_{F,F}$ where $G_{F,F} \equiv E_F(e|e > \hat{e})D_F(p_F)$ corresponds to Foreign emissions generated for Foreign consumption: only uncertified firms supply the Foreign domestic market so that the average emission rate for foreign domestic consumption is $E_F(e|e > \hat{e})$. Intuitively, both Home production, G_H, and emissions from Foreign production but exported to Home, $G_F - G_{F,F}$ are taxed under Home jurisdiction and can be taxed so as to eliminate the externality ($\tau_H = \tau_F = v$). In contrast, $G_{F,F}$ is never taxed by the Home policy maker but is still a function of $\hat{e}$ and prices.

The following Proposition, analogous to Corollary 3, gives the difference in welfare and emissions between an equilibrium with voluntary certification and one without. Here we take approximations in τ and take v and κ to be of the same order:

Proposition 7. *To a second-order approximation in $(\tau_F, \tau_H, v, \kappa)$, the difference between global welfare under certification for Foreign firms and under a uniform output-based tariff of $\tau_F E_F(e)$ is given by:*

$$
\begin{aligned}
&W^V - W^U \qquad (33)\\
= \;& \underbrace{s_F'\tau_F\left(v - \frac{\tau_F}{2}\right)Var_F(\varepsilon)}_{\text{Reallocation Effect}} + \underbrace{\tau_F\left(v - \frac{\tau_F}{2}\right)\frac{s_F\Psi_F(\hat{e})}{b''(0)}}_{\text{Abatement Effect}} \underbrace{-\left[(v-\tau_H)E_H(e)s_H' + (v-\tau_F)E_F(e)s_F'\right]\Delta p_H}_{\text{Price Effect on Untaxed Emissions}} \\
& \underbrace{-\frac{D_F'}{2}\left(\tau_F\left(E_F(e) + E_F(e|e>\hat{e})\right) - \rho\right)\Delta p_F}_{\text{Consumption Leakage Effect}} \underbrace{-\rho s_F'\left(1 - \Psi^F(\hat{e})\right)\left(\frac{\Delta p_H + \rho}{2} + (v-\tau_F)E(e|e>\hat{e})\right)}_{\text{Backfilling Effect}} \\
& \underbrace{-F\Psi_F(\hat{e})}_{\text{Cost of Certification}} + o(\tau^2),
\end{aligned}
$$

where $\rho \geq 0$ — defined by equation (29) — is the (net of taxes) price premium for uncertified Foreign firms of selling in Foreign compared with Home. We let $\Delta p_H \equiv p_H^V - p_H^U$ and $\Delta p_F \equiv p_F^V - p_F^U$ denote the changes in prices due to certification in each country.[24] *It holds that*

[24]We can use first and second order approximations to get expressions for the price changes. We use ϵ^D and ϵ^S to denote world demand and supply elasticities, which are the sum of share-weighted local elasticities, so $\epsilon^D = \epsilon_F^D\theta_F^D + \epsilon_H^D\theta_H^D$, and similarly for supply with $\theta_H^D = D_H/(D_F + D_H)$ and analogously for other θ expressions. In Appendix C.2.2 we show that such a voluntary certification program will have the following effect on the price at Home and in Foreign:

$$\Delta p_H = \underbrace{\frac{-\epsilon_F^D\theta_F^D}{\epsilon^S - \epsilon^D}\tau_F\left(E_F(e|e>\hat{e}) - E_F(e)\right)}_{>0} \underbrace{-\rho\frac{\left(1-\Psi_F(\hat{e})\right)\epsilon_F^S\theta_F^S - \epsilon_F^D\theta_F^D}{\epsilon^S - \epsilon^D}}_{\leq 0} + o(\tau), \qquad (34)$$

$$\Delta p_F = \underbrace{\frac{\epsilon_H^D\theta_H^D - \epsilon^S}{\epsilon^S - \epsilon^D}\tau_F\left(E_F(e|e>\hat{e}) - E_F(e)\right)}_{<0} + \underbrace{\rho\frac{\epsilon^S - \epsilon_F^S\theta_F^S(1-\Psi_F) - \epsilon_H^D\theta_D^H}{\epsilon^S - \epsilon^D}}_{\geq 0} + o(\tau), \qquad (35)$$

where elasticities are evaluated at p_0 for Home and $p_0 - \kappa$ for Foreign. Further:

$$\Delta p_H - \Delta p_F = \tau_F\left[E_F(e|e>\hat{e}) - E_F(e)\right] - \rho. \qquad (36)$$

In the pooling equilibrium $\rho = 0$ and in the separating equilibrium $\rho \geq 0$. In this latter case, ρ is given by combining equations (29) and (30)

$$\rho = \tau_F\left[E_F(e|e>\hat{e}) - \hat{e}\right] - \frac{F+f}{s^F(p_0 - \kappa)} + o(\tau). \qquad (37)$$

- The Reallocation and Abatement Effects are always positive for $2v > \tau_F$

- The Price Effect on Untaxed Emissions is zero for $\tau_H = \tau_F = v$

- The Consumption Leakage Effect has the same sign as Δp_F *when* $f \geq 0$*, and always negative in the pooling equilibrium.*

- The Backfilling Effect is 0 in the pooling equilibrium and negative in the separating equilibrium (if $v \geq \tau_F$*).*

- The total net welfare change is ambiguous.

The corresponding changes in emissions are given by:

Corollary 8. *To a first-order approximation change in emissions from moving to certification are given by:*

$$G_H^V - G_H^U = E_H(e)\, s_H^{'} \Delta p_H + o(\tau), \tag{38}$$

$$G_F^V - G_F^U = \begin{array}{c} s_F^{'} \left(E_F(e)\,\Delta p_H - \tau_F Var_F(\varepsilon) + \rho E_F(e|e > \hat{e})\,(1 - \Psi_F(\hat{e})) \right) \\ -\Psi_F(\hat{e})\, a_F s_F + o(\tau) \end{array}. \tag{39}$$

The Proposition mirrors and extends the analysis in the domestic case. The Reallocation and Abatement effects are as described in Section 2, and $F\Psi_F(\hat{e})$ continues to be the cost of certification. We add the Price Effect on Untaxed Emissions, which arises since prices are no longer constant and an increase in home prices p_H might encourage global production. This is harmful to welfare if taxes are lower than v. In addition, we have the two terms with which we opened this section:

The *Consumption Leakage Effect* is new from the domestic case. This is a leakage effect because when Home imposes an output-based tariff (equal to $\tau_F E_F(e)$) the Foreign price goes down which encourages Foreign consumption which is not taxed. With voluntary certification, this distortion increases further if the Foreign price decreases, which always occurs in the pooling equilibrium. In contrast, in the separating equilibrium, increased certification might reduce the pool of Foreign producers servicing their own market so much that the Foreign price increases, in which case the distortion is mitigated. The welfare cost is proportional to the underpricing which (to a first approximation) equals $(\tau_F(E_F(e) + E_F(e|e > \hat{e})) - \rho/2)$, the average of the price gap under no certification and certification, respectively. This effect disappears when foreign demand is inelastic.[25]

[25] An alternative intuition is to think of decreases in Foreign consumption as a zero-emission way of increasing international supply. As such this "supply" should be given no tax and consequently be facing a

The *Backfilling Effect* is also new from the domestic setting.[26] Whereas the Consumption Leakage Effect captures that Foreign consumption faces a price that is too low, this term captures that the most polluting producers might receive a price that is too high in the foreign market. This effect is therefore negative (if $\tau_F \leq v$). Recall that the reallocation effect captures that uncertified firms receive a lower price and correspondingly reduce their production. This is true in both the Domestic case and in the pooling equilibrium, in which case the backfilling effect disappears. However, in the separating equilibrium where $\rho > 0$, Foreign firms can divert their sales entirely to the Foreign market, which is untaxed. Because the relatively clean firms are exporters, those left to serve demand in Foreign are relatively pollution-intensive. Uncertified Foreign firms consequently receive a price that is too high by ρ. Their price changes by Δp_F and the size of the distortion depends on the gap between Δp_F and what the price change would have been had these Foreign firms been forced to export namely $\tau_F(E_F(e) - E_F(e|e > \hat{e}))$. Note that $\Delta p_F - \tau_F(E_F(e) - E_F(e|e > \hat{e})) = \Delta p_H + \rho$. The welfare changes of the backfilling effect is then given by $\rho(\Delta p_H + \rho)/2$, where the division by 2 occurs for standard "Harberger" triangle reasons (for $v = \tau_F$). This effect is amplified when Foreign emissions are undertaxed.

The overall welfare effect is the sum of several terms and is in general ambiguous. However, if we consider a pooling equilibrium ($\rho = 0$), where taxes are Pigovian ($\tau_F = \tau_H = v$), abatement effects are large compared to fixed costs of certification $\left(\frac{v^2}{2}\frac{s_F}{b''(0)} > F\right)$, then a large Foreign supply elasticity relative to the demand elasticity (or s'_F large relative to D'_F) is a sufficient condition. The relative size of supply and demand elasticities will play some role in the empirical application below.[27]

Changes in emissions in Corollary 8 can be interpreted along similar lines. An increase in the Home price, Δp_H, increases emissions for all firms. Emissions in Foreign are reduced through a reallocation effect and abatement, but increase because of a backfilling effect if $\rho > 0$.

Consequently, Proposition 7 shows how Foreign demand alters the conclusion from the

price $p_H - \kappa$. However, the actual price of Foreign consumption is p_F (less than $p_H - \kappa$) and consumption is consequently too high. This problem grows when the Foreign price declines.

[26] We distinguish here between the "backfilling" effect, which refers to the welfare implications of changes in output from unregulated firms serving the Foreign market, and "reshuffling" which focuses on the emissions implications from rearranging buyer-seller matches to avoid regulation (Bushnell et al. (2008, 2014)).

[27] For classic terms-of-trade reasons, the imposition of an output border tax leads to a (first order) welfare transfer from F to H when F exports the polluting good. The move to voluntary certification has then ambiguous effects on welfare distribution from that point: on one hand it changes the price gap between the two countries which can hurt F, but on the other hand, the certifying firms increase their profits which benefits F.

domestic model of Section 2. Though this policy continues to reallocate production from the more polluting to the less polluting firms, the presence of Foreign consumption poses limits to the taxing ability of the Home policy maker. This makes the welfare effects of the certification mechanism less clear than in the domestic case. Nevertheless, we will find that welfare gains can be substantial when the mechanism is implemented judiciously in the empirical exercise of section 3.4.

3.3 Alternative policy environments

This naturally raises the question of the second-best policy, that is the optimal program when Home cannot differentiate between Foreign producers unless they certify and where Foreign does not impose any tax. We find that the social planner sets the emission tax for Home firms and certified Foreign firms at the Pigovian level: $\tau_F = \tau_H = v$ (derivations in Appendix C.3). However, in an attempt to reduce the Consumption Leakage Effect described above, the social planner sets the output tax on uncertified firms, t^*, lower than $\tau_F E(e|e > \hat{e})$ at:

$$t^* = \frac{s_F^{'}(p_F)(1 - \Psi_F(\hat{e}))}{s_F^{'}(p_F)(1 - \Psi_F(\hat{e})) - D_F^{'}(p_F)} v E_F(e|e > \hat{e}).$$

The social planner only taxes the uncertified firms at the Pigovian level if the Consumption Leakage Effect is inactive, that is if $D_F^{'}(p_F) = 0$. 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 (Markusen, 1975; Hoel, 1996; Keen and Kotsogiannis, 2014; Balistreri et al., 2019; Kortum and Weisbach, 2020). Certification enables the Home government to effectively extend its jurisdiction to tax certified Foreign firms like domestic firms, and it optimally adjusts the output tax to account for consumption leakage.

The optimal level of certification is set through a positive tax $f = (t^* - v\hat{e})s_F(p_F)$ which mirrors the expression from the domestic setting. Additional certification creates convergence in Home and Foreign prices for the separating equilibrium, and divergence in the pooling equilibrium. Setting taxes on output and certification allows the government to control this process.

In Appendix C.3, we compare the welfare under this optimal policy to that with a laissez-faire setting of no taxes on Foreign ($\tau_F = t = 0$) and Pigouvian taxes on Home production ($\tau_H = v$). We replicate expressions for the reallocation effect, gains from Abatement and the fixed cost of certification. The expression also features a Consumption Leakage effect, though contrary to Proposition 7 it is always negative because the spread between Home and Foreign consumer prices always increases in the optimum. Finally, since the comparison takes as a starting point no output tax there is a gain from introducing a tax.

Interestingly, the policy maker can potentially set trade taxes such that Home exports to Foreign, thereby creating bilateral trade in a homogenous product.[28] This is not optimal when the transport cost κ is of the same order as the social cost v (as in the steel example below). Yet, it becomes more valuable as transportation costs fall. In effect, the Home policy maker broadens its scope of taxation since it can tax production if it is either consumed or produced at Home. The result is Home exporting its production to Foreign (covered under a domestic carbon tax), and importing certified production from Foreign to Home (covered under the voluntary program), which reduces the Backfilling effect. This case may be relevant in contexts with very low trade costs such as electricity sold on a grid that spans jurisdictional borders.

3.4 International Empirical Application: Brazilian Steel Trade

To illustrate the welfare implications of a voluntary certification program in the international setting, we now conduct an analysis based on our approximation formulae for trade in steel between Brazil and the OECD. The iron and steel sector is one of the most energy and carbon-intensive sectors responsible for 10% of global CO_2 emissions (IEA, 2020). Iron and steel are internationally traded, and it is therefore considered a key sector for carbon leakage. Steel is mostly produced through two different processes. In the blast furnace-basic oxygen furnace (BF-BOF) process, coke and iron ore are combined at high temperature to produce liquid steel. This process emits CO_2 emissions through the combustion of coke. Alternatively, steel can be produced with scrap steel using an electric arc furnace (EAF), a process which leads to fewer emissions. Within each process, there is still substantial heterogeneity in emissions depending on plant energy efficiency and fuel used to produce electricity. This makes steel an interesting sector to explore the costs and benefits of voluntary certification.

Calibration

We calibrate the model to the Brazilian steel sector in 2019 and consider a two-country world where OECD countries alone decide to implement carbon tariffs (either output-based or with voluntary certification) on Brazilian steel. We focus on Brazil because it is one the major steel producers in the world and its market is particularly geared toward exports.[29]

[28] In a similar vein, Kortum and Weisbach (2020) develop a model of unilateral policy, and find that an export subsidy to increase access of (cleaner) Home goods to Foreign markets is part of the optimum.

[29] Brazil is the 9^{th} largest steel producer in the World. Its ratio of domestic consumption over production in 2019 is 64% and only Russia has a slightly lower ratio among the top 10 producers. Brazil is the second

We will use the welfare formula given in Proposition 7. This exercise requires a handful of key statistics and economic parameters of supply and demand. We provide a brief overview here, and give additional details in Appendix D.2.

Production, trade and transport costs. Brazil produced 32.6 Mt of steel in 2019, exporting 8.5 Mt (on net) to OECD countries (including 6.1 Mt to the US, the largest net export market for Brazilian steel) at an average price of \$489/t, which we take as the laissez-faire price of steel in Foreign in our calibration (Instituto Aço Brasil, 2021).[30] We use data from World Steel (2020) to determine production of steel in the OECD. We set the transport cost κ at \$50/t (Eckett (2021)).

Emissions rates. The EAF shares are 22.2% in Brazil and 69.7% in the US (World Steel, 2020). The emissions rates using the more energy-intensive BF-BOF is 2.07 tCO_2/t and 0.46 for EAF ((Hasanbeigi and Springer, 2019)). Corresponding numbers for the US are 1.82 and 0.62, respectively, implying that EAF producers are relatively clean in Brazil but the average for the whole industry is dirtier. We use the same data sources to estimate mean emission rates in the OECD.

Of course, there is also substantial variation in emissions rates within each process. To account for this, we assume that for each process, the emissions rate distribution is a double-bounded log-normal distribution. We let the standard deviation of log emission rates within each technology be the same as the standard deviation of log productivity in the basic metal products sector in Brazil according to Schor (2004).[31] See Appendix D.2 for details.

Social cost of carbon, abatement cost and certification cost. We set the social cost of carbon at \$51 (Interagency Working Group on Social Cost of Greenhouse Gases (2021)). To parameterize abatement costs, we match abatement according to our formula with abatement according to a marginal abatement cost curve for the steel sector in Brazil from Pinto et al. (2018) for a \$51 tax. We find that firms abate 0.169 tCO_2/ t steel at this level. This

largest exporter to the US after Canada.

[30]We keep constant net exports of Brazil to non-OECD countries (2Mt) and remove them from Brazilian production. Net exports to the OECD therefore represent 27.7% of Brazilian production (excluding net exports to non-OECD countries).

[31]This assumes implicitly that i) Once one controls for the process-type, most heterogeneity reflects differences in energy intensity and that those move with TFP differences; ii) In the basic metal sector, within subsector heterogeneity dominates heterogeneity across sub-sectors. We adjust the total standard deviation of log productivity for the productivity premium enjoyed by the EAF process on average, using estimates for the US from Collard-Wexler and De Loecker (2015). This productivity premium is much smaller than the difference in log emission rates, so the standard deviation of log emission rates of the joint distribution is larger than that of log TFP—this is consistent with evidence from Lyubich et al. (2018) showing that emissions rates tend to be more heterogeneous than TFP.

Table 2: Emission and Welfare Effects from Environmental Trade Policies

	First Best	BCA	Voluntary Certification $f = 0$	$f = f^*$
Welfare				
Gains in M USD	1212	714	692	866
% of First Best Gains	100	58.9	58.4	71.5
Emissions				
Reduction in Mt	24.4	5.6	6.3	11.1

Note: All gains are calculated relative to a unilateral domestic carbon tax in the OECD without border adjustments. First best is a global carbon emissions tax. f is a tax on certification, with f^* denoting the optimum certification tax.

is conservative relative to McKinsey & Co. (2009), which predicts a marginal abatement cost of \$50 for 0.526 tCO_2/ t steel in 2030. To estimate the certification cost F, we rely on Gallaher and Depro (2002) from the EPA who find that the annualized cost of monitoring a variety of pollutants (but not CO_2) in the iron and steel sector was \$1.04M per plant in 2001. Assuming a constant fixed cost to total output ratio, this corresponds to \$16 M to certify all production in Brazil (or around \$0.49 per ton).

Elasticities. Finally, we use Fernandez (2018) who derives demand elasticities for steel in the US (-0.306) and in Brazil (-0.414). For the supply elasticity, we use a single value of 3.5, which is the supply elasticity used by EPA (2002). Data Appendix Table B.1 gives all parameters and their sources.

Results

Table 2 reports the effects on global welfare and emissions of introducing various taxation programs. Each of the calculations report the gains relative to a Pigovian emission tax in the OECD with no border adjustment. As a benchmark for comparison we calculate the welfare gains that would result from a universal carbon tax covering production in both the OECD and Brazil. To give some context to these numbers, note that the net export value of steel from Brazil to the OECD is $4.1B$ USD. Adding an output-based carbon border adjustment (i.e. a tariff on Brazilian exports of $vE_F(e)$) increases welfare by $714M$ USD which is already 59% of the gains that could be achieved by implementing the universal carbon tax.

Without any certification tax, the voluntary certification program leads to a high level of

Table 3: Decomposition of Welfare Effects from Voluntary Certification Relative to BCA

	Certification Fee	
	$f = 0$	$f = f^*$
Welfare Component		
Reallocation	150	150
Abatement	37	37
Consumption Leakage	40	-19
Backfilling	-228	-10
Certification Costs	-4	-4
Total Change in Welfare Relative to BCA	-6	152

Note: This table decomposes the welfare changes from voluntary certification relative to an output-based tariff (BCA) into the theoretical channels discussed previously for two cases: when there is no certification tax and under the optimal certification tax. All numbers in millions of USD.

certification. The economy ends up in the separating equilibrium with a price gap between Home and Foreign which is lower than in the output-based tariff case (105 USD versus 137 USD). As a result, without any certification tax, voluntary certification very slightly reduces welfare relative to an output-based tariff by $6M$ USD—though this ceases to be the case if abatement costs were just 25% lower (see Appendix D.2).

Table 3 decomposes the welfare change from the output-based tariff to the voluntary certification program into the different channels discussed above (Appendix D.2 shows how our estimates change with alternative parameters). It shows that the decrease in welfare comes from a large negative "backfilling effect"—a scaling up of production by the dirtiest Brazilian producers to serve the domestic market— of $-228M$ USD, while the "reallocation effect" leads to gains of $150M$ USD.

A tax or other program to limit certification to the optimal level, however, brings $152M$ USD of additional welfare gains relative to the output-based tariff. This corresponds to a third of the gap between the output-based tariff and the first best policy or 71% of the first best gain relative to a unilateral carbon tax in the OECD only. With the optimal certification tax, the equilibrium is still in the separating case but the price gap slightly increases relative to the output-based tariff at 148 USD. This leads to a "backfilling effect" of a much smaller magnitude at $-10M$ USD, while the reallocation effect remains $150M$

USD since the certification threshold is close in both cases.[32] With our baseline calibration, abatement gains are comparatively small at 37 M USD.[33] These gains would be substantially larger using the estimates of McKinsey & Co. (2009), so that even an unrestricted voluntary certification program would deliver gains over a standard BCA.

Finally, note that while the output-based tariff reduces emissions by $5.6M$t of CO_2, the optimal certification program reduces emissions by $11.1M$t of CO_2, nearly half of the reductions of a first best policy which also taxes emissions in Brazil. To give an idea of the magnitude involved, embodied carbon in the net exports of steel in laissez-faire represents $14.5M$t of CO_2.

To illustrate how the benefits of an opt-in emissions taxation program change with the extent of certification, Figure 3 displays the welfare gains of the certification program for different values of the certification tax relative to a Pigovian tax in the OECD only. The certification tax is expressed as a share of revenues in laissez-faire. For comparison, we also plot the gains of a standard border carbon adjustment. For a sufficiently high certification tax (corresponding to 15.6% of a firm's laissez-faire revenues), no firm certifies and the welfare gains are the same as with the output based-tariff. As the certification tax decreases (moving toward the left in the Figure), the share of firms certifying grows quickly, bringing most of the welfare gains from the certification program. With a certification tax as high as 15% of a firm's laissez-faire revenues, 16.3% of firms certify, close to 3/4 of EAF producers in Brazil. This reflects the presence of a sizable mass of relatively clean producers in Brazil. The welfare gains remain high (above $100M$ USD) as long as the certification fee remains above 4% of revenues, and they only disappear for trivial certification taxes.[34]

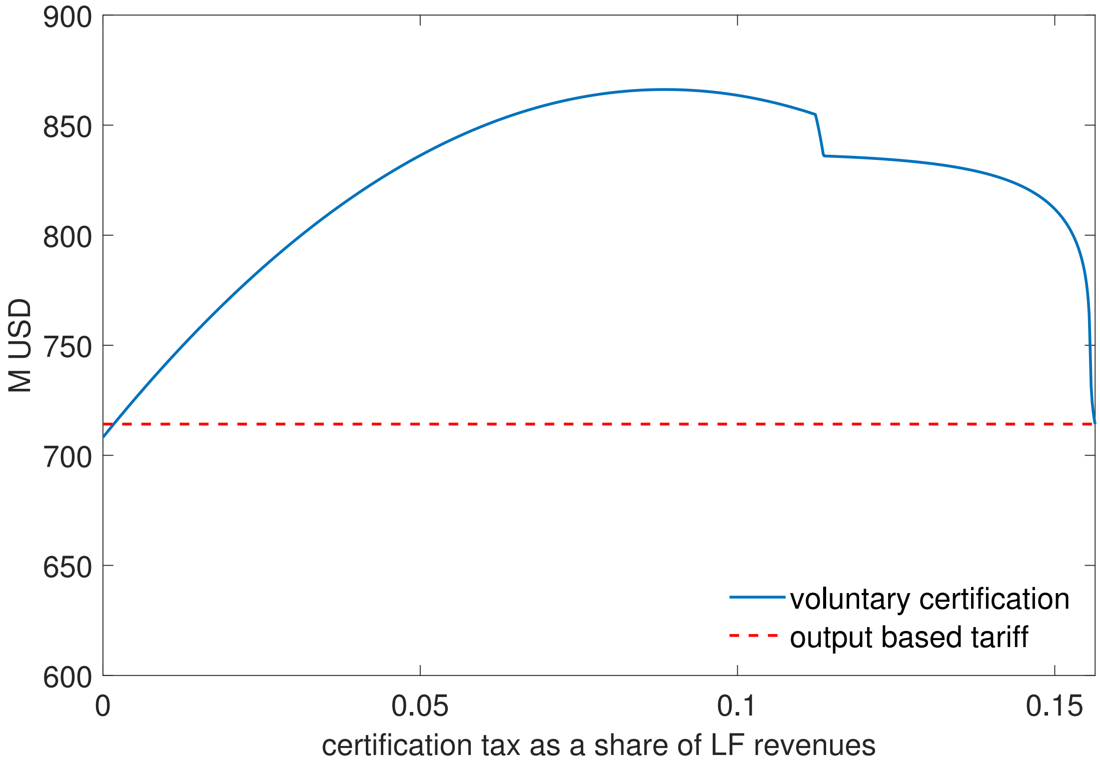
We stress that this exercise is a simple proof-of-concept and that our numbers are indicative of orders of magnitude but not exact values. It shows that there are potentially large emission reductions and significant welfare gains from such a certification program, though it may not be desirable to let certification occur without restriction if it leads to a large reduction in the price gap between the two countries. Fortunately, prevailing price gaps

[32] In the separating equilibrium the mass of certifying firms $\Psi(\hat{e})$ is close to the export share in laissez-faire, i.e. $\Psi(\hat{e}) = 1 - D_F(p_0 - \kappa)/S_F(p_0 - \kappa) + o(1)$. This is the reason why we show welfare gains as a function of the certification tax instead of $\hat{e}$ in Figure 3 below.

[33] The second best policy described in section 3.3 also leads to a separating equilibrium. In that case, the tariff t^* on uncertified exporters is irrelevant, so that the second best policy coincides with the optimal certification tax case reported here.

[34] The small jump in Figure 3 marks the point where we switch from the separating to the pooling equilibrium. This jump only occurs because we compute the welfare gains using Taylor approximations and would disappear if we were to compute welfare gains at higher orders.

Figure 3: Welfare Gains Relative to a Carbon Tax at Home Only for Different Levels of the Certification Tax



Note: The x-axis uses certification taxes to vary the extent of Foreign firm emissions taxation. A high certification fee yields $\hat{e} = 0$, and is equivalent to a standard BCA. The extent of certification $\Psi_F(\hat{e})$ rises as certification taxes fall

relative to border adjustment fees are observable. This makes it possible for policymakers to adjust certification criteria in order to avoid significant backfilling losses.

4 Conclusion

Settings in which a relatively small set of agents disproportionately contribute to a global externality perfectly encapsulate the problem of concentrated costs and diffuse benefits 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 achieved 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.

We apply these results to oil and gas production in the Permian basin of New Mexico and Texas, where methane emissions are a significant, largely unregulated problem. Coupling a royalty adder based on the average of uncertified emissions per barrel of oil equivalent with the option to certify emissions sets off an unraveling that converges on universal taxation, even with significant implementation costs. This would yield welfare gains of over \$1.2B per vintage from the Permian basin alone.

In the international setting, the voluntary emissions tax 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 heavily embodied in imports. We show that in addition to the variance of emissions, 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 (consumption leakage). In addition, with too much certification there is a risk that the most pollution-intensive foreign producers will expand operations to serve the foreign market (backfilling). We derive conditions that determines whether such a program would further increase welfare and reduce emissions beyond those achievable with border carbon adjustments. Applying our sufficient statistics formulas to the Brazilian-OECD steel trade, we find that a managed certification program could deliver nearly three quarters of the welfare gains of a universal carbon tax.

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.

Alquist, Ron and Lutz Kilian, "What do we learn from the price of crude oil futures?," *Journal of Applied Econometrics*, June 2010, *25* (4), 539–573.

Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant, "Hotelling Under Pressure," *Journal of Political Economy*, 2018, *126* (3), 984–1026. Publisher: National Bureau of Economic Research.

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. Publication Title: Environmental and Resource Economics Issue: 3 ISSN: 1573-1502.

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, "Strategic Environmental Policy and International Trade," *Journal of Public Economics*, 1994, *54*, 325–338.

Becker, Gary S. and George J Stigler, "Law enforcement, malfeasance, and compensation of enforcers," *The Journal of Legal Studies*, 1974, *3* (1), 1–18.

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.

Bó, Ernesto Dal, Pedro Dal Bó, and Erik Eyster, "The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects," *Review of Economic Studies*, 2018, *85*, 964–998.

Böhringer, Christoph, Knut Einar, and Halvor Briseid, "Robust policies to mitigate carbon leakage," *Journal of Public Economics*, 2017, *149*, 35–46. Publisher: Elsevier B.V.

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. ISBN: 1115101293.

Brazil Steel Institute, "Brazilian Steel Industry Yearbook," Technical Report 2021.

Bushnell, James B., Carla Peterman, and Catherine Wolfram, "Local Solutions to Global Problems: Climate Change Policies and Regulatory Jurisdiction," *Review of Environmental Economics and Policy*, 2008, *2* (2), 175–193.

_ **, Yihsu Chen, and Matthew Zaragoza-watkins**, "Downstream Regulation of CO2 Emissions in California' s Electricity Sector," *Energy Policy*, 2014, *64*, 313–323. Publisher: Elsevier.

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

Cicala, Steve, "Comments and discussion: Welfare and distributional implications of shale gas," *Brookings Papers on Economic Activity*, 2015, *Spring* (1), 126–139.

Collard-Wexler, Allan and Jan De Loecker, "Reallocation and technology: Evidence from the US steel industry," *American Economic Review*, 2015, *105* (1), 131–171.

Condon, Madison and Ada Ignaciuk, "Border Carbon Adjustment and International Trade: A Literature Review," Technical Report, OECD Trade and Environment Working Papers 2013.

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

Cosbey, Aaron, Susanne Droege, Carolyn Fischer, Julia Reinaud, John Stephenson, Lutz Weischer, and Peter Wooders, "A Guide for the Concerned: Guidance on the elaboration and implementation of border carbon adjustment," Technical Report, Entwined Policy Report 2012. Issue: November.

Council of the European Union, "Draft regulation of the European Parliament and of the Council establishing a carbon border adjustment mechanism," Technical Report 2021/0214(COD) March 2022.

Cropper, Maureen L. and Wallace E. Oates, "Environmental Economics : A Survey," *Journal of Economic Literature*, 1992, *30* (2), 675–740.

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," Technical Report 2010. Publication Title: Journal of Economic Literature Volume: 48 ISBN: 00220515 ISSN: 0022-0515.

Eckett, Sam, "EMEA Freight Markets," Technical Report, SP Global Platts 2021.

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 Perez Cervantes, and David Weisbach, "Trade and carbon taxes," *American Economic Review*, 2010, *100* (2), 465–469. ISBN: 0002-8282.

Energy Information Administration, "Trends in U.S. Oil and Natural Gas Upstream Costs," Technical Report 2016.

_ , "Drilling Productivity Report," Technical Report 2019.

_ , "Spot Prices for Petroleum and Other Liquids," 2021.

Farrokhi, Farid and Ahmad Lashkaripour, "Can Trade Policy Mitigate Climate Change?," Technical Report 2021.

Fernandez, Viviana, "Price and income elasticity of demand for mineral commodities," *Resources Policy*, 2018, *59* (1170037), 160–183. Publisher: Elsevier Ltd.

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," Technical Report 2016. Issue: February.

Fowlie, Meredith, Claire Petersen, and Mar Reguant, "Border Carbon Adjustments When Carbon Intensity Varies across Producers: Evidence from California," *AEA Papers and Proceedings*, 2021, *111*, 401–405.

Fullerton, Don, Inkee Hong, and Gilbert E. Metcalf, "A Tax on Output of the Polluting Industry Is Not a Tax on Pollution: The Importance of Hitting the Target," in Carlo Carraro and Gilbert E. Metcalf, eds., *Behavioral and Distributional Effects of Environmental Policy*, University of Chicago Press, 2000, pp. 13–44.

Gallaher, Michael P. and Brooks M. Depro, "Economic Impact Analysis of Final Integrated Iron and Steel NESHAP," Technical Report, US EPA September 2002. Issue: September.

Gerarden, Todd D., W Spencer Reeder, and James H Stock, "Federal Coal Program Reform, the Clean Power Plan, and the Interaction of Upstream and Downstream Climate Policies," *American Economic Journal: Economic Policy*, 2020, *12* (1), 167–199.

Gillingham, Kenneth, James B. Bushnell, Meredith L. Fowlie, Michael Greenstone, Charles Kolstad, Alan Krupnick, Adele Morris, Richard Schmalensee, and James Stock, "Reforming the U.S. coal leasing program," *Science*, 2016, *354* (6316), 1096–1098.

Glaeser, Edward L. and Andrei Shleifer, "A reason for quantity regulation," *American Economic Review*, 2001, *91* (2), 431–435.

González, Patrick, "Certification as a Rationale for Voluntary Agreements," 2011.

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.

Hasanbeigi, Ali and Cecilia Springer, "How Clean Is the U.S. Steel Industry? An International Benchmarking of Energy and CO2 Intensities," Technical Report, Global Efficiency Intelligence November 2019. Issue: November.

Hausman, Catherine H. and Ryan Kellogg, "Welfare and Distributional Implications of Shale Gas," *Brookings Papers on Economic Activity*, 2015. Publisher: NBER.

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

Holzer, Jorge, "Property Rights and Choice: The Case of the Fishery," *American Journal of Agricultural Economics*, 2015, *97* (4), 1175–1191.

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

Hyne, Norman J., *Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production*, Tulsa, OK: PennWell Corporation, 2001.

Interagency Working Group on Social Cost of Greenhouse Gases, "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990," Technical Report 2021. Issue: February.

International Energy Agency, "Iron and Steel Technology Roadmap Towards more sustainable steelmaking," Technical Report 2020.

Jacobs, Trent, "Life After Five," *Journal of Petroleum Technology*, January 2020.

Jacobsen, Mark R., Christopher R. Knittel, James M. Sallee, and Arthur A. Van Benthem, "The use of regression statistics to analyze imperfect pricing policies," *Journal of Political Economy*, 2020, *128* (5), 1826–1876.

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. ISBN: 00335533.

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

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

Keen, Michael and Christos Kotsogiannis, "Coordinating climate and trade policies: Pareto efficiency and the role of border tax adjustments," *Journal of International Economics*, September 2014, *94* (1), 119–128. Publisher: Elsevier.

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.

Kortum, Samuel and David A Weisbach, "Optimal Unilateral Carbon Policy," 2020.

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

Lade, Gabriel E and Ivan Rudik, "Costs of inefficient regulation: Evidence from the Bakken," *Journal of Environmental Economics and Management*, 2020, *102*, 102336. Publisher: Elsevier Ltd.

Laffont, Jean-Jacques, "Regulation of Pollution with Asymmetric Information," in "Nonpoint Source Pollution Regulation: Issues and Analysis," Netherlands: Springer, 1994, 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. Publisher: Elsevier B.V.

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

Lyubich, Eva, Joseph S. Shapiro, and Reed Walker, "Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Environmental Policy," *AEA Papers and Proceedings*, 2018, *108*, 136–142.

Marks, Levi, "The Abatement Cost of Methane Emissions from Natural Gas Production," *Journal of the Association of Environmental and Resource Economists*, 2022, *9* (2), 165–198.

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

McKinsey & Co., "Pathways to a Low-Carbon Economy," Technical Report 2009.

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. ISBN: 07416261.

Miller, Zeke and Josh Boak, "US, EU say deal on steel tariffs will help on climate change," *The Washington Post*, October 2021.

Newell, Richard G and Brian C Prest, "The unconventional oil supply boom: Aggregate price response from microdata," *Energy Journal*, 2019, *40* (3), 1–30.

— **,** — **, and Ashley B Vissing**, "Trophy hunting versus manufacturing energy: The price responsiveness of shale gas," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (2), 391–431.

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.

Petroleum Recovery Research Center, "GO-TECH: ONGARD Data by County," 2021.

Pinto, Raphael Guimaraes D., Alexandre S. Szklo, and Regis Rathmann, "CO2 emissions mitigation strategy in the Brazilian iron and steel sector—From structural to intensity effects," *Energy Policy*, 2018, *114* (November 2017), 380–393. Publisher: Elsevier Ltd.

Prest, Brian C and James H Stock, "Climate Royalty Surcharges," Technical Report 2021.

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

Robertson, Anna M, Rachel Edie, Robert A Field, David Lyon, Renee McVay, Mark Omara, Daniel Zavala-Araiza, and Shane M Murphy, "New Mexico Permian basin measured well pad methane emissions are a factor of 5–9 times higher than U.S. EPA estimates," *Environmental Science and Technology*, 2020, *54* (21), 13926–13934.

Schmutzler, Armin and Lawrence H. Goulder, "The Choice between Emission Taxes and Output Taxes under Imperfect Monitoring," *Journal of Environmental Economics and Management*, 1997, *32*, 51–64.

Schor, Adriana, "Heterogeneous productivity response to tariff reduction. Evidence from Brazilian manufacturing firms," *Journal of Development Economics*, 2004, *75* (2 SPEC. ISS.), 373–396.

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*, August 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.

Texas Railroad Commission, "Production Data Query Dump," 2021.

U.S. Environmental Protection Agency, "Oil and Natural Gas Sector: Emission Standards for New, Reconstructed, and Modified Sources Reconsideration," 2020. Publisher: Federal Register.

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

Whitehouse, Sheldon, "Methane Emissions Reduction Act of 2021," 2021.

World Steel Association, "Steel Statistical Yearbook 2020 Extended Version," Technical Report 2020.

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.

Zhang, Yuzhong, Ritesh Gautam, Sudhanshu Pandey, Mark Omara, Joannes D. Maasakkers, Pankaj Sadavarte, David Lyon, Hannah Nesser, Melissa P. Sulprizio, Daniel J. Varon, Ruix-

iong Zhang, Sander Houweling, Daniel Zavala-Araiza, Ramon A. Alvarez, Alba Lorente, Steven P. Hamburg, Ilse Aben, and Daniel J. Jacob, "Quantifying methane emissions from the largest oil-producing basin in the United States from space," *Science Advances*, 2020, *6* (17), 1–10.

Zimmerle, Daniel, Timothy Vaughn, Ben Luck, Terri Lauderdale, Kindal Keen, Matthew Harrison, Anthony Marchese, Anthony Marchese, Laurie Williams, and David Allen, "Methane Emissions from Gathering Compressor Stations in the U.S.," *Environmental Science and Technology*, 2020, *54* (12), 7552–7561.

A Theory Appendix: Domestic case

In the following we first present the model of the domestic setting with a fixed set of firms and an adjustment on the intensive margin. We proceed to the setting with an extensive margin in Section B.

A.1 Setup and proofs of uniqueness, Lemma 1 and Proposition 2.

We solve the model for the general case of certification at level $\hat{e}$ with the understanding that the case of no certification can be found at $\hat{e} = \underline{e}$. We note that the equilibrium is determined by the indifference condition on certification (4) with t defined by (5) together with the market clearing equation, which, using (7), can be written as:

$$D(p) = E(s(p - \tau\varepsilon)). \tag{40}$$

We label p^U the price without certification ($\hat{e} = \underline{e}$) and p^V the price with certification.

Welfare derived from the operation of this market is the sum of consumer utility, net of welfare costs from emissions, profits of firms, government revenue, and certification cost:

$$W = I - pC + u(D(p)) - vG + E\pi(p - \tau\varepsilon) + \tau G - F\Psi(\hat{e}), \tag{41}$$

where the representative agent has exogenous income I and emissions are given by equation (8).

A.1.1 Proof of uniqueness

Lemma. *The equilibrium is unique if i) s is weakly convex or τ is small and ii) $E\left(e|e > \hat{e}\right) - \hat{e}$ is decreasing in $\hat{e}$.*

To show this we define

$$g\left(\widehat{e}\right) \equiv \pi\left(p - \tau\widehat{e}\right) - \pi\left(p - \tau E\left(e|e > \hat{e}\right)\right), \tag{42}$$

where p is itself a function of $\widehat{e}$ through the market clearing equation (40). Thus we have

$$g'\left(\widehat{e}\right) = \left(s\left(p - \widehat{e}\tau\right) - s\left(p - \tau E\left(e|e > \hat{e}\right)\right)\right)\frac{dp}{d\widehat{e}} + \tau\left(\frac{dE\left(e|e > \hat{e}\right)}{d\widehat{e}} s\left(p - \tau E\left(e|e > \hat{e}\right)\right) - s\left(p - \tau\widehat{e}\right)\right)$$

Differentiating (40), one gets

$$\frac{dp}{d\widehat{e}}=\frac{\left(s\left(p-\tau\widehat{e}\right)-s\left(p-\tau E\left(e|e>\hat{e}\right)\right)\right)\psi\left(\widehat{e}\right)-\tau\frac{dE(e|e>\hat{e})}{d\widehat{e}}s'\left(p-\tau E\left(e|e>\hat{e}\right)\right)\left(1-\Psi\left(\widehat{e}\right)\right)}{D'\left(p\right)-\left(\int_0^{\widehat{e}}s'\left(p-\tau e\right)\psi\left(e\right)de+s'\left(p-\tau E\left(e|e>\hat{e}\right)\right)\left(1-\Psi\left(\widehat{e}\right)\right)\right)}$$

One further get that

$$\frac{dE\left(e|e>\hat{e}\right)}{d\widehat{e}}=\left(E\left(e|e>\hat{e}\right)-\widehat{e}\right)\frac{\psi\left(\widehat{e}\right)}{1-\Psi\left(\widehat{e}\right)}$$

Hence

$$\frac{dp}{d\widehat{e}}=-\psi\left(\widehat{e}\right)\frac{s\left(p-\tau\widehat{e}\right)-s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)-\tau E\left(e-\widehat{e}|e>\widehat{e}\right)s'\left(p-\tau E\left(e|e>\widehat{e}\right)\right)}{\int_0^{\widehat{e}}s'\left(p-\tau e\right)\psi\left(e\right)de+\left(1-\Psi\left(\widehat{e}\right)\right)s'\left(p-\tau E\left(e|e>\widehat{e}\right)\right)-D'\left(p\right)}. \tag{43}$$

If s is linear, p is constant. If s is convex then

$$\left(s\left(p-\tau\widehat{e}\right)-s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)\right)>\tau\left(E\left(e|e>\hat{e}\right)-\widehat{e}\right)s'\left(p-\tau E\left(e|e>\hat{e}\right)\right)$$

and $\frac{dp}{d\widehat{e}}<0$. On the other hand if s is concave $\frac{dp}{d\widehat{e}}>0$. We can then rewrite:

$$\begin{aligned}g'\left(\widehat{e}\right) &= \underset{<0 \text{ if s convex}}{\frac{dp}{d\widehat{e}}}\underbrace{s\left(p-\tau\widehat{e}\right)-s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)}_{>0}\\ &\quad+\tau\left(\underbrace{\left(\frac{dE\left(e|e>\hat{e}\right)}{d\widehat{e}}-1\right)}_{<0 \text{ if } E(e|e>\hat{e})-e \text{ decreasing}}s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)+\underbrace{s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)-s\left(p-\tau\widehat{e}\right)}_{<0}\right)\end{aligned}$$

Therefore if s is weakly convex and $E\left(e|e>\hat{e}\right)-e$ decreasing then there is a unique solution.

If τ is small, we get $\frac{dp}{d\widehat{e}}=o\left(\tau\right)$ so that

$$g'\left(\widehat{e}\right)=\tau\left(\left(\frac{dE(e|e>\hat{e})}{d\widehat{e}}-1\right)s\left(p-\tau E\left(e|e>\widehat{e}\right)\right)+s'\left(p\right)\tau\left(\widehat{e}-E\left(e|e>\widehat{e}\right)\right)\right)+\quad o\left(\tau^2\right),$$

which is negative if $E\left(e|e>\hat{e}\right)-e$ is decreasing.

Since firms certify as long as $g\left(e\right)>F+f$ and g is decreasing, there is a unique solution.

A.1.2 Proof of Lemma 1

Taking the difference between equations (8) and (2) gives equation (9). Since the no-certification case corresponds to $\hat{e} = \underline{e}$, we differentiate (8) with respect to $\widehat{e}$:

$$\frac{dG^V}{d\widehat{e}} = \psi(\widehat{e}) \underbrace{\left[\widehat{e}\left(s\left(p^V - \tau\widehat{e}\right) - s\left(p^V - \tau E\left(e|e > \widehat{e}\right)\right)\right) - \left(E\left(e|e > \widehat{e}\right) - \widehat{e}\right)\tau E\left(e|e > \widehat{e}\right) s'\left(p^V - \tau E\left(e|e > \widehat{e}\right)\right)\right]}_{\equiv A}$$

$$+ \underbrace{E\left(\varepsilon s'\left(p^V - \tau\varepsilon\right)\right)\frac{dp^V}{d\widehat{e}}}_{\equiv B}$$

Defining $h(e) \equiv es(p^V - \tau e)$, we note that the first term A can be rewritten as:

$$A = h(\widehat{e}) - h\left(E\left(e|e > \widehat{e}\right)\right) - \left(\widehat{e} - E\left(e|e > \widehat{e}\right)\right) h'\left(E\left(e|e > \widehat{e}\right)\right).$$

If h is concave in e then this term must be negative; in addition, if s is weakly convex we know that $\frac{dp^V}{d\widehat{e}} \leq 0$ so that the term B is negative. Therefore if both conditions are met $\frac{dG^V}{d\widehat{e}} < 0$ which implies that $G^V < G^U$ as stated in Lemma 1.

For a small τ, we have derived that $\frac{dp}{d\widehat{e}} = o(\tau)$, in addition get:

$$A = -\tau\left(E\left(e|e > \widehat{e}\right) - \widehat{e}\right)^2 s'(p_0) + o(\tau),$$

where p_0 is the laissez-faire price. This expression is negative for small τ, ensuring that in that case as well, $\frac{dG^V}{d\widehat{e}} < 0$ so that $G^V < G^U$.

A.1.3 Proof of Proposition 2

Using equation (41), we can write the welfare change following certification

$$\begin{aligned} & W^V - W^U \qquad (44) \\ = \; & u(D(p^V)) - u(D(p^U)) + \left[\pi(p^V - \tau E(e)) - \pi(p^U - \tau E(e))\right] - \left[p^V C^V - p^U C^U\right] \\ & -(v - \tau)\left[G^V - G^U\right] + E\pi(p^V - \tau\varepsilon) - \pi\left(p^V - \tau E(e)\right) - F\Psi(\hat{e}) \\ = \; & \underbrace{\int_{p^U}^{p^V} \left[s(p - \tau E(e)) - D(p)\right] dp}_{\text{Price Effect}} - \underbrace{(v - \tau)(G^V - G^U)}_{\text{Untaxed emissions effect}} + \underbrace{E\pi(p^V - \tau\varepsilon) - \pi\left(p^V - \tau E(e)\right)}_{\text{Reallocation Effect}} - F\Psi(\hat{e}) \end{aligned}$$

where we use $\partial\pi/\partial p = s$ and the first order condition of consumers demand: $u' = p$. We now sign each term in turn:

Price effect. The price effect is zero if $p^V = p^U$. Note that $s(p - \tau E(e)) - D(p)$ is increasing in p and takes the value zero at $p = p^U$ by definition of the equilibrium in the uncertified case. Therefore, $s(p - \tau E(e)) - D(p) > 0$ is positive if $p^U < p^V$ and the price effect itself is positive. Conversely, if $p^U > p^V$ then $s(p - \tau E(e)) - D(p) < 0$ for all $p \in [p^V, p^U)$, but since the lower bound is then larger than the upper bound, the integral is still positive.

Reallocation effect. The profit function is convex in ε. Using Jensen's inequality, we then get that the reallocation effect is positive.

A.1.4 Figure for concave supply functions

Figure 1 presents the equilibrium with linear or convex supply curves. Here we present an analogous figure for concave supply function. The $S^V(p)$ curve continues to intersect at a lower point than $S^U(p)$ since the least polluting firm will produce at a lower price under certification than without. It will, however, intersect the $S^U(p)$ such that, when all firms produce, $S^V(p) < S^U(p)$ and therefore $p^V > p^U$. Consider the case of $\tau = v$. Then the welfare effect from moving from $S^U(p)$ to $S^V(p)$ is $C - (B + E)$. The reallocation effect counts profits under p^V and the difference between the two is therefore $(C - B - E - D)$. This "over counts" the overall welfare change by $-D$ which the price effect, D, corrects for. It is apparent that D is always positive and the analysis above demonstrates that the reallocation effect is as well. The term A is now a reallocation from consumers to producers.

A.2 Proof of Corollary 3

We derive equations (11) and (12) using first order approximations in τ around $\tau = 0$. We also consider v to be of the same order. As in Section, p_0 is the equilibrium price in laissez-faire (i.e. for $\tau = 0$).

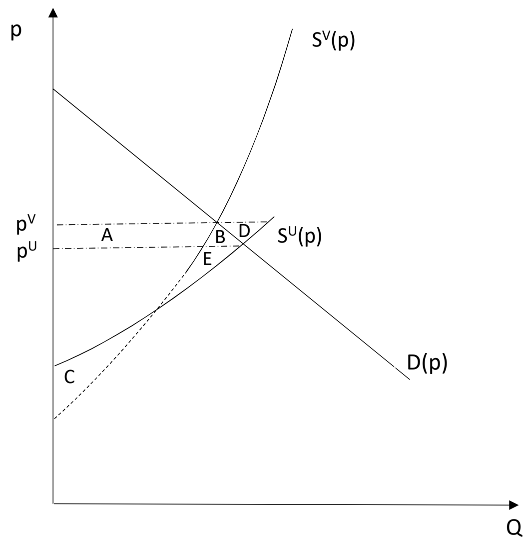
Taylor expansions of equilibrium prices. One can write the market clearing conditions in the output tax case and in the laissez-faire case:

$$s\left(p^U - \tau E\left(e\right)\right) = D\left(p^U\right) \text{ and } s\left(p_0\right) = D\left(p_0\right).$$

Taking a Taylor expansion of the difference between these two expressions gives that the equilibrium price change satisfies:

$$p^U - p_0 = \frac{E\left(e\right) s'\left(p_0\right)}{s'\left(p_0\right) - D'\left(p_0\right)}\tau + o\left(\tau\right). \tag{45}$$

Figure A.1: Equilibrium with concave supply functions.



Note: Concave supply functions implies that, when all firms produce, the supply is lower under certification than without. The sum of the reallocation and price effects is then $C-(B+E)$. A represents a reallocation from consumers to producers.

We take a similar approach with p^V using equation (40) which we write explicitly as:

$$\int_{\underline{e}}^{\widehat{e}} s\left(p^V-\tau e\right)\psi\left(e\right)de+\left(1-\Psi\left(\widehat{e}\right)\right)s\left(p^V-\tau E\left(e|e>\widehat{e}\right)\right)=D\left(p^V\right),$$

such that:

$$\begin{array}{c}\int_{\underline{e}}^{\widehat{e}}\left(s\left(p_0\right)+s'\left(p_0\right)\left(p^V-\tau e-p_0\right)\right)\psi\left(e\right)de+\\ \left(1-\Psi\left(\widehat{e}\right)\right)\left(s\left(p_0\right)+s'\left(p_0\right)\left(p^V-\tau E\left(e|e>\widehat{e}\right)-p_0\right)\right)\end{array}=D\left(p_0\right)+D'\left(p_0\right)\left(p^V-p_0\right)+o\left(\tau\right)$$

$$\Leftrightarrow s'\left(p_0\right)\left(p^V-\tau E\left(e\right)-p_0\right)=D'\left(p_0\right)\left(p^V-p_0\right)+o\left(\tau\right),$$

which, when reordered returns equation (45) (with p^Vreplacing p^U) such that we establish that, to a first order, prices are the same under certification and no certification:

$$p^V=p^U+o(\tau) \tag{46}$$

Taylor expansion of the welfare change W^V-W^U. We take Taylor expansions of each element of equation (44) at the second order (as we will see that $W^V-W^U+F\Psi(\hat{e})$ is in fact second order in τ). We start with the price effect and recall that $p^V - p^U = o(\tau)$

$$\begin{aligned}
& \int_{p^U}^{p^V} [s(p - \tau E(e)) - D(p)]\, dp \\
= & \int_{p^U}^{p^V} \left(\left(s'(p_0)(p - \tau E(e) - p_0) - D'(p_0)(p - p_0)\right) + o(\tau)\right) dp + o(\tau^2) \\
= & \left(s'(p_0) - D'(p_0)\right) \frac{\left(p^V - p_0\right)^2 - \left(p^U - p_0\right)^2}{2} - \tau E(e)\, s'(p_0)\left(p^V - p^U\right) + o\left(\tau^2\right) \\
= & \; o\left(\tau^2\right), \qquad (47)
\end{aligned}$$

such that the price effect is zero at second order.

We proceed with the reallocation effect and note that the integral here is zeroth order and develop profits correspondingly (at second order).

$$\begin{aligned}
& E\left(\pi\left(p^V - \tau\varepsilon\right)\right) - \pi\left(p^V - \tau E(e)\right) \\
= & \int_{\underline{e}}^{\widehat{e}} \begin{pmatrix} \pi(p_0) + s(p_0)\left(p^V - \tau e - p_0\right) \\ +s'(p_0)\frac{\left(p^V - \tau e - p_0\right)^2}{2} \end{pmatrix} \psi(e)\, de \\
& + (1 - \Psi(\widehat{e})) \begin{pmatrix} \pi(p_0) + s(p_0)\left(p^V - \tau E(e|e > \widehat{e}) - p_0\right) \\ +s'(p_0)\frac{\left(p^V - \tau E(e|e > \widehat{e}) - p_0\right)^2}{2} \end{pmatrix} - \begin{pmatrix} \pi(p_0) + s(p_0)\left(p^V - \tau E(e) - p_0\right) \\ +s'(p_0)\frac{\left(p^V - \tau E(e) - p_0\right)^2}{2} \end{pmatrix} + o\left(\tau^2\right) \\
= & \left(\int_{\underline{e}}^{\widehat{e}} \left(p^V - \tau e - p_0\right)^2 \psi(e)\, de + (1 - \Psi(\widehat{e}))\left(p^V - \tau E(e|e > \widehat{e}) - p_0\right)^2 - \left(p^V - \tau E(e) - p_0\right)^2\right) \frac{s'(p_0)}{2} + o\left(\tau^2\right) \\
= & \left(\int_{\underline{e}}^{\widehat{e}} \tau^2 e^2 \psi(e)\, de + (1 - \Psi(\widehat{e}))\, \tau^2 E(e|e > \widehat{e})^2 - \tau^2 E(e)^2\right) \frac{s'(p_0)}{2} + o\left(\tau^2\right) \\
= & \frac{s'(p_0)}{2} \tau^2 Var(\varepsilon) + o\left(\tau^2\right) > 0. \qquad (48)
\end{aligned}$$

We proceed with the change in emissions

$$\begin{aligned}
G^V - G^U &= E\left(\varepsilon s\left(p^V - \tau\varepsilon\right)\right) - E(e)\, s\left(p^V - \tau E(e)\right) \qquad (49) \\
&= s'(p_0)\,\tau \left(E(e)^2 - \int_{\underline{e}}^{\widehat{e}} e^2 \psi(e)\, de - (1 - \Psi(\widehat{e}))\, E(e|e > \widehat{e})^2\right) + o\left(\tau^2\right) \\
&= -s'(p_0)\,\tau Var(\varepsilon) + o(\tau),
\end{aligned}$$

which is equation (12). Noting that our approach supposes that v is of the same order of τ, we can then write the untaxed emissions effect as:

$$-(v - \tau)\left(G^V - G^U\right) = (v - \tau)\, s'(p_0)\,\tau Var(\varepsilon) + o\left(\tau^2\right). \qquad (50)$$

We combine (47), (48) and (50), and add the certification costs to get equation (11), which establishes that the expression $W^V - W^U + F\Psi(\hat{e})$ is indeed second order.

Proof of statements a)-c) in Corollary 3: Parts a) and b) directly follow from the fact that $Var(\varepsilon)$ is increasing in $\widehat{e}$ (as long as distribution of ε is not degenerative). We have already established that $p^V = p^U$ when supply is linear, so that the price effect is null. We note that in the linear case $s'' = 0$ so that (48) and (50) hold exactly. This implies Part c).

A.3 Optimal policy

We solve for the optimal allocation. The social planner solves the problem given in equation (14). The first-order condition with respect to $q^V(e)$ is given by

$$u'(C) - c'(q^V(e)) - ev = 0, \tag{51}$$

where market clearing gives $C = \int_{\underline{e}}^{\widehat{e}} q^V(e)\,\psi(e)\,de + q^U(1 - \Psi(\widehat{e}))$. The first-order condition with respect to q^U is given by:

$$u'(C) - c'(q^U) - vE(e|e > \hat{e}) = 0. \tag{52}$$

Finally, the first-order condition with respect to $\widehat{e}$ leads to:

$$u'(C)\left(q^V(\widehat{e}) - q^U\right) - \left(c\left(q^V(\widehat{e})\right) - c\left(q^U\right)\right) - \tau\widehat{e}\left(q^V(\widehat{e}) - q^u\right) - F = 0. \tag{53}$$

This allocation can be decentralized using an emission tax $\tau = v$ for certified firms and an output tax $t = \tau E(e|e > \hat{e})$ for uncertified firms, together with a tax on certification f. Indeed, under these policy instruments, we get that the consumer price $p = u'(C)$ and we recover from (51) that $q^V(e) = s(p - \tau e)$ and from (52) that $q^U = s(p - \tau E(e|e > \hat{e}))$. We can then rewrite (53) as (16) such that the tax on certification has to be given by (15). We can then directly check that the IC constraint is satisfied for all emission rates e.

A.4 Proof of Proposition 4

We prove Proposition 4, first without and then with a certification tax.

Case with no certification tax. $\widehat{e}_n$ is defined according to

$$\pi(p - \tau\widehat{e}_n) - F = \pi(p - \tau E(e|e > \widehat{e}_{n-1})), \tag{54}$$

which we can rewrite as

$$\pi\left(p-\tau\widehat{e}_n\right)-\pi\left(p-\tau\widehat{e}_{n-1}\right)=F-g\left(\widehat{e}_{n-1}\right)$$

with g defined as in (42). For p exogenous and $E\left(e|e>\widetilde{e}\right)-\widetilde{e}$ decreasing, we have established that g is decreasing. Further for the equilibrium $\widehat{e}$, we get $g\left(\widehat{e}\right)=F$. Therefore if $\widehat{e}_{n-1}<\widehat{e}$, we have

$$\pi\left(p-\tau\widehat{e}_n\right)-\pi\left(p-\tau\widehat{e}_{n-1}\right)<0,$$

which ensures that $\widehat{e}_n>\widehat{e}_{n-1}$.

Let us define the function

$$h\left(\widetilde{e},\overrightarrow{e}\right)\equiv\pi\left(p-\tau\overrightarrow{e}\right)-\pi\left(p-\tau E\left(e|e>\widetilde{e}\right)\right).$$

Note that h is decreasing in $\overrightarrow{e}$ and increasing in $\widetilde{e}$. Therefore,

$$h\left(\widehat{e}_{n-1},\widehat{e}\right)<h\left(\widehat{e},\widehat{e}\right)=F=h\left(\widehat{e}_{n-1},\widehat{e}_n\right),$$

so that we must have $\widehat{e}_n<\widehat{e}$.

To summarize, we have that if $\widehat{e}_{n-1}<\widehat{e}$, then $\widehat{e}_{n-1}<\widehat{e}_n<\widehat{e}$. Given that $\widehat{e}_0<\widehat{e}$, we get that by induction $\widehat{e}_n$ must converge monotonically towards the fixed point of (54), namely $\widehat{e}$.

Case with a certification tax. In the certification case with $f_n=\tau\left(E\left(e|e>\widehat{e}_{n-1}\right)-\widehat{e}_{n-1}\right)s\left(p-\tau E\left(e|e>\widehat{e}_{n-1}\right)\right)$, (54) is replaced by:

$$\begin{aligned}&\pi\left(p-\tau\widehat{e}_n\right)-F-\tau\left(E\left(e|e>\widehat{e}_{n-1}\right)-\widehat{e}_{n-1}\right)s\left(p-\tau E\left(e|e>\widehat{e}_{n-1}\right)\right)\\=\;&\pi\left(p-\tau E\left(e|e>\widehat{e}_{n-1}\right)\right).\end{aligned}\tag{55}$$

The fixed point of this equation is the threshold under the social planner problem, which we will denote here $\widehat{e}^*$. We can rewrite (55) as

$$\pi\left(p-\tau\widehat{e}_n\right)-\pi\left(p-\tau\widehat{e}_{n-1}\right)=F-g^*\left(\widehat{e}_{n-1}\right),$$

where

$$g^*\equiv\pi\left(p-\tau\widetilde{e}\right)-\pi\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)-\tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right).$$

We get

$$g^{*\prime}(\widetilde{e}) \quad = \quad -\tau\left(s\left(p-\tau\widetilde{e}\right)-s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)+\tau\frac{dE\left(e|e>\widetilde{e}\right)}{d\widetilde{e}}\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)\right).$$

If s is weakly convex then $s\left(p-\tau\widetilde{e}\right)-s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)\geq\tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)$, so if $E\left(e|e>\widetilde{e}\right)-\widetilde{e}$ is decreasing in $\widetilde{e}$ (which implies that $\frac{dE(e|e>\widetilde{e})}{d\widetilde{e}}<1$) then $g^{*\prime}(\widetilde{e})<0$.

For small τ, we get:

$$g^{*\prime}(\widetilde{e})=\tau^2\left(\frac{dE\left(e|e>\widetilde{e}\right)}{d\widetilde{e}}-1\right)\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p_0\right)+o\left(\tau^2\right),$$

which again will be negative if $E\left(e|e>\widetilde{e}\right)-\widetilde{e}$ is decreasing in $\widetilde{e}$.

Under these assumptions, we therefore get that if $\widehat{e}_{n-1}<\widehat{e}^*$ then $\widehat{e}_n>\widehat{e}_{n-1}$. We can then define

$$h^*\left(\widetilde{e},\overrightarrow{e}\right)\equiv\pi\left(p-\tau\overrightarrow{e}\right)-\pi\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)-\tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right).$$

h^* is decreasing in $\overrightarrow{e}$ and increasing in $\widetilde{e}$ and $h^*\left(\widehat{e}_{n-1},\widehat{e}_n\right)=F=h^*\left(\widehat{e}^*,\widehat{e}^*\right)$. As above this ensures that if $\widehat{e}_{n-1}<\widehat{e}^*$ then $\widehat{e}_{n-1}<\widehat{e}_n<\widehat{e}^*$, which in return implies that $\widehat{e}_n$ converges monotonically toward $\widehat{e}^*$.

A.5 Extensions

We consider in turn three extensions: abatement, heterogeneity in TFP and free entry. The first extension is discussed in the text in Section 2.6.

A.5.1 Abatement

In this section we give additional details on the case with abatement and formally prove Corollary 5.

With abatement, the supply of a certified firm with emissions rate e is given by $q=s\left(p-\tau e+A(\tau)\right)$,where $A\left(\tau\right)\equiv\tau a^*(\tau)-b(a^*(\tau))$, so that production is higher when firms abate. We can then write the profits of certifying firms as $\pi(p-\tau e+A(\tau))$. As a result, the certification threshold $\hat{e}$ is now implicitly defined by:

$$\pi(p-\tau\hat{e}+A(\tau))-F-f=\pi(p-\tau E(e|e>\hat{e})).$$

Though such threshold may not exist as it is now possible that all firms wish to certify even for $F+f>0$.

Difference in emissions. We can then write changes in emissions as:

$$G^V - G^U = \left(G^V - G^U\right)|_{\text{no abatement}} - a^*\Psi(\hat{e})E\left[s\left(p^V - \tau e + A(\tau)\right)|e < \hat{e}\right] + \Psi(\hat{e})E\left\{e\left[s\left(p^V - \tau e + A(\tau)\right) - s\left(p^V - \tau e\right)\right]|e \leq \hat{e}\right\}, \tag{56}$$

where $\left(G^V - G^U\right)|_{\text{no abatement}}$ corresponds to the change in emissions without abatement given in Lemma 1 (except that the price p^V now represents the equilibrium price under certification with abatement). Abatement therefore adds two terms. The first new term in expression (56) is the direct impact of a mass $\Psi(\hat{e})$ certifying and thereby abating 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 additional effect pulls in the direction of higher emissions, hence the second new term. However, emissions decrease if s is weakly convex, $es(p^V - \tau e)$ is increasing and concave in e and $\underline{e} > a^*$.[35] As stipulated in Corollary 5 and proved below emissions must also decline if τ is small.

Difference in welfare. We use equation (41) and follow similar steps as in Section A.1.3, and obtain that the welfare change from certification can be written as:

$$W^V - W^U = \left(W^V - W^U\right)|_{\text{no abatement}} + \Psi\left(\widehat{e}\right)\left(E\left(\pi\left(p^V - \tau e + A(\tau)\right) - \pi\left(p^V - \tau e\right)|e \leq \widehat{e}\right)\right), \tag{57}$$

were $\left(W^V - W^U\right)|_{\text{no abatement}}$ corresponds to the expression given in Proposition 2 for the no-abatement case.[36] The new term captures the increase in profits for firms that receive a higher net price after abatement. This benefit accrues only to the share of firms $\Psi(\hat{e})$ that certify. For given price p^V, profit maximization ensures that this term is positive.

Proof of Corollary 5. We follow an approach analogous to Section A.2. We start out by taking a Taylor expansion of the optimal level of abatement $b'(a) = \tau$ to get $b'(0) + ab''(a) = \tau + o(\tau)$. Using that $b'(0) = 0$, we write:

$$a^* = \frac{\tau}{b''(0)} + o(\tau), \tag{58}$$

[35] If s is weakly convex in p then without abatement and for given prices, certification would lead to an increase in production. Given that certified firms increase their production further with abatement, this is still true in this model. As a result, the price decreases, $p^V < p^U$. With $es(p^V - \tau e)$ concave in e, we can then follow the same steps as in Section A.1.2 and get that $\left(G^V - G^U\right)|_{\text{no abatement}}$ is negative. In addition, we have that $(e - a^*)\,s\left(p^V - \tau e + A(\tau)\right) < (e - a^*)\,s\left(p^V - \tau\left(e - a^*\right)\right)$ (as long as $e > a^*$), therefore with $es(p^V - \tau e)$ increasing in e, the sum of the abatement terms is negative.

[36] The price p^V changes with abatement, and as we have seen, the change in emissions $G^V - G^U$ as well. In particular, there are additional welfare gains relative to a world without abatement through a larger untaxed emissions effect if $v > \tau$ and emissions decrease.

which implies that:

$$A(\tau) = \tau a^* - b(a^*) = \frac{1}{2}\frac{1}{b''(0)}\tau^2 + o(\tau^2). \tag{59}$$

Since $A(\tau)$ is second order, the amount produced by certified firms does not change at first order, i.e. $s\left(p - \tau e + A(\tau)\right) = s\left(p - \tau e\right) + o\left(\tau\right)$. We therefore recover the result that certification does not change prices at first order: equations (45) and (46) continue to hold. The threshold $\widehat{e}$ also remains the same at first order. We therefore recover the same expressions as before for the price effect (47) and the reallocation effect (48). The abatement effect can be written as:

$$\begin{aligned} & \Psi\left(\widehat{e}\right)\left(E\left(\pi\left(p^V - \tau e + A(\tau)\right) - \pi\left(p^V - \tau e\right)|e < \widehat{e}\right)\right) \\ = & \int_{\underline{e}}^{\widehat{e}} \left(\begin{array}{c} s\left(p_0\right)\left(\left(p^V - \tau e + A(\tau) - p_0\right) - \left(p^V - \tau e - p_0\right)\right) \\ +\frac{s'(p_0)}{2}\left(\left(p^V - \tau e + A(\tau) - p_0\right)^2 - \left(p^V - \tau e - p_0\right)^2\right) + o\left(\tau^2\right) \end{array} \right) de \\ = & \; s\left(p_0\right)\frac{\tau^2}{2b''\left(0\right)}\Psi\left(\widehat{e}\right) + o\left(\tau^2\right). \end{aligned} \tag{60}$$

Using (56) and that (49) applies for $\left(G^V - G^U\right)|_{\text{no abatement}}$, we can write the change in emissions as:

$$G^V - G^U = -s'\left(p_0\right)\tau Var\left(\varepsilon\right) + \int_{\underline{e}}^{\widehat{e}} \left[e\left(s'\left(p_0\right)\left(\tau a^* - b\left(a^*\right)\right)\right) - a^* s\left(p_0\right)\right]\psi\left(e\right)de + o\left(\tau\right).$$

We then use (58) and (59) to obtain (17).

Plugging (47), (48), (60) and (17) into (57) gives the welfare change at second order in v, τ, equation (18).

Proportional abatement. We now assume that by spending $b\left(a\right)$ per unit of output a firm can reduce its emission rate by a factor a (i.e. from e to $e\left(1-a\right)$). Certifying firms therefore solve the problem

$$max_{q,a} pq - c(q) - \tau e(1-a)q - b(a)q,$$

so that they choose an abatement rate which depends on the emission rate: $a^*\left(e\right) = b'^{-1}(\tau e)$. The effective price per unit received by a certifying firm, $p\left(e\right) = p - \tau e\left(1-a\right) - b\left(a\right)$, is still decreasing in e. Therefore profits are decreasing in e and certification still occurs on an interval of the type $[\underline{e}, \hat{e}]$.

For small τ, we then obtain that the abatement rate is proportional to the initial emission rate: $a^*(e) = \tau e / b''(0) + 0(\tau)$, and the firm's savings per unit of output can be written as $A(\tau, e) \equiv \tau e a - b(a) = (\tau e)^2 / (2b''(0))$. Equation (56) becomes

$$G^V - G^U = \left(G^V - G^U\right)|_{\text{no abatement}} + \Psi(\hat{e}) E\left\{e\left(1 - a^*(e)\right) s\left(p^V - \tau e + A(\tau, e)\right) - es\left(p^V - \tau e\right) | e \leq \hat{e}\right\},$$

while equation (57) is still valid provided that A is adjusted accordingly. Taking Taylor expansions as before, we then obtain

$$G^V - G^U = -s'(p_0)\,\tau Var(\varepsilon) - \Psi(\hat{e}) \frac{\tau s(p_0)}{b''(0)} E\left(e^2 | e \leq \hat{e}\right) + o(\tau),$$

and

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right)\left[s'(p_0) Var(\varepsilon) + \Psi(\hat{e}) \frac{s(p_0)}{b''(0)} E\left(e^2 | e \leq \hat{e}\right)\right] + o\left(\tau^2\right).$$

These adjustments directly reflect that on average certifying firms now abate $\tau E\left(e^2 | e \leq \hat{e}\right) / b''(0)$ per unit.

A.5.2 Heterogeneous Productivity

We now let firms have heterogeneous productivity levels (and for simplicity here preclude abatement). We assume that firm i has costs of production of $c(q)/\varphi_i$, where $c(q)$ has the same properties as before but $\varphi_i > 0$ differs across firms. We let $\Psi(e, \varphi)$ denote the joint distribution and allow unrestricted covariance between e and φ. We consider supply functions with constant elasticity such that a firm i that certifies produces $q_i = s(\varphi_i(p - \tau e))$ with $s(x) \equiv s_0 x^\alpha$ where $\alpha > 0$ is the common elasticity of production with respect to prices. The uncertified firms produce $q^u = s(\varphi_i(p - t))$, where t is τ times the production-weighted average emission rate of uncertified firms. This is also the optimal quantity tax rate under no certification when the government does not observe total revenues (otherwise it could infer some information on the emission rate from the correlation between φ and e).[37]

The profits of a certified firm (gross of certification costs) can then be written as $\frac{1}{\varphi}\pi(\varphi(p - \tau e))$ and the profits of an uncertified firm as $\frac{1}{\varphi}\pi(\varphi(p - t))$ where π is the profit function previously defined. In this setup, firms differ both in their productivity and their emissions. Equation (4) is then replaced by:

$$\frac{1}{\varphi}\pi\left(\varphi\left(p - \tau \widehat{e}(\varphi)\right)\right) - \frac{1}{\varphi}\pi\left(\varphi(p - t)\right) = F + f,$$

[37] Our approach could also be used with other supply functions but the analysis becomes more complicated, in particular because t is no longer the optimal quantity tax.

which defines $\hat{e}(\varphi)$. Firms with an emission rate $e < \hat{e}(\varphi)$ certify while other firms do not. The cut-off function $\hat{e}(\varphi)$ depends positively on productivity because production increases with productivity whereas the certification cost, $F + f$, does not.

The Taylor expansions for the changes in emissions and welfare of Corollary 3 take the same form, except $Var(\varepsilon)$ is replaced by the output-weighted variance of emissions:

$$\tilde{Var(\varepsilon)} = \int_{\varphi} \int_{e} (\varepsilon - \tilde{E}(\varepsilon))^2 \tilde{\psi}(\varphi, e) de d\varphi, \tag{61}$$

where $\tilde{\psi}(\varphi, e) = \varphi^{\alpha}\psi(\varphi, e) / \left(\int_{\varphi} \int_{e} \varphi^{\alpha}\psi(\varphi, e) de d\varphi \right)$ is a density distribution rescaled by output (proportional to φ^{α} at price p_0) such that $\tilde{E}(\varepsilon)$ equals the average emissions per unit without certification:

$$\tilde{E}(\varepsilon) = \int_{\varphi} \int_{e} \varepsilon \tilde{\psi}(\varphi, e) de d\varphi = G^U / S^U.$$

Intuitively, the output reallocation effect is still the driving force behind our results, though firms' emissions are now weighted by their size.

Therefore, the simple expressions we derived previously are still valid approximations of the emissions and welfare changes brought about by certification as long as production is close to isoelastic.

Proof. We show that Corollary 3 still applies with quantity-weighted distribution of emissions for the model with heterogenous productivity level described in Section 2.6.

We can write aggregate output as:

$$S^V(p) = \int_{\varphi} \left(\int_{\underline{e}}^{\widehat{e}(\varphi)} s_0 \left[\varphi \left(p - \tau e \right) \right]^{\alpha} \psi \left(e | \varphi \right) de + s_0 \left[\varphi \left(p - t^V \right) \right]^{\alpha} \left(1 - \Psi \left(\widehat{e}(\varphi) \right) \right) \right) \psi_{\varphi}(\varphi) d\varphi. \tag{62}$$

The output tax on uncertified firm t is defined by total emissions of uncertified firms divided by total output of uncertified firms. We denote by $\psi_{\varphi}(\varphi)$ the (unconditional) distribution of productivity and by $\psi_e(e|\varphi)$ the distribution of emissions conditional on productivity. We can then write:

$$t = \tau \frac{\int_{\varphi} \psi_{\varphi}(\varphi) \int_{\widehat{e}_{\varphi}}^{\infty} es\left(\varphi\left(p - t\right)\right) \psi_e\left(e|\varphi\right) de d\varphi}{\int_{\varphi} \psi_{\varphi}(\varphi) s\left(\varphi\left(p - t\right)\right) \left(1 - \Psi_e\left(\widehat{e}(\varphi)|\varphi\right)\right) d\varphi}. \tag{63}$$

We let t^U denote the tax when no certification is in place ($\hat{e}(\varphi) = \underline{e}$ for all φ) and t^V the tax on uncertified firms when certification is in place.

Welfare is still given by (41), so that we can follow steps the same steps as in (41) and write the welfare change as:

$$\begin{aligned}
& W^V - W^U \qquad (64)\\
= & \underbrace{\int_{p^U}^{p^V} \left[\left(\int_\varphi s\left(\varphi\left(p - t^U\right)\right) \psi_\varphi\left(\varphi\right) \right) d\varphi - D\left(p\right) \right] dp}_{\text{price effects}} \\
& \underbrace{+ \int_\varphi \frac{1}{\varphi} \left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}(\varphi)} \pi\left(\varphi\left(p^V - \tau e\right)\right) \psi_e\left(e|\varphi\right) de \\ + \left(1 - \Psi_e\left(\hat{e}(\varphi)|\varphi\right)\right) \pi\left(\varphi\left(p^V - t^V\right)\right) - \pi\left(\varphi\left(p^V - t^U\right)\right) \end{array} \right) \psi_\varphi\left(\varphi\right) d\varphi}_{\text{reallocation gains}} \\
& - \underbrace{\left(v - \tau\right)\left(G^V - G^U\right)}_{\text{untaxed emissions}} - F \int_\varphi \psi_\varphi\left(\varphi\right) \Psi_\varphi\left(\widehat{e}_\varphi|\varphi\right) d\varphi.
\end{aligned}$$

The change in emissions itself is given by:

$$\begin{aligned}
& G^V - G^U \qquad (65)\\
= & \underbrace{\int_\varphi \psi_\varphi\left(\varphi\right) \left(s_0\left(\varphi\left(p^V - t^U\right)\right)^\alpha - s_0\left(\varphi\left(p^U - t^U\right)\right)^\alpha\right) E\left(e|\varphi\right) d\varphi}_{\text{price effect}} \\
& \underbrace{+ \int_\varphi \psi_\varphi\left(\varphi\right) d\varphi \left[\begin{array}{c} \int_{\underline{e}}^{\widehat{e}(\varphi)} s\left(\varphi\left(p^V - \tau e\right)\right) e \psi_e\left(e|\varphi\right) de \\ + \left(1 - \Psi_e\left(\widehat{e}(\varphi)|\varphi\right)\right) s_0\left(\varphi\left(p^V - t^V\right)\right)^\alpha E\left(e|e > \hat{e}(\varphi), \varphi\right) \\ - E\left(e|\varphi\right) s\left(\varphi\left(p^V - t^V\right)\right) \end{array} \right]}_{\text{reallocation}}.
\end{aligned}$$

We now use Taylor expansions to simplify these expressions. First, using (63), we can write:

$$\begin{aligned}
t^V &= \tau \frac{\int_\varphi \psi_\varphi\left(\varphi\right) \int_{\widehat{e}_\varphi}^\infty e \varphi^\alpha \psi_e\left(e|\varphi\right) de d\varphi}{\int_\varphi \psi_\varphi\left(\varphi\right) \varphi^\alpha \left(1 - \Psi_e\left(\widehat{e}_\varphi|\varphi\right)\right) d\varphi} + o\left(\tau\right), \qquad (66)\\
&= \tau \frac{\int_\varphi E(e\varphi^\alpha|\varphi, e > \hat{e}(\varphi)) \psi_\varphi(\varphi) d\varphi}{\int_\varphi \varphi^\alpha E(e|\varphi, e > \hat{e}(\varphi)) \psi_\varphi(\varphi) d\varphi} + o\left(\tau\right).
\end{aligned}$$

We differentiate equation (62) to get:

$$S^V\left(p\right) = S\left(p\right)\left(1 - \tau \frac{\alpha}{p} \tilde{E}(\varepsilon)\right) + o(\tau),$$

where $S\left(p\right) = s_0 p^\alpha \int_\varphi \varphi_i^\alpha \psi_\varphi(\varphi) d\varphi$ is total production at price p in laissez-faire, ε is still the revealed emission rate (i.e. $\varepsilon = e$ for $e < \widehat{e}_\varphi$ and $\varepsilon = t^V/\tau$ for $e > \widehat{e}_\varphi$), and $\tilde{E}(\varepsilon)$ is the φ^α-weighted (i.e. production-weighted) expectation of ε:

$$\tilde{E}(\varepsilon) = \frac{\int_\varphi \int_e \varepsilon \varphi^\alpha \psi_e(e|\varphi)\psi_\varphi(\varphi)}{\int_\varphi \int_e \varphi^\alpha \psi(\varphi, e) de d\varphi} d\varepsilon d\varphi = \int_\varphi \int_e \varepsilon \tilde{\psi}(\varphi, e) de d\varphi.$$

$\tilde{\psi}(\varphi, \varepsilon) = \frac{\varphi^\alpha \psi_e(e|\varphi)\psi_\varphi(\varphi)}{\int_\varphi \int_\varepsilon \varphi^\alpha \psi(\varphi,\varepsilon) d\varepsilon d\varphi}$ is the output-scaled joint distribution of φ and ε. Using (66), we note that $\tilde{E}(\varepsilon) = \tilde{E}(e)$ and it is therefore independent of the threshold function $\widehat{e}_\varphi$. As a result, we get that $\tilde{E}(\varepsilon) = G_0/S_0 = G^U/S^U$, so that the average revealed emission rate is the average emission rate in laissez-faire or under an output tax.

Furthermore, we obtain that $S^V(p) = S^U(p) + o(\tau)$,so that certification does not change supply at first order. This immediately implies that prices also stay constant at first order $p^V = p^U + o(\tau)$. We then get that price effect in (64) is still 0 at second order. Further, a second order Taylor expansion of the reallocation gains gives:

$$\text{Reallocation gains} = \tau^2 S(p_0) \ \tfrac{1}{2}\tfrac{\alpha}{p_0} V\tilde{ar}(\varepsilon) \ + o(\tau^2),$$

where $V\tilde{ar}(\varepsilon)$ given in (61), is the output-weighted variance of the revealed emissions. $S(p_0)$ is total production under p_0 and since and $S'(p_0) = \alpha S(p_0)/p_0$, we replicate the slope of aggregate supply from Corollary 3.

In a first order expansion of (65), the price effect term drops and we get:

$$G^V - G^U = -\tau S'(p_0) V\tilde{ar}(\varepsilon) + o(\tau),$$

Combining these terms, we get the change in welfare as:

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right) S'(p_0) V\tilde{ar}(\varepsilon) - F E_\varphi \Psi_\varphi(\widehat{e}(\varphi)) + o(\tau^2).$$

Both expressions mirror those of Corollary 3. □

A.5.3 Free Entry

We now allow for free entry. Firms must pay an entry cost F_E before drawing an emission rate. We consider small τ's such that all firms which enter produce and the set of entering firms N also denotes the set of producing firms. Further, for this section, instead of introducing a certification tax f, we assume that the government can decide on a maximal certification level $\tilde{e}$, so that a firm will certify if $e \leq \tilde{e}$ and $\pi(p - \tau e) - F \geq \pi(p - t)$.[38] In addition, we assume that the certification costs F are second order in τ (otherwise certification leads to welfare losses). As a result, the constraint $e \leq \tilde{e}$ binds and $\tilde{e} = \hat{e}$ as long as $\tilde{e}$ is not close to $E\left(e|e \geq \tilde{e}\right)$.[39]

In equilibrium, firms are indifferent between entering or not, which given our previous assumptions, leads to the free entry condition:

$$\int_{\underline{e}}^{\hat{e}} \pi(p - \tau e)\psi(e)de - F\Psi(\hat{e}) + \pi(p - t)(1 - \Psi(\hat{e})) = F_E. \tag{67}$$

This condition determines the endogenous mass of firms N. Consequently, market clearing is given by:

$$D(p) = N\left(\int_{\underline{e}}^{\hat{e}} s(p - \tau e)\psi(e)de + (1 - \Psi(\hat{e}))s(p - t)\right). \tag{68}$$

We then obtain the following corollary:

Corollary 9. *The change in emissions brought about by certification can be written as:*

$$G^V - G^U = -N_0 s'\left(p_0\right)\tau Var\left(\varepsilon\right) + o(\tau), \tag{69}$$

and the change in welfare as:

$$W^V - W^U = \left(v - \frac{\tau}{2}\right) N_0 s'\left(p_0\right) Var\left(\varepsilon\right)\tau - N_0 F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \tag{70}$$

N_0 denotes the mass of firms in laissez-faire. These expressions are identical to those without free-entry given in Corollary 3, so that our initial results are robust to considering free-entry. One important difference, however, is that total profits net of entry costs equal 0, so that the welfare benefits accrue to consumers. Intuitively, certification (ignoring the certification costs) initially increases aggregate profits, which leads to an increase in entry, which drives down the price (and profits) but raises consumers' welfare. Certification costs themselves have the opposite effects.

Proof. Since aggregate profits net of entry (and certification) costs are null, we can write

[38] We do not introduce the certification tax f because it would lead to a distortion in the entry decision of firms unless the government decides to introduce another policy to adjust entry.

[39] Otherwise, we get nearly full certification and $\Psi\left(\hat{e}\right) = \Psi\left(\tilde{e}\right) + O\left(\tau\right) = 1 + O(\tau)$, and we can ignore the difference between these three values in the coming expressions.

welfare as:

$$W = I + u(D(p)) - pC - vG.$$

Taking an approach analogous to Section A.2, we find the change in welfare as

$$W^V - W^U = -\int_{p^U}^{p^V} D(p)dp - (v-\tau)\left(G^V - G^U\right). \tag{71}$$

Emissions are given by:

$$\begin{aligned} & G^U - G^V \qquad (72) \\ = \; & N^V \underbrace{\left[E(e)\, s\left(p^U - \tau E(e)\right) - E(e)\, s\left(p^V - \tau E(e)\right)\right]}_{\text{price effect}} + \underbrace{\left(N^U - N^V\right) E(e)\, s\left(p^U - \tau E(e)\right)}_{\text{entry effect}} \\ & + N^V \underbrace{\left[E(e)\, s\left(p^V - \tau E(e)\right) - \left[\int_{\underline{e}}^{\widehat{e}} es\left(p - \tau e\right)\psi(e)\, de + s\left(p - t\right) E\left(e|e > \widehat{e}\right)\left(1 - \Psi\left(\widehat{e}\right)\right)\right]\right]}_{\text{reallocation effect}} + o(\tau). \end{aligned}$$

Taking a first-order Taylor expansion of equation (67), we get

$$p^V = p^U + o(\tau) = p_0 + \tau E(e) + o(\tau), \tag{73}$$

where we used that F is second order. Therefore the price difference between the certification equilibrium and the output tax equilibrium is second order. We then take a first-order Taylor expansion of (68) and get:

$$D'(p_0)\left(p^V - p_0\right) = N_0 s'(p_0)\left(p^V - \tau E(e) - p_0\right) + \left(N^V - N_0\right) s(p_0) + o(\tau).$$

Using (73), we get:

$$N^V = N^U + o(\tau) = N_0 + \tau \frac{D'(p_0)\, E(e)}{s(p_0)} + o(\tau^2). \tag{74}$$

so that the number of firms does not change at first order from the introduction of certification.

A first order Taylor expansion of (72), using (73) and (74) gives (69).

Furthermore the welfare changes given by equation (71) are zero at first order so we need to develop $p^V - p^U$ further. Therefore, we write $p^V - p_0 = \tau E(e) + \left(p^V - p_0\right)_2 + o(\tau^2)$ where $\left(p^{IB} - p_0\right)_2$ denotes the second order term in the price difference $p^V - p_0$. We then take a second order expansion of (67) and get

$$\left(p^V - p_0\right)_2 = \frac{F\Psi\left(\widehat{e}\right)}{s\left(p_0\right)} - \frac{1}{2}\frac{s'\left(p_0\right)}{s\left(p_0\right)}\tau^2 Var\left(\varepsilon\right),$$

from which we obtain:

$$p^V - p_0 = \tau E\left(e\right) + \frac{F\Psi\left(\widehat{e}\right)}{s\left(p_0\right)} - \frac{1}{2}\frac{s'\left(p_0\right)}{s\left(p_0\right)}\tau^2 Var\left(\varepsilon\right) + o\left(\tau^2\right),$$

and

$$p^V - p^U = \frac{F\Psi\left(\widehat{e}\right)}{s\left(p_0\right)} - \frac{\tau^2}{2}\frac{s'\left(p_0\right)}{s\left(p_0\right)} Var\left(\varepsilon\right) + o\left(\tau^2\right). \tag{75}$$

Using (73), (75) and that $D\left(p_0\right) = Ns\left(p_0\right)$, we then derive

$$\int_{p^{LB}}^{p^{IB}} \left(-D\left(p\right)\right) dp = \frac{1}{2}N_0 s'\left(p_0\right)\tau^2 Var\left(\varepsilon\right) - N_0 F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right).$$

Combining this expression with (71) and (69) delivers (70). □

B Adjustment on the extensive margin

We now study the case of adjustment along the extensive margin. Section B.1 solves the model presented in Section 2.6.2 and proves Proposition 6. Section B.2describes how we adjust the model for our study of the Permian Basin.

B.1 Baseline model

We consider the model described in Section 2.6.2. Section B.1.1 derives the slopes of the supply (20) and emission (21) functions under an output tax. Section B.1.2 establishes that the output tax given by (22) is optimal. Section B.1.3 describes the equilibrium with certification. Section B.1.4 derives expressions for welfare under certification and the output tax and derives equation (25). Section B.1.5 similarly derives equation (26).

B.1.1 Deriving slopes: a quick review of Dirac Delta Functions

Under laissez-faire, aggregate output and emissions are given by

$$S = \int_{\underline{u}}^{\bar{u}} \int_{\underline{q}}^{\bar{q}} \int_{\underline{e}}^{\bar{e}} q\mathbf{1}_{pq \geq uq} d\Psi(u, q, e) \text{ and } G = \int_{\underline{u}}^{\bar{u}} \int_{\underline{q}}^{\bar{q}} \int_{\underline{e}}^{\bar{e}} eq\mathbf{1}_{pq \geq uq} d\Psi(u, q, e), \tag{76}$$

which correspond to the expressions given by equation (19) when the output tax $t = 0$. These expressions rely on the indicator function$\mathbf{1}_{p-t\geq u}$, which is not differentiable at $pq = uq$.

To carry out Taylor approximations, we therefore employ the Dirac Delta Function and briefly remind the reader how this function operates. Strictly speaking it is a functional with the properties that

$$\delta_{t-a} = 0 \text{ for } t \neq a \text{ and } \int_{a-\varepsilon}^{a+\varepsilon} f(t)\delta_{t-a}dt = f(a) \text{ for } \varepsilon > 0,$$

that is a function that "picks out" the point $t = a$ such that the integral around a still sums to 1.

For some vector (x, y) distributed according to $\Psi(x, y)$ with pdf of $\psi(x, y)$ and $\psi_x(x)$ and $\psi_y(y|x)$, we consider the function:

$$F(a) = \int_x \int_y \mathbf{1}_{ax\geq y} f(y, x, a) d\Psi(x, y),$$

for some parameter a. We can differentiate with respect to a to get:

$$F'(a) = \int_x \int_y x\delta_{ax=y} f(ax, x, a) d\Psi(x, y) + \int_x \int_y \mathbf{1}_{ax\geq y} \frac{\partial f}{\partial a}(y, x, a) d\Psi(x, y).$$

The second term is standard. The first term can be written as:

$$\int_x \int_y x\delta_{ax=y} f(ax, x, a) d\Psi(x, y) = \int_x x f(ax, x, a)\psi_y(ax|x)\psi_x(x) dy dx,$$

that is, the Dirac Delta Function "picks" out the point $y = ax$ and the density $\psi_y(y|x)$ is evaluated at this point. We will repeatedly exploit this property in what follows.

In particular, we differentiate equation (76) to get:

$$S'(p) = \int_{c,q,e} q\delta_{p=u}\psi_u(p)\psi_{q,e}(q, e) dq de = \psi_u(p) \int_e \psi_e(e) \int_q q\psi_q(q|e) dq de, \tag{77}$$

where $\psi_e(e)$ is the unconditional pdf of e and $\psi_q(q|e)$ is the conditional pdf of q on e. Analogously for $G'(p)$, one gets:

$$G'(p) = \psi_u(p) \int_e \psi_e(e) \int_q eq\psi_q(q|e) dq de.$$

Similar derivations give equations (20) and (21) in the main text.

B.1.2 Optimal output tax

We can write welfare under an output tax as

$$W(t,v) = pS\left(p-t\right) - C\left(p-t\right) - vG\left(p-t\right), \tag{78}$$

where $C\left(p-t\right)$ denotes aggregate extraction cost:

$$C(p-t) = \int uq\mathbf{1}_{(p-t)\geq u}d\Psi(u,q,e). \tag{79}$$

To obtain the optimal output tax, we differentiate (78) with respect to t and get:

$$-pS'\left(p-t\right) + C'\left(p-t\right) + vG'\left(p-t\right) = 0. \tag{80}$$

Differentiating (79), we get

$$C'(p-t) = \psi_u(p-t)\left(p-t\right)\int_e\int_q q\psi_{q,e}(q,e)dqde.$$

Plugging this expression and (20) into (80) delivers the optimal output tax as:

$$t^* = v\frac{G'\left(p-t^*\right)}{S'\left(p-t^*\right)},$$

which at first order gives (22).

B.1.3 Equilibrium under certification

We now derive expressions for the equilibrium under certification. As mentioned in the text, instead of imposing a certification tax of f, we permit firms to certify to the point of $\tilde{e}$. A firm with $e > \tilde{e}$ will produce if $(p-t)q \geq 0$. A firm with $e < \tilde{e}$ can certify, if it does so, it pays τ per unit of emission for a total tax of τeq and a certification cost F. Therefore, a firm with $e \leq \tilde{e}$ produces if:

$$max\{(p-\tau e)q - c - F, (p-t)q - c\} \geq 0.$$

Such a firm would then prefer to certify over not certifying provided that

$$t - \tau e > \frac{F}{q}. \tag{81}$$

For reasons analogous to the model on the intensive margin, welfare effects will only be positive if F is second order in τ, which we assume from now on. The tax on uncertified firms, t, is implicitly defined by (23). If $\tilde{e}$ is not close to the $q-$weighted expectation of e above $\tilde{e}$, $\frac{E(eq|e\geq\tilde{e})}{E(q|e\geq\tilde{e})}$, then the LHS in (81) is first order in τ. The inequality is always satisfied so all firms with $e \leq \tilde{e}$ certify such that $\hat{e} = \tilde{e}$ and we can take $\hat{e}$ as an exogenous variable.

We can then write the aggregate supply as (24) and emissions are given by

$$G(p, \tau, t) = \int eq \left(\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{p-\tau e\geq u+F/q} + \mathbf{1}_{e>\hat{e}}\mathbf{1}_{p-t\geq u}\right) d\Psi(u, q, e). \tag{82}$$

where the integral is over (u, q, e) when not otherwise specified. The mass of firms that certify is given by:

$$M = \int \mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{p-\tau e\geq u+F/q} d\Psi(u, q, e).$$

If instead $\tilde{e}$ is close to $\frac{E(eq|e\geq\tilde{e})}{E(q|e\geq\tilde{e})}$, then the constraint $\hat{e} \leq \tilde{e}$ may not bind. The previous expressions still apply but $\hat{e}$ is now a function of q bounded by $\tilde{e}$, with $\hat{e}(q)$ close to $\frac{E(eq|e\geq\hat{e}(q))}{E(q|e\geq\hat{e}(q))}$ for all q. As a result, the q-weighted mass of firms which certify conditional on entry is close to 1 (and the q-weighted mass of firms with $e < \tilde{e}$ is also close to 1). As long as the correlation between e and q is not extreme, then the (unweighted) mass of certifying firms (conditional on entry) is also close to 1. We get $M = \Psi_u(p)\,\Psi_{q,e}(\tilde{e}) + o(1) = \Psi_u(p) + o(1)$ and all of our following expressions based on Taylor expansions apply as we can replace $\hat{e}(q)$ with an exogenous $\tilde{e}$ and ignore the dependence of $\hat{e}$ on q.

B.1.4 Change in Welfare

We find the welfare under certification and without certification. We write welfare as a function of the tax on emissions, τ, and social cost of emissions, v:

$$\begin{aligned} W(\tau, v) &= pS - C - vG - FM \\ &= \int \left[(pq - c - veq - F)\,\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{(p-\tau e)\geq u+\frac{F}{q}} + (pq - c - veq)\,\mathbf{1}_{e>\hat{e}}\mathbf{1}_{(p-t)\geq u}\right] d\Psi(u, q, e). \end{aligned}$$

We seek Taylor approximations to W in τ. t itself is a function of τ and we write explicitly $t(\tau)$ here. F is a parameter, which we assume is second order in τ. To make our Taylor expansions, we temporarily write $F(\tau)$—this is purely a technical convention which allows us to differentiate only once with respect to τ instead of differentiating with respect to both τ and F but it has no bearing on our results. With F second order, $F'(0) = 0$. We also use that $c = uq$ and differentiate to get:

$$\begin{aligned}
\frac{dW}{d\tau} &= \int\left[-\left((p-u-ve)-\frac{F}{q}\right)q\left(e+\frac{F'(\tau)}{q}\right)\mathbf{1}_{e\leq\hat{e}}\delta_{(p-\tau e)=u+F/q}-(p-u-ve)\,qt'(\tau)\mathbf{1}_{e>\hat{e}}\delta_{(p-t)=u}\right]d\Psi(u,q,e)\\
&\quad-\int F'(\tau)\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{(p-\tau e)\geq u+\frac{F}{q}}d\Psi(u,q,e).\\
&= \int\left[-(\tau-v)\,eq\left(e+\frac{F'(\tau)}{q}\right)\mathbf{1}_{e\leq\hat{e}}\delta_{(p-\tau e)=u+F/q}-(t-ve)\,qt'(\tau)\mathbf{1}_{e>\hat{e}}\delta_{(p-t)=u}\right]d\Psi(u,q,e) \qquad (83)\\
&\quad-\int F'(\tau)\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{(p-\tau e)\geq u+\frac{F}{q}}d\Psi(u,q,e)
\end{aligned}$$

Note that here we ignore the fact that $\hat{e}$ may differ from $\tilde{e}$ and depend on q,τ and F. This is consistent with our previous discussion and we verify below that this has no bearing on our approximations. We evaluate this expression at $\tau=0$ and use that F is second order in τ (so that $F'(0)=0$) as well as $t(0)=0$ to get:

$$\begin{aligned}
\frac{dW}{d\tau}|_{\tau=0} &= \int q\left[ve^2\mathbf{1}_{e\leq\hat{e}}\delta_{p=u}+veqt'(0)1_{e>\hat{e}}\delta_{p=u}\right]d\Psi(u,q,e) \qquad (84)\\
&= v\psi_u(p)\int_q q\psi_q(q)\left[\int_{e\leq\hat{e}}e^2\psi_e(e|q)de+\int_{e>\hat{e}}et'(0)\psi_e(e|q)de\right]dq.
\end{aligned}$$

To get the second order derivative we rewrite equation (83) as:

$$\begin{aligned}
\frac{dW}{d\tau} &= \int_q\psi_q(q)q\left[\int_{e\leq\hat{e}}(v-\tau)e\left(e+\frac{F'(\tau)}{q}\right)1_{e\leq\hat{e}}\psi_u\left(p-\tau e-\frac{F}{q}\right)de+\int_{e>\hat{e}}(ve-t(\tau))t'(\tau)\psi_u(p-t)de\right]dq\\
&\quad-\int_q\psi_q(q)F'(\tau)\int_{e\leq\hat{e}}\psi_e(e|q)\Psi_u\left(p-\tau e-\frac{F}{q}\right)dedq.
\end{aligned}$$

Differentiating and evaluating at $\tau=0$ (for which $F\left(0\right)=F'\left(0\right)=t\left(0\right)=0$) one gets:

$$\frac{dW^2}{d\tau^2}|_{\tau=0} = \int_q\psi_q(q)\left[\begin{array}{c}
q\int_{e\leq\hat{e}}-e^2 1_{e\leq\hat{e}}\psi_u\left(p\right)\psi_e(e|q)de+\int_{e\leq\hat{e}}veF''(0)1_{e\leq\hat{e}}\psi_u\left(p\right)\psi_e(e|q)de\\
-q\int_{e\leq\hat{e}}ve^3 1_{e\leq\hat{e}}\psi_u'\left(p\right)\psi_e(e|q)de-q\int_{e>\hat{e}}\left(t'(0)\right)^2\psi_u(p)\psi_e(e|q)de\\
+q\int_{e>\hat{e}}vet''(0)\psi_u(p)\psi_e(e|q)de-q\int_{e>\hat{e}}vet'(0)^2\psi_u'(p)\psi_e(e|q)de\\
-\int_{e\leq\hat{e}}\psi_e(e|q)F''(0)\Psi_u\left(p\right)de.
\end{array}\right]dq \qquad (85)$$

We can then perform a Taylor expansion of the welfare in τ following:

$$W(\tau,v)=W(0,v)+\frac{dW}{d\tau}|_{\tau=0}\tau+\frac{d^2W}{d\tau^2}|_{\tau=0}\frac{\tau^2}{2}+o(\tau^2). \qquad (86)$$

Using (84) and (85) and noting that many terms are of higher than second order (since v is small too), we can then write:

$$W(\tau,v) = W(0,v)+\tau\psi_u(p)\left[\begin{array}{c}\int_{e\le\hat{e}}\left(v-\frac{\tau}{2}\right)\psi_e(e)e^2\int_q q\psi_q(q|e)dqde\\ +\int_{e>\hat{e}}\left(ve-\frac{\tau}{2}t'(0)\right)t'(0)\psi_e(e)\int_q q\psi_q(q|e)dqde\end{array}\right] \tag{87}$$
$$-F\int_{e\le\hat{e}}\psi_e(e)\int_q\psi_q(q|e)\Psi_u(p)dedq+o(\tau^2).$$

Differentiating (23) with t expressed as a function of τ and evaluating this at $\tau=0$ gives $t=t'(0)\tau+o(\tau)$ with

$$t'(0)=\frac{\int eq\mathbf{1}_{e>\hat{e}}\mathbf{1}_{p\ge u}d\Psi(u,q,e)}{\int q\mathbf{1}_{e>\hat{e}}\mathbf{1}_{(p-t)\ge u}d\Psi(u,q,e)}.$$

We further define the q-weighted distribution of emission rates as:

$$\tilde{\psi}_e(e) = \frac{\psi_e(e)\int_q\psi_q(q|e)qdq}{\int_{e'}\psi_e(e')\int_q\psi_q(q|e')qdqde'}=\frac{\psi_u(p)\psi_e(e)\int_q\psi_q(q|e)qdq}{\frac{dS}{dp}|_{\tau=0}},$$

where $dS/dp|_{\tau=0}$ is given by equation (77). We can then rewrite (87) as:

$$W(\tau,v) = W(0,v)+\frac{dS}{dp}|_{\tau=0}\left[\int_{e\le\hat{e}}\tau\left(v-\frac{\tau}{2}\right)\tilde{\psi}(e)e^2de+\int_{e>\hat{e}}\left(ve-\frac{t}{2}\right)t\tilde{\psi}(e)de\right] \tag{88}$$
$$-F\int_{e\le\hat{e}}\psi_e(e)\int_q\psi_q(q|e)\Psi_u(p)dedq+o(\tau^2).$$

We then introduce ε defined as $\varepsilon=e$ if $e\le\hat{e}$ and $\varepsilon=t'(0)=E|_{\tilde{\psi}_e}(e|e>\hat{e})$ if $e>\hat{e}$ (i.e. $\varepsilon=t/\tau+o(1)$) where $E|_{\tilde{\psi}_e}$ is the expectation operator over $\tilde{\psi}_e$, with a variance $Var|_{\tilde{\psi}_e}$ defined on the same distribution. We consider the welfare expression for $\tau=v$ to get:

$$W^V(v,v) = W(0,v)+\frac{v^2}{2}S'(p)|_{\tau=0}\left[Var|_{\tilde{\psi}_e}(\varepsilon)+\left(E|_{\tilde{\psi}_e}(e)\right)^2\right] \tag{89}$$
$$-F\int_{e<\widehat{e}}\psi_e(e)\int_q\psi_q(q|e)\Psi_u(u)dedq+o\left(\tau^2\right),$$

where we make explicit that this is welfare for certification, W^V, in anticipation for welfare under no certification.

Setting $\hat{e}$ to the lower bound in (88), we get welfare under an output tax as:

$$\begin{aligned} W^U(t,v) &= W(0,v) + t^U \int q\delta_{p=u}\left[ve - \frac{1}{2}t^U\right] d\Psi(u,c,e) + o(t^2) \\ &= W(0,v) + S'(p)|_{\tau=0} E|_{\tilde{\psi}_e}\left(ve - \frac{1}{2}\tau\frac{G(0)}{S(0)}\right)\tau\frac{G(0)}{S(0)} + o\left(t^2\right), \end{aligned} \tag{90}$$

where we make it explicit that this is the tax without certification, t^U. Here $G(0)$ and $S(0)$ are emissions and supply under $\tau = 0$, and we note that average emissions $G(0)/S(0) = E|_{\tilde{\psi}_e}(e)$.

Taking the difference between (89) and (90) delivers:

$$W^V - W^U = S'(p)|_{\tau=0}\frac{v^2}{2}Var|_{\tilde{\psi}_e}(\varepsilon) - F\Psi_u(p)\int_{e\leq\hat{e}}\psi_e(e)\int_q \psi_q(q|e)dedq + o(\tau^2),$$

which can be rewritten as equation (25) of Proposition 6. As mention earlier, we note that if $\tilde{e}$ is close to $\frac{E(eq|e\geq\tilde{e})}{E(q|e\geq\tilde{e})}$ then nearly all active firms certify so that certification costs $FM = F\Psi_u(p) + o\left(\tau^2\right)$ and our previous analysis does apply.

B.1.5 Change in emissions

We proceed in a similar way for emissions. We differentiate (82) with respect to τ, writing again $t(\tau)$ and $F(\tau)$ and evaluate at $\tau = 0$ to get:

$$\begin{aligned} \frac{\partial G}{\partial \tau} &= -\int_{e\leq\hat{e}} e^2 \int_q q\psi_u(u)\psi_q(q|e)\psi_e(e)dqde - t'(0)\int_{e>\hat{e}} e\int_q q\psi_u(u)\psi_q(q|e)\psi_e(e)dqde \\ &= -S'(p)\int_{e\leq\hat{e}} e^2\tilde{\psi}_e(e)de - t'(0)S'(p)\int_{e>\hat{e}} e\tilde{\psi}_e(e)de. \end{aligned}$$

We can then write that

$$G^V(\tau) = G(0) - S'(p)\tau\left(\int_{e\leq\hat{e}} e^2\tilde{\psi}_e(e)de + E|_{\tilde{\psi}_e}(e|e>\hat{e})\int_{e>\hat{e}} e\tilde{\psi}_e(e)de\right) + o(\tau)$$

and

$$G^U(\tau) = G(0) - S'(p)\tau\left(E|_{\tilde{\psi}_e}(e)\right)^2 + o(\tau).$$

Taking the difference gives

$$G^V - G^U = -S'(p)\tau Var_{\tilde{\psi}_e}(\varepsilon) + o(\tau),$$

which is equation (26) in Proposition 6.

B.2 Special case for application to the Permian Basin

B.2.1 Adjustment to the extensive margin model

We make two adjustments to the model presented in Section 2.6.2. First, we introduce abatement. Given that some wells have very small emission rates, we consider the case of proportional abatement presented at the end of Section 2.6.1. As derived in Section A.5.1, certifying firms abate at rate $a^{*}(e) = b'^{-1}(\tau e) = \tau e / b"(0) + o(\tau)$.

Second, wells produce a mix of gas and oil which have different prices, which we denote $p^g + \rho$ and $p^o + \rho$, where ρ is a common shock which affect both fossil fuel prices, and p^g and p^o are fixed. Since firms face different prices, we introduce ρ for notional convenience, in order to derive the aggregate supply elasticity with respect to a common fossil fuel price shock. If a firm has a weight ω on gas, then that firm faces, in laissez-faire, a price of $p + \rho$ with $p = \omega p^g + (1-\omega) p^o$. Therefore each firm is described by a vector (u, q, e, p) with an associated c.d.f Ψ. Drilling costs are assumed to be proportional to revenues (excluding the shock ρ) so that $c = upq$ where u is independent of p, q and e. Moreover, we assume that p is also independent of q and e.

Finally, as before, we assume that $\hat{e}$ is exogenous and equal to $\tilde{e}$ since either the constraint $\hat{e} \leq \tilde{e}$ binds or nearly all entering firms certify.

We can then write supply as:

$$S = \int q \left(\mathbf{1}_{e \leq \hat{e}} \mathbf{1}_{p+\rho-\tau e(1-a)-b(a) > up + F/q} + \mathbf{1}_{e > \hat{e}} \mathbf{1}_{p+\rho-t > up} \right) d\Psi(u, q, e, p).$$

We define Laissez-faire supply without taxes as:

$$S^{LF}(\rho) \equiv \int q \mathbf{1}_{p+\rho > up} d\Psi(u, q, e, p).$$

The laissez-faire supply with no price shock is then equal to

$$S^{LF}(0) = \Psi_u(1) E|_{\psi_{e,q}}(q). \tag{91}$$

The derivative of aggregate supply in laissez-faire with respect to a common price shock (evaluated at 0), is then given by

$$\frac{\partial S^{LF}}{\partial \rho}(0) = \int_{u,q,e,p} \frac{q}{p} \delta_{u=1} d\Psi(u,q,e,p) = \psi_u(1) E|_{\psi_p} \left(\frac{1}{p}\right) E|_{\psi_{e,q}} (q), \tag{92}$$

where in the following it is understood that $\partial S^{LF}/\partial \rho$ is evaluated at zero.

B.2.2 Analytical derivation of the relevant expressions.

We now derive the changes in output, emissions, tax revenues, producer surplus and welfare brought about by output taxation and voluntary taxation.

Output. To find the change in output brought about by taxation, we perform calculations similar to those in the previous section. We find that

$$\frac{\partial S}{\partial \tau}|_{\tau=0} = -\psi_u(1) \int_{q,e,p} \frac{q}{p} \left(\mathbf{1}_{e \leq \hat{e}} e + t'(0) \mathbf{1}_{e > \hat{e}}\right) \psi_{q,e,p}(q,e,p)\, dqdedp. \tag{93}$$

The output tax can now be expressed as:

$$t(\tau) = \tau \frac{\int eq \mathbf{1}_{e>\hat{e}} \mathbf{1}_{p-t \geq up} d\Psi(u,q,e,p)}{\int q \mathbf{1}_{e>\hat{e}} \mathbf{1}_{p-t \geq up} d\Psi(u,q,e,p)},$$

which leads to

$$t'(0) = \frac{\int e \frac{q}{p} \mathbf{1}_{e>\hat{e}} d\Psi(q,e,p)}{\int \frac{q}{p} \mathbf{1}_{e>\hat{e}} d\Psi(q,e,p)} = E|_{\widetilde{\psi}_e} (e | e > \hat{e}), \tag{94}$$

where we used that p is independent of e, q and $\widetilde{\psi}_e$ is the $q-$weighted distribution of e. We therefore get that $t = \tau E|_{\widetilde{\psi}_e} (e | e > \hat{e}) + o(\tau)$. Plugging (94) in (93), we get:

$$\frac{\partial S}{\partial \tau}|_{\tau=0} = -\psi_u(1) E|_{\psi_p} \left(\frac{1}{p}\right) E|_{\psi_{e,q}} (eq) \tag{95}$$

We then obtain that the change in supply relative to laissez-faire is given by:

$$S^V = S^U + o(\tau) = S^{LF} - \tau \dot{S}^{LF}(0) E|_{\widetilde{\psi}_e} (e) + o(\tau). \tag{96}$$

Emissions and taxes. We can express emissions as (setting $\rho = 0$ here):

$$G = \int q \left(e(1-a) \mathbf{1}_{e \leq \hat{e}} \mathbf{1}_{p - \tau e(1-a) - b(a) > up + \frac{F}{q}} + e \mathbf{1}_{e>\hat{e}} \mathbf{1}_{p-t>up} \right) d\Psi(u,q,e,p).$$

In laissez-faire, we have $G^{LF} = E|_{\widetilde{\psi}_e}(e)\, S^{LF}$. We then get

$$\frac{\partial G}{\partial \tau}|_{\tau=0} = -\psi_u(1) E|_{\psi_p}\left(\frac{1}{p}\right)\int_{q,e} q\varepsilon^2 d\Psi_{q,e}(q,e) - \frac{\Psi_u(1)\,\Psi_e(\hat{e})\,E|_{\psi_{q,e}}\left(e^2 q|e<\hat{e}\right)}{b"(0)}.$$

Combining this expression with (91) and (92), we then obtain the change in emissions from voluntary certification and from an output tax as:

$$G^V - G^{LF} = -\tau\frac{\partial S^{LF}}{\partial \rho}E|_{\widetilde{\psi}_e}\left(\varepsilon^2\right) - \tau S^{LF}\frac{\widetilde{\Psi}_e(\hat{e})}{b"(0)}E|_{\widetilde{\psi}_e}\left(e^2|e<\hat{e}\right) + o(\tau); \tag{97}$$

and similarly, we get

$$G^U - G^{LF} = -\tau\frac{\partial S^{LF}}{\partial \rho}E|_{\widetilde{\psi}_e}(e)^2 + o(\tau). \tag{98}$$

From these expressions, one can derive tax revenues at second order:

$$T^V = \tau\left(G^{LF} - \tau\frac{\partial S^{LF}}{\partial \rho}E|_{\widetilde{\psi}_e}\left(\varepsilon^2\right) - \tau S^{LF}\frac{\widetilde{\Psi}_e(\hat{e})}{b"(0)}E|_{\widetilde{\psi}_e}\left(e^2|e<\hat{e}\right)\right) + o\left(\tau^2\right), \tag{99}$$

$$T^U = \tau\left(G^{LF} - \tau\frac{\partial S^{LF}}{\partial \rho}E|_{\widetilde{\psi}_e}(e)^2\right) + o\left(\tau^2\right). \tag{100}$$

Producer surplus. The producer surplus can be expressed as (for $\rho = 0$):

$$PS^V = \int q\left(\mathbf{1}_{e\leq\hat{e}}\left(p(1-u) - \tau e(1-a) - b(a) - \frac{F}{q}\right)^+ + \mathbf{1}_{e>\hat{e}}\left(p(1-u) - t\right)^+\right) d\Psi(u,q,e,p);$$

where we denote $x^+ = \max(x,0)$. As in section B.1.4, we take first-order derivative:

$$\begin{aligned}\frac{\partial PS^V}{\partial \tau} &= -\int q\mathbf{1}_{e\leq\hat{e}}\mathbf{1}_{p(1-u)-\tau e(1-a)-b(a)-\frac{F}{q}>0}\left(e(1-a) + \frac{F'(\tau)}{q}\right)d\Psi(u,q,e,p) \\ &\quad -\int q\mathbf{1}_{e>\hat{e}}\mathbf{1}_{p(1-u)-t>0}t'(\tau)\,d\Psi(u,q,e,p)\end{aligned} \tag{101}$$

Then one gets

$$\frac{\partial PS^V}{\partial \tau}|_{\tau=0} = -G^{LF}. \tag{102}$$

Further differentiating (101) and evaluating at $\tau = 0$, one further gets:

$$\frac{\partial^2 PS^V}{\partial \tau^2} = \psi_u(1) E|_{\psi_p}\left(\frac{1}{p}\right) E|_{\psi_{q,e}}\left(\varepsilon^2 q\right) + \Psi_u(1)\left[\frac{E|_{\psi_{q,e}}\left(e^2 q|e<\hat{e}\right)}{b''(0)} - F''(0)\right]\Psi_e(\hat{e}).$$

Combining this expression with (91) and (92) and using (102), we then obtain the change in producer surplus from voluntary certification and from an output tax as:

$$PS^V - PS^{LF} = -\tau G^{LF} + \frac{\tau^2 \frac{\partial S^{LF}}{\partial \rho} E|_{\widetilde{\psi}_e}\left(\varepsilon^2\right)}{2} + \frac{\tau^2 S^{LF} E|_{\widetilde{\psi}_e}\left(e^2|e<\hat{e}\right)\widetilde{\Psi}_e(\hat{e})}{2b''(0)} - F\Psi_e(\hat{e}) + o\left(\tau^2\right); \tag{103}$$

$$PS^U - PS^{LF} = -\tau G^{LF} + \frac{\tau^2 \frac{\partial S^{LF}}{\partial \rho} E|_{\widetilde{\psi}_e}(e)^2}{2} + o\left(\tau^2\right). \tag{104}$$

Welfare. We can then express the change in welfare relative to laissez-faire for Pigovian taxation ($\tau = v$) by combining (97), (99) and (103):

$$W^V(v,v) - W^{LF} = \frac{v^2}{2}\left(\frac{\partial S^{LF}}{\partial \rho} E|_{\widetilde{\psi}_e}\left(\varepsilon^2\right) + \frac{S^{LF}\widetilde{\Psi}_e(\hat{e})}{b''(0)} E|_{\widetilde{\psi}_e}\left(e^2|e<\hat{e}\right)\right) - F\Psi_e(\hat{e}) + o\left(v^2\right); \tag{105}$$

$$W^U(v) - W^{LF} = \frac{v^2}{2}\frac{\partial S^{LF}}{\partial \rho} E|_{\widetilde{\psi}_e}(e)^2 + o\left(v^2\right); \tag{106}$$

from this we can express the gains from certification relative to an output tax as:

$$W^V(v,v) - W^U = \frac{v^2}{2}\left(\frac{\partial S^{LF}}{\partial \rho} Var|_{\widetilde{\psi}_e}\left(\varepsilon^2\right) + \frac{S^{LF}\widetilde{\Psi}_e(\hat{e})}{b''(0)} E|_{\widetilde{\psi}_e}\left(e^2|e<\hat{e}\right)\right) - F\Psi_e(\hat{e}) + o\left(v^2\right).$$

B.2.3 Expressions for Permian Basin Calculations

In this section we collect the expressions based on sufficient statistics as presented in Table 1. Supply under laissez-faire, denoted S^{LF}, is the total estimated eight-year production of sites drilled in 2019. We then obtain that $\frac{\partial S^{LF}}{\partial \rho}(0) = \epsilon^S \frac{S^{LF}}{p_0}$, where p_0 is the average price and ϵ^S is the drilling elasticity equal to 1.26 (Newell and Prest (2019); Newell et al. (2019)). There are N sites, indexed with i. In these exercises methane is taxed at the social cost, v. Outcomes under an emissions tax correspond to a voluntary certification program with $\Psi(\widehat{e}) = 1$. Our assumption that drilling costs c are proportional to pq ensures that the observed emission rates and quantities in laissez-faire are representative from the distribution Ψ. We can then use the expressions previously derived. We get that $E|_{\widetilde{\psi}_e}(e|e>\hat{e}) = \frac{\sum_{i|e>\hat{e}} e_i q_i}{\sum_{i|e>\hat{e}} q_i}$, which is necessary to compute ε.

To calibrate the slope of the marginal abatement curve, $b"(0)$, we use the quantity-weighted estimate of methane abatement share from Marks (2022). This represents the aggregate emissions reduction from abatement when $\tau = v$, and is found to be 0.513. Under the assumption of proportional abatement, we have $a^*(e) = \tau e/b"(0)$. Using emissions-weighted moments to deliver the equivalent of the aggregate estimate under Marks (2022), we calibrate $b"(0) = \frac{v}{0.513}\frac{E|_{\tilde{\psi}_e}(e^2)}{E|_{\tilde{\psi}_e}(e)}$. Further, we get:

Output Tax Relative to Laissez Faire (Approximation)

Outcome	Formula	Reference
Production	$-\frac{\epsilon^S S^{LF}}{p_0} v \frac{\sum_i e_i q_i}{\sum_i q_i}$	eq (96)
Emissions	$-\frac{\epsilon^S S^{LF}}{p_0} v \left(\frac{\sum_i e_i q_i}{\sum_i q_i}\right)^2$	eq (98)
Producer Surplus	$-S^{LF} v \frac{\sum_i e_i q_i}{\sum_i q_i} + \frac{1}{2}\frac{\epsilon^S S^{LF}}{p_0}\left(v \frac{\sum_i e_i q_i}{\sum_i q_i}\right)^2$	eq (104)
Tax Revenue	$vS^{LF}\left(1 - \frac{\epsilon^S}{p_0} v \frac{\sum_i e_i q_i}{\sum_i q_i}\right)\frac{\sum_i e_i q_i}{\sum_i q_i}$	eq (100)
External Cost	$-v^2 \frac{\epsilon^S S^{LF}}{p_0}\left(\frac{\sum_i e_i q_i}{\sum_i q_i}\right)^2$	$v\times$eq (98)
Welfare	$\frac{v^2 \epsilon^S S^{LF}}{2p_0}\left(\frac{\sum_i e_i q_i}{\sum_i q_i}\right)^2$	eq (106)

Voluntary certification Tax Relative to Laissez Faire (Approximation)

Outcome	Formula	Reference
Production	$-\frac{\epsilon^S S^{LF}}{p_0} v \frac{\sum_i e_i q_i}{\sum_i q_i}$	eq (96)
Emissions	$-vS^{LF}\left(\frac{\epsilon^S}{p_0}\frac{\sum_i \varepsilon_i^2 q_i}{\sum_i q_i} + \frac{1}{b''(0)}\frac{\sum_{i\|e_i<\hat{e}} e_i^2 q_i}{\sum_i q_i}\right)$	eq (97)
Gross Prod. Surplus	$vS^{LF}\left(-\frac{\sum_i e_i q_i}{\sum_i q_i} + \frac{v\epsilon^S}{2p_0}\frac{\sum_i \varepsilon_i^2 q_i}{\sum_i q_i} + \frac{v}{2b''(0)}\frac{\sum_{i\|e_i<\hat{e}} e_i^2 q_i}{\sum_i q_i}\right)$	eq (103)
Tax Revenue	$vS^{LF}\left(\frac{\sum_i e_i q_i}{\sum_i q_i} - \frac{v\epsilon^S}{p_0}\frac{\sum_i \varepsilon_i^2 q_i}{\sum_i q_i} - \frac{v}{b''(0)}\frac{\sum_{i\|e_i<\hat{e}} e_i^2 q_i}{\sum_i q_i}\right)$	eq (99)
External Cost	$-v^2 S^{LF}\left(\frac{\epsilon^S}{p_0}\frac{\sum_i \varepsilon_i^2 q_i}{\sum_i q_i} + \frac{1}{b''(0)}\frac{\sum_{i\|e_i<\hat{e}} e_i^2 q_i}{\sum_i q_i}\right)$	$v\times$eq (97)
Gross Welfare	$\frac{v^2}{2} S^{LF}\left(\frac{\epsilon^S}{p_0}\frac{\sum_i \varepsilon_i^2 q_i}{\sum_i q_i} + \frac{1}{b''(0)}\frac{\sum_{i\|e_i<\hat{e}} e_i^2 q_i}{\sum_i q_i}\right)$	eq (105)

The expressions given here correspond to the gross producer surplus and gross welfare in the sense that they ignore potential certification costs.

B.2.4 Emissions Tax Relative to Laissez Faire with Uniformly Distributed Costs

We now solve exactly the model when the idiosyncratic cost shocks, u are distributed uniformly over the interval $[\underline{u},\bar{u}]$, where we assume the break-even cost shock is in the interval: $\underline{u} < 1 < \overline{u}$. Continuing the extensive margin setting with price heterogeneity of subsection B.2.1, aggregate supply in laissez-faire as a function of ρ (assuming that ρ is small such that $\underline{u} > \frac{p+\rho}{p} > \underline{u}$) is:

$$S^{LF}(\rho) = \frac{1}{\overline{u} - \underline{u}} \left[\int_{q,e,p} q \left(\frac{p+\rho}{p} - \underline{u} \right) d\Psi_{q,e,p}(q,e,p) \right].$$

Hence

$$S^{LF}(0) = \frac{1-\underline{u}}{\overline{u} - \underline{u}} \left[\int_q q d\Psi_q(q) \right]$$

and

$$\frac{\partial S^{LF}}{\partial \rho} = \frac{1}{\overline{u} - \underline{u}} \left[\int_{q,p} \frac{q}{p} d\Psi_{q,p}(q,e,p) \right]. \tag{107}$$

We then obtain that the elasticity in laissez-faire is given by

$$\epsilon^S = \frac{\frac{\partial S^{LF}}{\partial \rho} E(p)}{S^{LF}(0)} = \frac{E\left(\frac{q}{p}\right) E(p)}{(1-\underline{u}) E(q)}; \tag{108}$$

from which we can obtain

$$\underline{u} = 1 - \frac{1}{\epsilon^S} \frac{\left(\frac{1}{N}\sum_i \frac{q_i}{p_i}\right)\left(\frac{1}{N}\sum_i p_i\right)}{\left(\frac{1}{N}\sum_i q_i\right)}.$$

Note that we do not impose that p is independent of e and q when solving the model exactly.

To solve the model exactly, we need to specify an abatement cost function. Recall that by assumption, $a \in [0,1]$, $b(0) = 0$, $b'(0) = 0$, and $b''(0) > 0$. This is satisfied with a single parameter to calibrate with the abatement cost function

$$b(a) = \beta(-\ln(1-a) - a)$$

When choosing abatement levels, the first order condition $\beta \frac{a}{1-a} = ve$ yields an optimal abatement of $a^* = \frac{ve}{\beta+ve}$. We then get that

$$ve\left(1-a^*\right)+b\left(a^*\right)=\beta\ln\left(1+\frac{ve}{\beta}\right) \tag{109}$$

Using Marks (2022) quantity-weighted estimate of abatement, we calibrate β such that

$$\sum_i\left(\frac{ve_i}{\beta+ve_i}\frac{eq_i}{\sum_j eq_j}\right)=0.513$$

We only use the exact model in the case of full certification, therefore we simply assume that the government imposes an emission tax $\tau = v$ (and $F = 0$ here). We can then write output as

$$S=\int q\left(\mathbf{1}_{p-\tau e(1-a)-b(a)>p\underline{u}}\frac{p\left(1-\underline{u}\right)-\tau e\left(1-a\right)-b\left(a\right)}{p\left(\overline{u}-\underline{u}\right)}\right)d\Psi\left(q,e,p\right).$$

Using (107) and (109), we obtain:

$$S=\frac{\frac{\partial S^{LF}}{\partial\rho}\int\frac{q}{p}\left(p\left(1-\underline{u}\right)-\beta\ln\left(1+\frac{ve}{\beta}\right)\right)^{+}d\Psi\left(q,e,p\right)}{\int_{q,p}\frac{q}{p}d\Psi_{q,p}\left(q,e,p\right)}. \tag{110}$$

In sample, we can compute using (108)

$$\frac{\partial S^{LF}}{\partial\rho}=\frac{\epsilon^S S(0)}{\frac{1}{N}\sum_i p_i}\text{ and }S^{LF}(0)=\frac{\frac{\partial S^{LF}}{\partial\rho}\left(1-\underline{u}\right)\frac{1}{N}\sum_i q_i}{\frac{1}{N}\sum_i\frac{q_i}{p_i}}.$$

Using (110), we then get:

$$S-S^{LF}=\frac{\frac{\partial S^{LF}}{\partial\rho}\frac{1}{N}\sum_i q_i\left(\max\left\{\left(1-\underline{u}-\frac{\beta}{p_i}\ln\left(1+\frac{ve}{\beta}\right)\right),0\right\}-\left(1-\underline{u}\right)\right)}{\frac{1}{N}\sum_i\frac{q_i}{p_i}}. \tag{111}$$

We can similarly compute the change in emissions, producer surplus, tax revenues, external costs of emissions and welfare. We collect the expressions in the table below:

Change in:	Formula
Production	$\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\frac{q_i}{p_i}-q_i\left(1-\underline{u}\right)\right)$
Emissions	$\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\frac{q_i}{p_i}e_i\frac{\beta}{\beta+ve_i}-q_i\left(1-\underline{u}\right)e_i\right)$
Producer Surplus	$\frac{1}{2}\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\left[\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\right]^2\frac{q_i}{p_i}-p_iq_i\left(1-\underline{u}\right)^2\right)$
Tax Revenue	$v\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\frac{q_i}{p_i}e_i\frac{\beta}{\beta+ve_i}\right)$
External Cost	$v\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\frac{q_i}{p_i}e_i\frac{\beta}{\beta+ve_i}-q_i\left(1-\underline{u}\right)e_i\right)$
Welfare	$\frac{1}{2}\frac{\partial S^{LF}/\partial\rho}{\frac{1}{N}\sum_i \frac{q_i}{p_i}}\frac{1}{N}\sum_i\left(\begin{array}{c}\left[\max\left\{p_i\left(1-\underline{u}\right)+\beta\ln\left(\frac{\beta}{\beta+ve_i}\right),0\right\}\right]^2\frac{q_i}{p_i}\\ -q\left(1-\underline{u}\right)\left[p\left(1-\underline{u}\right)-2ve\right]\end{array}\right)$

B.2.5 Welfare change along the algorithm

Figure 2 shows the welfare gains with progressive unraveling along the lines of the algorithm presented in Section 2.5. Welfare at the different steps of the algorithm differs from welfare in a certification equilibrium with the same threshold of certification $\hat{e}$ because in a certification equilibrium non-certified firms pay the output tax $t=\tau E\left(e|e>\widehat{e}\right)$ while under the algorithm, non-certified firms pay an output tax which corresponds to the certification threshold of the previous round, $t_{-1}=\tau E\left(e|e>\widehat{e}_{-1}\right)$, where $\widehat{e}_{-1}$ was the certification threshold in the previous round. We therefore need to adjust our formula for the welfare changes accordingly.

We note that welfare under the algorithm can be written as:

$$W^{alg}\left(\tau,v,t_{-1}\right)=\int q\left[\begin{array}{c}\left(p\left(1-u\right)-ve\left(1-a\right)-b\left(a\right)-\frac{F}{q}\right)1_{e\leq\widehat{e}}1_{p(1-u)-\tau e(1-a)-b(a)-\frac{F}{q}\geq 0}\\ +\left(p\left(1-u\right)-ve\right)1_{e>\widehat{e}}1_{p(1-u)\geq t_{-1}}\end{array}\right]d\Psi\left(u,q,e,p\right).$$

Therefore the welfare difference under the algorithm and in the certification equilibrium can be written as:

$$\begin{aligned}W^{alg}\left(\tau,v,t_{-1}\right)-W^{cert}\left(\tau,v\right) &= \int q\left(p\left(1-u\right)-ve\right)1_{e>\widehat{e}}1_{1-u\geq\frac{t_{-1}}{p}}d\Psi\left(u,q,e,p\right)\\ &\quad-\int q\left(p\left(1-u\right)-ve\right)1_{e>\widehat{e}}1_{1-u\geq\frac{t}{p}}d\Psi\left(u,q,e,p\right).\end{aligned}$$

We follow a strategy similar to that in Appendix B.1.4 and compute

$$\begin{aligned}
& \frac{\partial\left(W^{alg}\left(\tau, v, t_{-1}\right)-W^{cert}\left(\tau, v\right)\right)}{\partial \tau} \\
= & \int -\frac{q}{p} E|_{\widetilde{\psi}_e}\left(e \mid e>\widehat{e}_{-1}\right)\left(t_{-1}-v e\right) \psi_u\left(1-\frac{t_{-1}}{p}\right) 1_{e>\widehat{e}} d \Psi_{q, e, p}(q, e, p) \\
& +\int \frac{q}{p} E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e})(t-v e) \psi_u\left(1-\frac{t}{p}\right) 1_{e>\widehat{e}} d \Psi_{q, e, p}(q, e, p) .
\end{aligned}$$

We then get:

$$\begin{aligned}
& \frac{\partial\left(W^{alg}\left(\tau, v, t_{-1}\right)-W^{cert}(\tau, v)\right)}{\partial \tau}|_{\tau=0} \\
= & v \psi_u(1)\left(E|_{\widetilde{\psi}_e}\left(e \mid e>\widehat{e}_{-1}\right)-E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e})\right) E|_{\psi_p}\left(\frac{1}{p}\right)\left(1-\widetilde{\Psi}_e(\widehat{e})\right) E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e}) E|_{\psi_{e, q}}(q) .
\end{aligned}$$

Further

$$\begin{aligned}
& \frac{\partial^2\left(W^{alg}\left(\tau, v, t_{-1}\right)-W^{cert}(\tau, v)\right)}{(\partial \tau)^2}|_{\tau=0} \\
= & \psi_u(1)\left(E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e})^2-E|_{\widetilde{\psi}_e}\left(e \mid e>\widehat{e}_{-1}\right)^2\right) E|_{\psi_p}\left(\frac{1}{p}\right)\left(1-\widetilde{\Psi}_e(\widehat{e})\right) E|_{\psi_{e, q}}(q)+O(v) .
\end{aligned}$$

Therefore, for our case of interest with $\tau=v$, we get that:

$$\begin{aligned}
& W^{alg}\left(v, v, t_{-1}\right)-W^{cert}(v, v) \\
=- & \frac{v^2}{2} \psi_u(1) E|_{\psi_p}\left(\frac{1}{p}\right)\left(1-\widetilde{\Psi}_e(\widehat{e})\right)\left(E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e})-E|_{\widetilde{\psi}_e}\left(e \mid e>\widehat{e}_{-1}\right)\right)^2 E|_{\psi_{e, q}}(q)+o\left(v^2\right) .
\end{aligned}$$

Using (92), we can rewrite this correction term

$$\begin{aligned}
& W^{alg}\left(v, v, t_{-1}\right)-W^{cert}(v, v) \\
=- & \frac{v^2}{2} \frac{\partial S^{LF}}{\partial \rho}\left(1-\widetilde{\Psi}_e(\widehat{e})\right)\left(E|_{\widetilde{\psi}_e}(e \mid e>\widehat{e})-E|_{\widetilde{\psi}_e}\left(e \mid e>\widehat{e}_{-1}\right)\right)^2+o\left(v^2\right) .
\end{aligned}$$

This correction is very natural: relative to the certification equilibrium, welfare is distorted because the uncertified firms are not paying the optimal output tax.

In the data, we can compute:

$$
\begin{aligned}
& W^{alg}\left(v, v, t_{-1}\right) - W^{cert}\left(v, v\right) \\
= - & \frac{v^2 \epsilon^S S^{LF}}{2p_0} \frac{\sum_{i|e_i>\widehat{e}} q_i}{\sum_i q_i} \left(\frac{\sum_{i|e_i>\widehat{e}} e_i q_i}{\sum_{i|e_i>\widehat{e}} q_i} - \frac{\sum_{i|e_i>\widehat{e}_{-1}} e_i q_i}{\sum_{i|e_i>\widehat{e}_{-1}} q_i} \right)^2 + o\left(v^2\right).
\end{aligned}
$$

C Theory Appendix: International

We consider the international setting. We predominantly solve the model for the pooling equilibrium ($\rho = 0$). Analogous calculations give the results for the separating equilibrium. In anticipation of the following results, we first establish the effect of $\hat{e}$ on Home prices, p_H, and Foreign prices, p_F.

C.1 Effect of changes in $\hat{e}$ on (p_H, p_F)in the pooling equilibrium

Lemma 10. *In the pooling equilibrium, the effect of rising $\hat{e}$ on p_H and p_F is given by:*

$$
\frac{dp_H}{d\hat{e}} = \frac{\left(D_F' - (1-\Psi_F(\hat{e}))s_F'(p_F)\right)\tau_F \frac{\partial E_F(e|e>\hat{e})}{\partial \hat{e}} + \left[s_F(p_F + A_F) - s_F(p_F)\right]\psi_F(\hat{e})}{\left\{D_H'(p_H) + D_F'(p_F) - \int_{\underline{e}}^{\hat{e}} \left[s_F'(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right] - (1-\Psi_F(\hat{e}))s_F'(p_F)\right\}}, \tag{112}
$$

$$
\frac{dp_F}{d\hat{e}} = \frac{-D_H'(p_H) + \int_{\underline{e}}^{\hat{e}} \left[s_F'(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right]\tau_F \frac{\partial E_F(e|e>\hat{e})}{\partial \hat{e}} + \left[s_F(p_F + A_F) - s_F(p_F)\right]\psi_F(\hat{e})}{\left\{D_H'(p_H) + D_F'(p_F) - \int_{\underline{e}}^{\hat{e}} \left[s_F'(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right] - (1-\Psi_F(\hat{e}))s_F'(p_F)\right\}} < 0, \tag{113}
$$

where $\partial E(e|e > \hat{e})/\partial \hat{e} > 0$. p_F decreases following an increase in certification; and p_H increases if abatement is small (A_F is small), which is the case when τ is small.

Proof. In the pooling equilibrium (29) holds with equality, leading to:

$$
p_F = (p_H - \tau_F E_F(e|e > \hat{e}) - \kappa). \tag{114}
$$

Plugging this expression into equation (31) and differentiating with respect to $\hat{e}$ gives (112) from which one can get (113). The sign of (113) is negative, while $\frac{dp_H}{d\hat{e}} > 0$ as long the term $\left[s_F(p_F + A_F) - s_F(p_F)\right]\psi_F(\hat{e})$ is dominated which occurs if A_F is small enough. This is in turn the case when τ is small. □

C.2 Proof of Proposition 7

We seek to establish Proposition 7 which gives the difference in world welfare between the uncertified equilibrium and the certified equilibrium. We prove Corollary 8 in the same process. We proceed in four steps: We first write explicitly the expressions for welfare and emissions. We then compute the price changes at first order, which allows us to take Taylor expansion of the welfare and emission changes. Finally, we derive the signs of the different effects.

C.2.1 The Welfare Expressions

We denote W^V world welfare under certification and W^U world welfare without certification when home imposes only an output-based tariff on its imports (and domestic taxation). Combining (27) and (28), we can write W^V(up to a constant equal to world exogenous income) as:

$$W^V = \underbrace{CS_H + CS_F}_{\text{consumer surpluses}} + \underbrace{PS_H + PS_F}_{\text{producer surpluses}} \underbrace{- \left[(v - \tau_H) G_H + (v - \tau_F) G_F + \tau_F G_{F,dom} \right]}_{\text{non-internalized emissions}} - F\Psi_F(\hat{e}) . \tag{115}$$

Consumer and producer surpluses are given by:

$$CS_H = u_H(C_H) - p_H C_H \text{ and } CS_F = u_F(C_F) - p_F C_F, \tag{116}$$

$$\begin{aligned} PS_H = & \int_{\underline{e}}^{\infty} (p_H - \tau_H e + A_H) s_H (p_H - \tau_H e + A_H) \psi_H(e) \, de \\ & - \int_{\underline{e}}^{\infty} c_H (s_H (p_H - \tau_H e + A_H)) \psi_H(e) \, de, \end{aligned} \tag{117}$$

$$\begin{aligned} PS_F = & \int_{\underline{e}}^{\hat{e}} (p_H - \tau_F e + A_F - \kappa) s_F (p_H - \tau_F e + A_F - \kappa) \psi_F(e) \, de \\ & - \int_{\underline{0}}^{\hat{e}} c_F (s_F (p_H - \tau_F e + A_F)) \psi_F(e) \, de + (p_F s_F(p_F) - c_F(s_F(p_F))) (1 - \Psi_F(\hat{e})), \end{aligned} \tag{118}$$

where c_F and c_H are the cost functions associated with Foreign and Home supply functions. Further emissions at home and foreign are given by:

$$G_H = \int_{\underline{e}}^{\bar{e}} (e - a_H) s_H (p_H - \tau_H e + A_H) \psi_H(e) \, de,$$

$$G_F = \int_{\underline{e}}^{\hat{e}} (e - a_F) s_F (p_H - \tau_F e + A_F - \kappa) \psi_F(e) \, de + s_F(p_F) E(e | e > \hat{e}) (1 - \Psi_F(\hat{e})),$$

while $G_{F,dom}$ which correspond to the foreign emissions from domestic consumptions are given by:

$$G_{F,dom} = \tau_F E_F(e | e > \hat{e}) D_F(p_F)$$

since domestic foreign producers are uncertified.

We assume that the Home government gives an export subsidy which is equivalent to the output based tariff, namely $\tau^F E^F\left(e|e>\widehat{e}^S\right)$, when Home firms export to foreign. This subsidy is only active in an equilibrium where there is enough certification that Home firms export to Foreign. For brevity, we ignore that case throughout this Appendix but briefly discuss its implications at the end of section 3.3.

C.2.2 Taylor Approximations of price changes

We denote by p_0 the price at home in an equilibrium without any taxes. We are considering a case where Foreign exports to Home, so that $p_0-\kappa$ is the Foreign price in this equilibrium. Taylor approximations are undertaken assuming that τ, v, κ are of the same order.

A first-order Taylor approximation of the global market clearing equation (31) gives

$$\begin{aligned}
& D'_H(p_0)\left(p_H^V-p_0\right)+D'_F(p_0-\kappa)\left(p_F^V-p_0+\kappa\right) \\
= \quad & s'_H(p_0)\left(p_H^V-\tau_H E_H(e)-p_0\right)+s'_F(p_0-\kappa) \\
& +\left(\Psi_F(\hat{e})\left(p_H^V-\tau_F E_F(e|e<\hat{e})\right)+(1-\Psi_F(\hat{e}))\left(p_F^V+\kappa\right)-p_0\right)+o(\tau),
\end{aligned} \tag{119}$$

where we used that $A_F=A_H=o(\tau)$.

In the uniform output-based tariff equilibrium, we get:

$$\begin{aligned}
& D'_H(p_0)\left(p_H^U-p_0\right)+D'_F(p_0-\kappa)\left(p_F^U-p_0+\kappa\right) \\
= \quad & s'_H(p_0)\left(p_H^U-\tau_H E_H(e)-p_0\right)+s'_F(p_0-\kappa)\left(p_F^U-p_0+\kappa\right)+o(\tau),
\end{aligned} \tag{120}$$

and

$$p_F^U=p_H^U-\kappa-\tau_F E_F\left(e\right). \tag{121}$$

We can then express the H price in that equilibrium as:

$$p_H^U-p_0=\frac{s'_H\left(p_0\right)\tau_H E_H\left(e\right)+s'_F\left(p_0-\kappa\right)\tau_F E_F\left(e\right)-D'_F\left(p_0-\kappa\right)\tau_F E_F\left(e\right)}{s'_H\left(p_0\right)+s'_F\left(p_0-\kappa\right)-D'_H\left(p_0\right)-D'_F\left(p_0-\kappa\right)}+o\left(\tau\right). \tag{122}$$

Separating equilibrium. We first focus on a separating equilibrium. At order 0, (32) implies that:

$$s_F(p_0-\kappa)(1-\Psi_F(\hat{e}))=D_F(p_0-\kappa)+o(1), \tag{123}$$

which pins down $\hat{e}$ at zeroth order:

$$\hat{e} = \hat{e}_0 + \tilde{e} + o(\tau), \quad \text{where } \hat{e}_0 = \Psi_F^{-1}\left(1 - D_F(p_0 - \kappa)/s_F(p_0 - \kappa)\right),$$

and $\tilde{e}$ is the "first order" term of $\hat{e}$. If $\hat{e}$ is such that equation (123) holds with "$>$" we are in the pooling equilibrium. If it holds with "$<$, we are in the case where Home firms export to Foreign, which we ignore here. Note, $\hat{e}$ and $\hat{e}_0$ differ only to a first order and consequently whether we evaluate functions at $\hat{e}$ or $\hat{e}_0$ will be equivalent in our Taylor expansions. For ease of exposition we will use $\hat{e}$ both here and in equation (33) in the main text. This is correct, though a more stringent adherence to convention would have us evaluate at $\hat{e}_0$ for the separating equilibrium.

From the certification condition (30), one gets that at first order:

$$p_H - \tau_F \hat{e} - p_F = \kappa + \frac{F+f}{s_F\left(p_0 - \kappa\right)} + o\left(\tau\right), \tag{124}$$

note that we implicitly assumed that $\frac{F+f}{s_F(p_0-\kappa)}$ is first order or smaller otherwise no firm would want to certify. Plugging this expression in the definition of ρ gives (37).

The separating equilibrium is then characterized by the following condition:

$$0 < \tau_F\left(E_F\left(e|e > \hat{e}\right) - \hat{e}\right) - \frac{F+f}{s_F\left(p_0 - \kappa\right)} < 2\kappa.$$

The first inequality reflects the condition $\rho > 0$, and ensures that profits from selling domestically are strictly higher for uncertified Foreign firms than exporting. The second inequality ensures that Home firms would not want to export to Foreign when they obtain an export subsidy $\tau^F E^F\left(e|e > \widehat{e}^S\right)$ (i.e. $p_H > p_F + \tau^F E^F\left(e|e > \widehat{e}^S\right) - \kappa$).

We use equation (124), label p_H and p_F, p_H^V and p_F^V, respectively and substitute for p_F^V in equation (119) to get:

$$p_H^V - p_0 = \frac{\begin{pmatrix} s_H'(p_0)\tau_H E_H(e) + s_F'(p_0-\kappa)\left(\Psi_F(\hat{e})\tau_F E_F(e|e<\hat{e}) + (1-\Psi_F(\hat{e}))\left(\tau_F\hat{e} + \frac{F+f}{s_F(p_0-\kappa)}\right)\right) \\ -D_F'(p_0-\kappa)\left(\tau_F\hat{e} + \frac{F+f}{s_F(p_0-\kappa)}\right)\end{pmatrix}}{s_H'(p_0) + s_F'(p_0-\kappa) - D_H'(p_0) - D_F'(p_0-\kappa)} + o(\tau),$$

and combine this with equation (122) to get

$$p_H^V - p_H^U = \frac{\begin{pmatrix} \left(1-\Psi_F\left(\widehat{e}\right)\right) s_F^{'}\left(p_0-\kappa\right)\left(\left(\tau_F \widehat{e} + \frac{F+f}{s_F(p_0-\kappa)}\right) - \tau_F E_F\left(e|e>\widehat{e}\right)\right) \\ -D_F^{\prime}\left(p_0-\kappa\right)\left(\tau_F \widehat{e} + \frac{F+f}{s(p_0-\kappa)} - \tau_F E_F\left(e\right)\right) \end{pmatrix}}{s_H^{\prime}\left(p_0\right) + s_H^{\prime}\left(p_0-\kappa\right) - D_H^{\prime}\left(p_0\right) - D_F^{\prime}\left(p_0-\kappa\right)} + o\left(\tau\right), \tag{125}$$

and similarly:

$$p_F^V - p_F^U = \frac{\begin{pmatrix} s_H^{'}\left(p_0\right)\left(\tau_F E_F\left(e\right) - \left(\tau_F \hat{e} + \frac{f}{s_F(p_0)}\right)\right) \\ +s_F^{'}\left(p_0-\kappa\right)\left(\tau_F E_F\left(e\right) - \left(1-\Psi_F\left(\hat{e}\right)\right)\tau_F E_F\left(e|e>\hat{e}\right) - \Psi_F\left(\hat{e}\right)\left(\tau_F \hat{e} + \frac{f}{s_F(p_0-\kappa)}\right)\right) \\ -\left(\tau_F E_F\left(e\right) - \left(\tau_F \hat{e} + \frac{f}{s^F(p_0-\kappa)}\right)\right) D_H^{'}\left(p_0\right) \end{pmatrix}}{s_H^{'}\left(p_0\right) + s_F^{'}\left(p_0-\kappa\right) - D_H^{'}\left(p_0\right) - D_F^{'}\left(p_0-\kappa\right)} + o(\tau). \tag{126}$$

We define ϵ_F^D as the elasticity of demand wrt. prices in foreign as $\epsilon_F^D = D_F^{'}(p_0-\kappa)(p_0-\kappa)/D_F(p_0-\kappa) = D_F^{\prime}(p_0-\kappa)p_0/D_F(p_0-\kappa) + o(\tau)$ where the equality follows because κ is of the same order as τ. We further define Foreign share in demand $\theta_F^D = D_F(p_0-\kappa)/\left(D_F(p_0-\kappa)+D_H(p_0)\right)$. Analogous versions of ϵ and θ exist for Home and supply and we let $\epsilon^D = \theta_F^D\epsilon_F^D + \theta_H^D\epsilon_H^D$ be the elasticity of world demand wrt. price. We then write:

$$\begin{aligned} & p_H^V - p_H^U + o\left(\tau\right) \\ = & \frac{\begin{pmatrix} \left(1-\Psi_F\left(\widehat{e}_0\right)\right)\theta_F^S\epsilon_F^S\left(\left(\tau_F \widehat{e}_0 + \frac{F+f}{s_F(p_0-\kappa)}\right) - \tau_F E_F\left(e|e>\widehat{e}_0\right)\right) \\ -\epsilon_F^D\theta_F^D\left(\tau_F \widehat{e}_0 + \frac{F+f}{s(p_0-\kappa)} - \tau_F E_F\left(e\right)\right) \end{pmatrix}}{\epsilon_S - \epsilon_D}. \end{aligned}$$

Using (37), we obtain (34). Using (121) and (29), we then get (35).

Pooling equilibrium. In the pooling equilibrium (124) still applies and $\rho = 0$, so that

$$E_F(e|e>\hat{e}) - \hat{e} = \frac{1}{\tau_F}\frac{F+f}{s_F\left(p_0-\kappa\right)} + o\left(1\right),$$

which defines $\hat{e}$ at order 0. Note that to be in the pooling equilibrium $\frac{F+f}{s_F(p_0-\kappa)}$ must be first order: if it is larger than no firm would want to certify, while if it is smaller, nearly all firms would certify which contradicts the assumption that we are in a pooling equilibrium.

Using that $\rho = 0$, such that (114) holds with equality, in (119) delivers:

$$p_H^V - p_0 = \frac{s_H^{'}(p_0)\tau_H E_H(e) + s_F^{'}(p_0-\kappa)\tau_F E_F(e) - D_F^{'}(p_0-\kappa)\tau_F E_F(e|e>\hat{e})}{s_H^{'}(p_0) + s_F^{'}(p_0-\kappa) - D_H^{'}(p_0) - D_F^{'}(p_0-\kappa)} + o(\tau).$$

Combine this with equation (122) to get

$$p_H^V - p_H^U = \frac{D_F^{'}(p_0-\kappa)\tau_F\left(E_F\left(e\right) - E_F(e|e>\hat{e})\right)}{s_H^{'}(p_0) + s_F^{'}(p_0-\kappa) - D_H^{'}(p_0) - D_F^{'}(p_0-\kappa)} + o(\tau).$$

This implies that (34) still holds in the pooling equilibrium. Using that $\rho = 0$, we then further get that (35) also holds in the pooling equilibrium.

C.2.3 Taylor approximations for welfare and emission changes

We now derive Taylor approximations for the welfare and emissions changes. First, we sum the changes in consumer and producer surplus (from (116), (117) and (118)) to get:

$$\begin{aligned}
& CS_H^V + CS_F^V + PS_H^V + PS_F^V - \left(CS_H^U + CS_F^U + PS_H^U + PS_F^U\right) \\
= \; & \underbrace{\begin{array}{c} E_H\left(\int_{p_H^U}^{p_H^V} s_H\left(\widetilde{p} - \tau_H e + A_H\right) d\widetilde{p}\right) + E_F\left(\int_{p_H^U}^{p_H^V} s_F\left(\widetilde{p} - \tau_F E_F\left(e\right) - \kappa\right) d\widetilde{p}\right) \\ - \left(\int_{p_F^U}^{p_F^V} D_H\left(\widetilde{p}\right) d\widetilde{p} + \int_{p_F^U}^{p_F^V} D_F\left(\widetilde{p}\right) d\widetilde{p}\right) \end{array}}_{\equiv \text{price effect}} \\
& + \underbrace{\left(1 - \Psi_F\left(\hat{e}\right)\right)\left[\pi_F\left(p_F^V\right) - \pi_F\left(p_H^V - \tau_F E_F\left(e|e>\hat{e}\right) - \kappa\right)\right]}_{\equiv \text{adjustment term}} \\
& + \underbrace{\Psi_F\left(\hat{e}\right) E_F\left(\left(\pi_F\left(p_H^V - \tau_F e + A_F - \kappa\right) - \pi_F\left(p_H^V - \tau_F e - \kappa\right)\right)|e<\hat{e}\right)}_{\equiv \text{abatement gains}} \\
& + \underbrace{E\left(\pi_F\left(p_H^V - \tau_F \varepsilon - \kappa\right)\right) - \pi_F\left(p_H^V - \tau_F E_F\left(e\right) - \kappa\right)}_{\equiv \text{reallocation term}},
\end{aligned}$$

where the *adjustment term* reflects the "extra" profits uncertified firms receive from Foreign prices being higher than what they would earn selling to Home. The remaining terms are explained in the main text.

We handle each of these terms in order:

$$\begin{aligned}
& \text{price effects} \qquad (127) \\
= \; & -D_F\left(p_0 - \kappa\right)\left(p_F^V - p_H^V + \tau_F E_F\left(e\right) + \kappa\right) \\
& + \left(p_H^V - p_H^U\right)\left[\begin{array}{c} s_H^{'}\left(p_0\right)\left(\frac{p_H^V + p_H^U}{2} - \tau_H E_H\left(e\right) - p_0\right) \\ + s^{F\prime}\left(p_0 - \kappa\right)\left(\frac{p_H^V + p_H^U}{2} - \tau_F E_F\left(e\right) - p_0\right) - D_H^{'}\left(p_0\right)\left(\frac{p_H^V + p_H^U}{2} - p_0\right) \end{array}\right] \\
& - D_F^{'}\left(p_0 - \kappa\right)\left(p_F^V - p_F^U\right)\left(\frac{p_F^V + p_F^U}{2} - p_0 + \kappa\right) + o\left(\tau^2\right).
\end{aligned}$$

Summing up (119) and (120) gives:

$$
\begin{aligned}
& s_H^{'}\left(p_0\right)\left(\frac{p_H^V+p_H^U}{2}-\tau_H E_H\left(e\right)-p_0\right) \\
& +s_F^{'}\left(p_0-\kappa\right)\left(\frac{\Psi_F\left(\hat{e}\right)\left(p_H^V-\tau_F E_F\left(e|e<\hat{e}\right)\right)+\left(1-\Psi_F\left(\hat{e}\right)\right)\left(p_F^V+\kappa\right)+p_H^U-\tau_F E_F\left(e\right)}{2}-p_0\right) \\
& -D_H^{'}\left(p_0\right)\left(\frac{p_H^V+p_H^U}{2}-p_0\right)-D_F^{'}\left(p_0-\kappa\right)\left(\frac{p_F^V+p_F^U}{2}-p_0+\kappa\right) \\
= \; & o\left(\tau\right).
\end{aligned}
$$

Plugging this expression in (127) delivers:

$$
\begin{aligned}
& \textbf{price effects} \qquad (128)\\
= \; & -D_F\left(p_0-\kappa\right)\left(p_F^V-p_H^V+\tau_F E_F\left(e\right)+\kappa\right) \\
& +\left(p_H^V-p_H^U\right)s_F^{'}\left(p_0-\kappa\right)\frac{\left(1-\Psi_F\left(\hat{e}\right)\right)\left(p_H^V-p_F^V-\kappa-\tau_F E_F\left(e|e>\widehat{e}^F\right)\right)}{2} \\
& +D_F^{'}\left(p_0-\kappa\right)\left(p_H^V-p_F^V-\tau_F E_F\left(e\right)-\kappa\right)\left(\frac{p_F^V+p_F^U}{2}-p_0+\kappa\right)+o\left(\tau^2\right)
\end{aligned}
$$

We continue with the reallocation term:

$$
\textbf{reallocation term} = \frac{s_F^{\prime}\left(p_0-\kappa\right)}{2}\left(\tau_F\right)^2 Var_F\left(\varepsilon\right)+o\left(\tau^2\right), \qquad (129)
$$

the abatement term:

$$
\textbf{abatement term} = \Psi_F\left(\hat{e}\right)s_F\left(p_0-\kappa\right)A_F+o\left(\tau^2\right), \qquad (130)
$$

and the adjustment term:

$$
\begin{aligned}
& \textbf{adjustment term} \qquad (131)\\
= \; & \left(1-\Psi_F\left(\hat{e}\right)\right)\left(p_F^V+\kappa-p_H^V+\tau_F E\left(e|e>\hat{e}\right)\right)\left[\begin{array}{c} s_F\left(p_0-\kappa\right) \\ +s_F^{'}\left(p_0-\kappa\right)\left[\frac{p_F^V+\kappa+p_H^V-\tau_F E_F\left(e|e>\hat{e}\right)}{2}-p_0\right]\end{array}\right]+o\left(\tau^2\right).
\end{aligned}
$$

We now look at terms corresponding to the non-internalized emissions. We first directly derive the change in emissions at home at first order as (38) and remark that we only need this expression at first order to get a second order approximation of the welfare change. The change in Foreign emissions can be written as:

$$
\begin{aligned}
& G_F^V - G_F^U \\
= \; & s_F^{'}\left(p_0-\kappa\right)\left[\int_{\underline{e}}^{\hat{e}} e\left(p_H^V - \tau_F e - \kappa - p_F^U\right)\psi_F\left(e\right)de + \left(p_F^V - p_F^U\right)E_F\left(e|e>\hat{e}\right)\left(1-\Psi_F\left(\hat{e}\right)\right)\right] + o(\tau) \\
& -\Psi_F\left(\hat{e}\right)a_F s_F\left(p_0-\kappa\right) + o(\tau)
\end{aligned}
$$

Using (121) and (29) then gives the first order change in Foreign emissions as (39), which again is sufficient to compute the second order change in welfare. We have therefore established Corollary 8.

We note that we can write:

$$
\begin{aligned}
& \tau_F\left(G_{F,dom}^V - G_{F,dom}^U\right) \\
= \; & \tau_F\left(E_F\left(e|e>\hat{e}\right) - E_F\left(e\right)\right)D_F\left(p_0-\kappa\right) \\
& + D_F^{'}\left(p_0-\kappa\right)\left[\tau_F E_F\left(e|e>\hat{e}\right)\left(p_F^V - p_0 + \kappa\right) - \tau_F E_F\left(e\right)\left(p_F^U - p_0 + \kappa\right)\right] + o\left(\tau^2\right)
\end{aligned}
$$

Plugging this term, (128), (129), (130) and (131) in (115), and using the definition of ρ, we obtain the welfare change from certification as:

$$
\begin{aligned}
& W^V - W^U \qquad (132) \\
= \; & \left(1-\Psi_F\left(\hat{e}\right)\right)s_F^{'}\left(p_0-\kappa\right)\rho\left(\tfrac{p_F^V+\kappa+p_H^U-\tau_F E_F\left(e|e>\hat{e}\right)}{2} - p_0\right) \\
& + D_F^{'}\left(p_0-\kappa\right)\left(-\rho\left(\frac{p_F^V+p_F^U}{2} - p_0 + \kappa\right) + \tau_F\frac{E_F\left(e|e>\hat{e}\right)+E_F\left(e\right)}{2}\left(p_F^U - p_F^V\right)\right) \\
& + \rho\,\left[\left(1-\Psi_F\left(\hat{e}\right)\right)s_F\left(p_0-\kappa\right) - D_F\left(p_0-\kappa\right)\right] \\
& + \frac{s_F^{\prime}\left(p_0-\kappa\right)}{2}\left(\tau_F\right)^2 Var_F\left(\varepsilon\right) + \Psi_F\left(\hat{e}\right)s_F\left(p_0-\kappa\right)A_F + o\left(\tau^2\right) \\
& - \left(v-\tau_H\right)\left(G_H^V - G_H^U\right) - \left(v-\tau_F\right)\left(G_F^V - G_F^U\right) - F\Psi_F\left(\hat{e}\right)
\end{aligned}
$$

Taking a first-order expansion of the market-clearing equation in Foreign in the separating equilibrium (32), we get

$$
\begin{aligned}
& D_F(p_0-\kappa) + D_F^{'}\left(p_0-\kappa\right)\left(p_F^V - p_0 + \kappa\right) + o\left(\tau\right) \\
= \; & s_F(p_0-\kappa)(1-\Psi_F(\hat{e})) + s_F^{\prime}\left(p_0-\kappa\right)\left(p_F^V - p_0 + \kappa\right)\left(1-\Psi_F(\hat{e})\right).
\end{aligned}
$$

Since $\rho = 0$ in the pooling equilibrium, then we always have

$$\begin{aligned} & \rho\ \left[\left(1-\Psi_F\left(\hat{e}\right)\right) s_F\left(p_0-\kappa\right)-D_F\left(p_0-\kappa\right)\right]+o\left(\tau^2\right) \\ = & \ \rho\left(D_F^{\prime}\left(p_0-\kappa\right)\left(p_F^V-p_0+\kappa\right)-s_F^{\prime}\left(p_0-\kappa\right)\left(p_F^V-p_0+\kappa\right)\left(1-\Psi_F(\hat{e})\right)\right). \end{aligned}$$

Plugging this expression in (132) and using (29), we get:

$$\begin{aligned} & W^V - W^U \\ = & -\left(1-\Psi_F\left(\hat{e}\right)\right) s_F^{\prime}\left(p_0-\kappa\right) \rho\left(\ \frac{\rho+\Delta p_H}{2}\ \right) \\ & -\frac{D_F^{\prime}\left(p_0-\kappa\right)}{2} \Delta p_F\left(\tau_F\left(E_F\left(e \mid e>\hat{e}\right)+E_F\left(e\right)\right)-\rho\right) \\ & +\frac{s_F^{\prime}\left(p_0-\kappa\right)}{2}\left(\tau_F\right)^2 Var_F\left(\varepsilon\right)+\Psi_F\left(\hat{e}\right) s_F\left(p_0-\kappa\right) A_F+o\left(\tau^2\right) \\ & -\left(v-\tau_H\right)\left(G_H^V-G_H^U\right)-\left(v-\tau_F\right)\left(G_F^V-G_F^U\right)-F \Psi_F\left(\hat{e}\right). \end{aligned} \tag{133}$$

Plugging in the expressions for the change in emissions (38) and (39) and using $A_F = \tau_F^2/\left(2b''\left(0\right)\right)+o\left(\tau_F^2\right)$,we get the change in welfare expressed as in (33).

C.2.4 The signs of the Backfilling and Consumption Leakage Effects

Sign of the backfilling effect. We first show that the Backfilling effect is always weakly negative when $v \geq \tau_F$. Recall that the Backfilling effect is given by:

$$-s_F^{\prime}\left(1-\Psi_F\left(\widehat{e}\right)\right)\left(\frac{\Delta p_H+\rho}{2}+(v-\tau_F) E_F(e \mid e>\hat{e})\right) \rho.$$

This is zero in the pooling equation where $\rho = 0$. Consequently, consider a separating equilibrium where $\rho > 0$. For $v \geq \tau_F$ this expression is negative if $\Delta p_H + \rho > 0$. Using (29) and (121), we get

$$\Delta p_H+\rho=\Delta p_F+\tau_F\left[E_F(e \mid e>\hat{e})-E_F(e)\right].$$

Next using the expression for Δp_F (equation 126), we get:

$$\begin{aligned} \Delta p_H+\rho= & \frac{\rho\left[s^{H\prime}\left(p_0\right)+s_F^{\prime} \Psi_F-D_H^{\prime}\right]}{s_H^{\prime}\left(p_0\right)+s^{F\prime}\left(p_0-\kappa\right)-D^{H\prime}\left(p_0\right)-D^{F\prime}\left(p_0-\kappa\right)} \\ & -\frac{\tau_F D^{\prime}\left[E_F(e \mid e>\hat{e})-E_F(e)\right]}{s^{H\prime}\left(p_0\right)+s^{F\prime}\left(p_0-\kappa\right)-D^{H\prime}\left(p_0\right)-D^{F\prime}\left(p_0-\kappa\right)}>0, \end{aligned}$$

Both ρ and $[E_F(e|e > \hat{e}) - E_F(e)]$ are positive. Consequently $\Delta p_H + \rho > 0$ and the Backfilling effect (which contains a minus) is always negative in the separating equilibrium.

Sign of the consumption leakage effect. The consumption leakage effect is given by:

$$-D'_F \left(\frac{\tau_F \left(E_F(e) + E_F(e|e > \hat{e}) \right) - \rho}{2} \right) \Delta p_F.$$

In the pooling equilibrium $\Delta p_F < 0$ and $\rho = 0$ ensuring that the consumption leakage effect is negative.

Combining (29) and (124), we get that:

$$\tau_F \left(E_F(e) + E_F(e|e > \hat{e}) \right) - \rho = \tau_F \left(E_F(e) + \hat{e} \right) + \frac{F + f}{s_F \left(p_0 - \kappa \right)},$$

which is positive if $f \geq 0$. Therefore, in that case, the sign of the consumption leakage is the same as the sign of Δp_F (this is also true in the pooling equilibrium). Our numerical example for the steel industry shows that this sign is ambiguous (see Table 3).

C.3 Optimal policy

In the following, we look at the optimal unilateral policy for a Home policy maker who maximizes world welfare. For simplicity, we directly restrict attention to a set of policies here instead of starting from a general allocation problem as in the domestic case (section 2.4), though such an approach would lead to the same results.

Specifically, we permit the Home policy maker to offer a price $p_E^R - \tau_F(e - a)$ to a Foreign firm which reveals its e, exports and undertakes abatement. The Home policy maker can also tax / subsidize certification at rate f. The Home policy maker offers p_E^U to exporters who do not reveal. She can set any allocation at Home but cannot set other policies for Foreign. With this price schedule, certified foreign firms profits are given by $\pi_F \left(p_E^R - \tau_F(e - a) - \kappa - b_F(a) \right)$ where π is the profit function previously defined, and they abate $a_F^* = {b'_F}^{-1}(\tau_F)$. Uncertified Foreign firm receive $\pi_F(p_F)$ by selling domestically or $\pi_F \left(p_E^U - \kappa \right)$ by exporting. We assume that there is always some foreign demand, so that $\pi_F(p_F) \geq \pi_F \left(p_E^U - \kappa \right)$ and $p_F \geq p_E^U - \kappa$ with equalities if there are uncertified foreign exports. As a result, foreign firms certify and export if

$$\pi_F \left(p_E^R - \tau_F(e - a) - \kappa - b_F(a) \right) - F - f \geq \pi_F(p_F),$$

which naturally implies the existence of a threshold $\hat{e}$ so that firms certify for $e \leq \hat{e}$ and the previous inequality is an equality for $e = \hat{e}$. Since f allows to freely adjust $\hat{e}$ and plays no other role, we let the social planner choose $\hat{e}$ directly.

The Home policy maker is constrained by market forces in Foreign which imply that $C_F = D_F(p_F)$ with the demand function defined as before and lead to the market clearing equation:

$$D\left(p^F\right) = s^F\left(p^F\right)\left(1 - \Psi^F\left(\hat{e}^F\right)\right) - M, \tag{134}$$

where M denotes export by uncertified Foreign firms (for $M > 0$), and imports by Foreign from Home (if $M < 0$).

The Home social planner chooses $q_H(e), a_H(e), \hat{e}, \tau_F, p_E^R, M^U$ and p^F in order to maximize world welfare

$$\begin{aligned}
& W \qquad (135)\\
= \; & u_H\left(\int_0^\infty q_H(e)\,\psi_H(e)\,de + \int_0^{\hat{e}} s_F\left(p_E^R - \tau_F(e - a_F) - b_F(a_F) - \kappa\right)\psi_F(e)\,de + M\right)\\
& - \int_0^\infty \left(c_H(q_H(e)) + b_H(a_H(e))\,q_H(e)\right)\psi_H(e)\,de + u_F(D_F(p_F))\\
& - \int_0^{\hat{e}} \left(c_F\left(s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right) + (b_F(a_F) + \kappa)\,s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right)\psi_F(e)\,de\\
& - c_F(s_F(p_F))(1 - \Psi_F(\hat{e})) - \kappa|M| - \int_0^\infty (e - a_H(e))\,q_H(e)\,\psi_H(e)\,de\\
& - v\left(\int_0^{\hat{e}} (e - a_F)\,s_F\left(p_E^R - \tau_F(e - a_F) - b_F(a_F) - \kappa\right)\psi_F(e)\,de + E_F(e|e > \hat{e})\,s_F(p_F)(1 - \Psi_F(\hat{e}))\right)\\
& - F\Psi_F(\hat{e}),
\end{aligned}$$

with the constraint that Foreign markets clear (equation 134). We solve the corresponding Lagrange problem and let λ be the Lagrange multiplier on (134).[40] We derive first order conditions and subsequently check the three cases of M (< 0, $= 0$ and > 0).

The first order conditions of this problem are:

$$\text{wrt. } q_H(e)\text{: } u_H'(C_H) - c_H'(q_H(e)) - b_H(a_H(e)) - v(e - a_H(e)) = 0,$$

$$\text{wrt. } a_H(e) : b_H'(a_H(e)) = \nu.$$

Hence, the optimal policy has firms in Home paying a tax $\tau_H = \nu$, undertaking optimal abatement and facing a price $p_H = u_H'(C_H)$.

[40] Note that we can eliminate p_E^U: either $M > 0$ and $p_E^U = p_F + \kappa$ or $M \leq 0$ and there are no uncertified exporting firms so that the exact value of p_E^U does not matter as long as $p_E^U < p_F + \kappa$.

$$\text{wrt. } p_E^R : \int_0^{\hat{e}} \left(p_H - p_E^R + (\tau_F - v)(e - a_F) \right) s_F' \left(p_E^R - \tau_F e + A_F - \kappa \right) \psi_F(e)\, de = 0, \quad (136)$$

where as before $A_F \equiv \tau_F a_F - b_F(a_F)$,

$$\text{wrt. } p_F{:}D_F'(p_F)\, p_F - s_F'(p_F)\left(p_F + vE_F(e|e > \hat{e})\right)(1 - \Psi_F(\hat{e})) + \lambda \left(s_F'(p_F)(1 - \Psi_F(\hat{e})) - D_F'(p_F)\right) = 0, \quad (137)$$

FOC wrt. to M:

$$\lambda = p_H - \kappa \text{ if } M > 0; \text{ and } \lambda = p_H + \kappa \text{ if } M < 0, \quad (138)$$

or we have $D_F(p_F) = s_F(p_F)(1 - \Psi_F(\hat{e}))$ if $M = 0$.

FOC wrt. τ_F leads to:

$$\int_0^{\hat{e}} \left[\begin{array}{c} \left(p_E^R - p_H + (\nu - \tau_F)(e - a_F)\right) s_F'\left(p_E^R - \tau_F e + A_F - \kappa\right)(-(e - a_F)) \\ + \frac{da_F}{d\tau^F}\left(b_F'(a_F) - \nu\right)\left(s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right) \end{array} \right] \psi_F(e)\, de = 0 \quad (139)$$

FOC wrt. $\hat{e}$

$$\begin{aligned} p_H s_F\left(p_E^R - \tau^F \hat{e} + A_F - \kappa\right) - c_F\left(s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right)\right) \\ - \left(b_F(a_F) + \kappa\right) s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right) + c_F(s_F(p_F)) + c_F(s_F(p_F)) \\ -v\left((e - a_F)\, s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right) - \hat{e} s_F(p_F)\right) - F - \lambda s_F(p_F) \quad = 0. \end{aligned} \quad (140)$$

Together equations (136) and (139) imply that:

$$p_H = p_E^R \text{ and } \tau_F = \tau_H = \nu,$$

so that certified Foreign firms simply face a Pigouvian emission tax as they export to Home. We can then rewrite (140) as

$$\pi_F\left(p_H - \tau_F(\hat{e} - a_F) - b_F(a_F) - \kappa\right) - F = (\lambda - v\hat{e})\, s_F(p_F) - c_F(s_F(p_F))\,. \quad (141)$$

We then solve for the system of (137), (138) and (141) for the three different cases where $M > 0$, $M < 0$ and $M = 0$:

In the pooling case ($M > 0$), we get $\lambda = p_H - \kappa$. Plugging this in (137), gives $p_F =$

$p_H - t^* - \kappa$, with

$$t^* = v\frac{s'_F\left(p_F\right)E_F\left(e|e>\hat{e}\right)\left(1-\Psi_F\left(\hat{e}\right)\right)}{s'_F\left(p_F\right)\left(1-\Psi_F\left(\hat{e}\right)\right)-D'_F\left(p_F\right)}, \tag{142}$$

so that uncertified Foreign firms face an output-based tariff given by t^*. We can then rewrite (141) as

$$\pi_F\left(p_H - \tau_F\left(\hat{e}-a_F\right) - b_F\left(a_F\right) - \kappa\right) - F - f^* = \pi_F\left(p_F\right), \tag{143}$$

with

$$f^* = \left(t^* - v\hat{e}\right)s_F\left(p_F\right). \tag{144}$$

Therefore Foreign firms pay a certification tax given by f.

In the case where Home firms export, $M < 0$, then $\lambda = p_H + \kappa$. Plugging this in (137), gives $p_F = p_H - t^* + \kappa$, with t^* still defined by (142), that is Home firms receive an export subsidy given by t^* when they export to Foreign. This again delivers (143), so that Foreign firms still pay a certification tax given by (144).

Finally, in the separating case, $M = 0$, we get that p_H, p_F and $\hat{e}$ are defined by Foreign market clearing (134) which becomes $D\left(p^F\right) = s^F\left(p^F\right)\left(1-\Psi^F\left(\hat{e}^F\right)\right)$, a Home market clearing equation, and the first order condition on $\hat{e}$, (141) which can be written as (143) with f^* given by (144). This allocation can be implemented with the trade tax t^* defined in (143) as long as $p_H - p_F - \kappa < t^* < p_H - p_F + \kappa$ (though in that case the trade tax is inactive and other values can implement the same allocation).

How t^* compares to $p_H - p_F - \kappa$ and $p_H - p_F + \kappa$ therefore determines the type of allocation that is optimal. Bringing the three cases together, we obtain:

Proposition 11. *A Home policymaker who can implement any allocation at Home, has access to an output-based tariff for uncertified Foreign firms, an emission-based tariff for certified Foreign firms and a certification tax, but is otherwise constrained by market forces in Foreign, maximizes welfare by implementing the following policy: an emission tax at Home and for certified Foreign firms given by v, an output-based tariff on uncertified Foreign firms given by $t^* = v\frac{s'_F(p_F)E_F(e|e>\hat{e})(1-\Psi_F(\hat{e}))}{s'_F(p_F)(1-\Psi_F(\hat{e}))-D'_F(p_F)}$ or an export subsidy on Home firms given by the same formula, and a certification tax given by $f^* = \left(t^* - v\hat{e}\right)s_F\left(p_F\right)$.*

We compare the resulting welfare from solving problem (135) to that of a setting with no Home taxes on Foreign exports, $\tau_F = t = 0$, which we label W^{LF} (for laissez-faire). We perform calculations of a Taylor expansion along the lines of those in previous sections (details omitted) to find:

$$
\begin{aligned}
W^* - W^{LF} \;\; = \;\; & s_F^{'}(p_0-\kappa)\frac{v^2}{2}Var(\varepsilon) + \Psi_F(\hat{e})A_F s_F(p_0-\kappa) - F\Psi_F(\hat{e}) \\
& + D_F^{'}(p_0-\kappa)\frac{t^*}{2}E_F(e|e>\hat{e}) - s_F^{'}(p_0-\kappa)\frac{v}{2}E_F(e)\left(p_H^* - vE_F(e) - p_H^{LF}\right),
\end{aligned}
$$

where t^* obeys (142). This welfare expression holds in all three cases, though strictly speaking t^* only applies as a tax in the pooling equilibrium.

D Data Appendix

D.1 Domestic Application

Lease-level annual production in the Permian basin is pulled from online portals of the states of Texas and New Mexico. In 2019 daily production was about 4.9 million barrels per day of oil and 18 million barrels of oil-equivalent of natural gas. There were approximately 75,000 leases actively producing, two thirds of which were in Texas. Texas wells were more productive on average, yielding 75-80% of the basin's oil and gas. To focus on the lifetime production of unconventional wells, we include data from sites whose first production began between 2012 and 2019. We remove leases that produce sporadically ($\sim$ 1% of production in 2019).

While 20% of wells produce only gas, and 13% only oil, the overwhelming majority of output comes from wells that produce a mix of the two. In Figure B.1b this is presented as a kernel density, in which the oil share is weighted by well production. Most output comes from wells producing about 80% oil, which may make flaring an attractive option for economically disposing of methane in the absence of sufficient collection infrastructure.

The drilling decision weighs the cost of drilling against the lifetime revenue a well will produce. While production across wells is highly heterogeneous, the profile of production decay is a well-understood phenomenon that depends on falling pressure as resources are extracted (Hyne (2001)). In Figure B.2 we plot the mean production-age profile by vintage. While there has been incredible growth in initial production, rates of decay have been remarkably stable over time. We use this fact to estimate 8-year production volumes, which account for the overwhelming majority of a well's output (Jacobs (2020)). Letting y_{it} denote the natural log of production in barrels of oil-equivalent (BOE) from site i in month t, we conduct an exponential decline curve analysis (DCA) according to

$$y_{it} = \gamma_i + \sum_{y=1}^{4+} [\eta_y + \beta_y \mathbf{1}_y age_{it}] + u_{it}$$

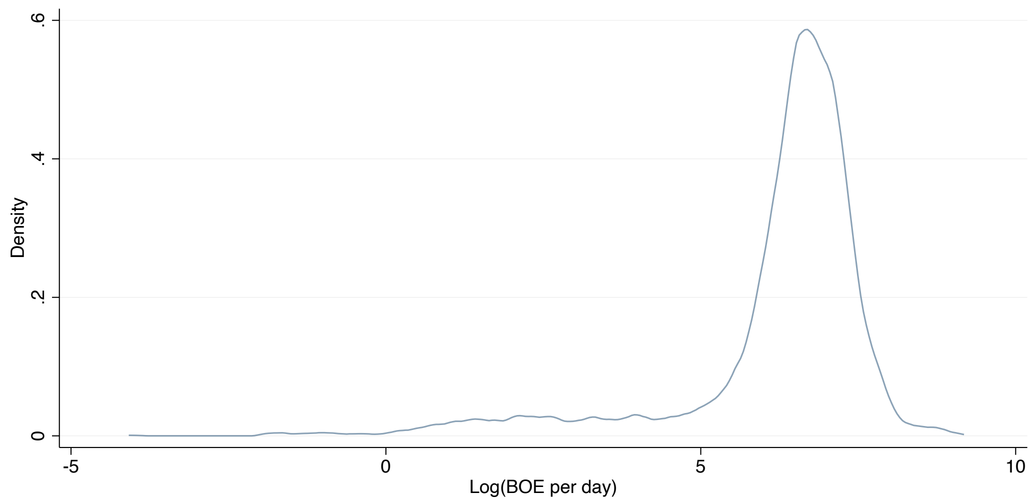
where γ_i are site-level fixed effects, we include separate age fixed effects and within-year slopes for each of the first three years of production, and estimate a single decay rate thereafter. For each site we then project the expected ultimate recovery by combining the site-level fixed effects and estimated decay parameters.

To estimate the distribution of emissions rates, we distinguish between 'simple' and 'complex' sites, following Robertson et al. (2020). Complex sites have some combination of storage tanks and compressors, which are disproportionate sources of emissions. Simple sites are composed of wellheads connected via pipes and hoses to a centralized collection and gathering (C&G) stations. We account for the emissions that occur at these C&G stations using the data from Zimmerle et al. (2020). Emissions from complex sites are drawn from the data in Robertson et al. (2020).

We bootstrap the distribution of site emissions rates as follows: For each bootstrap sample, we draw from a linear probability model based on the data collected by Robertson et al. (2020) to determine whether a site is simple or complex as a function of oil-to-gas ratio and production volume. Simple sites are assigned emissions rates randomly drawn from Zimmerle et al. (2020). For complex sites, bootstrap samples are drawn from Robertson et al. (2020) and matched based on production rates to wells in the Texas and New Mexico production data. For the set of wells drilled in 2019, a single bootstrap iteration consists of estimated lifetime production of oil and gas, simple/complex status, and lifetime methane emissions. We use fuel prices in 2019 as expected future prices following Alquist and Kilian (2010)

Figure B.1: Permian Basin Lease Production and Oil Share

(a) *Log*(BOE per day)



(b) Oil Share

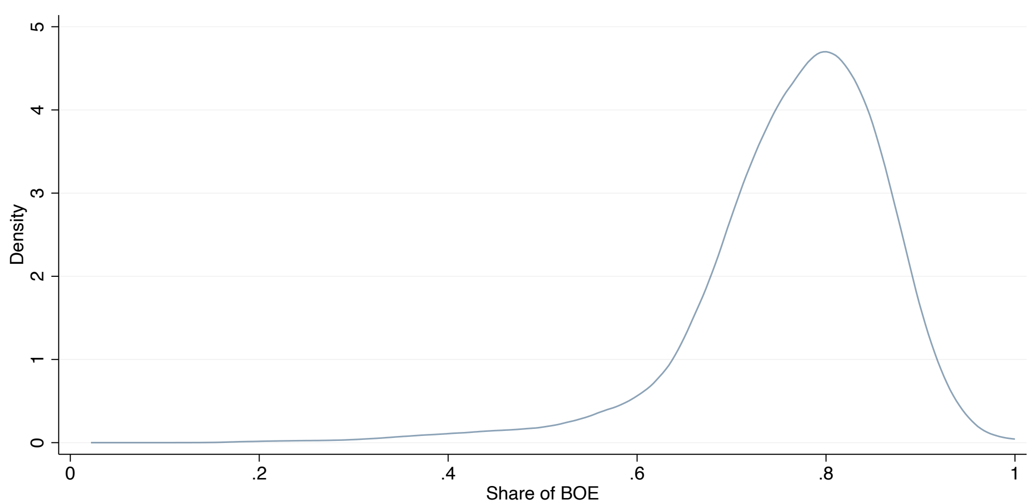


Figure B.2: Monthly Production Profile by Vintage

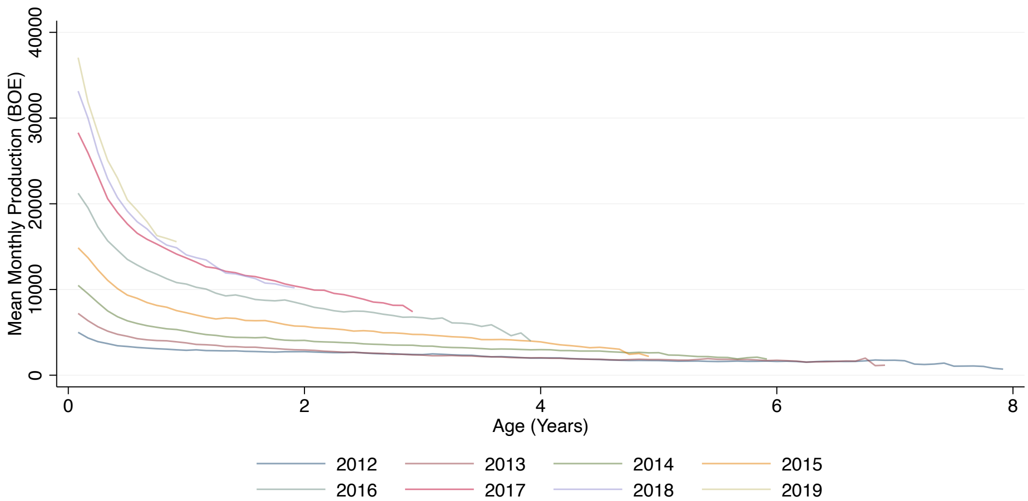
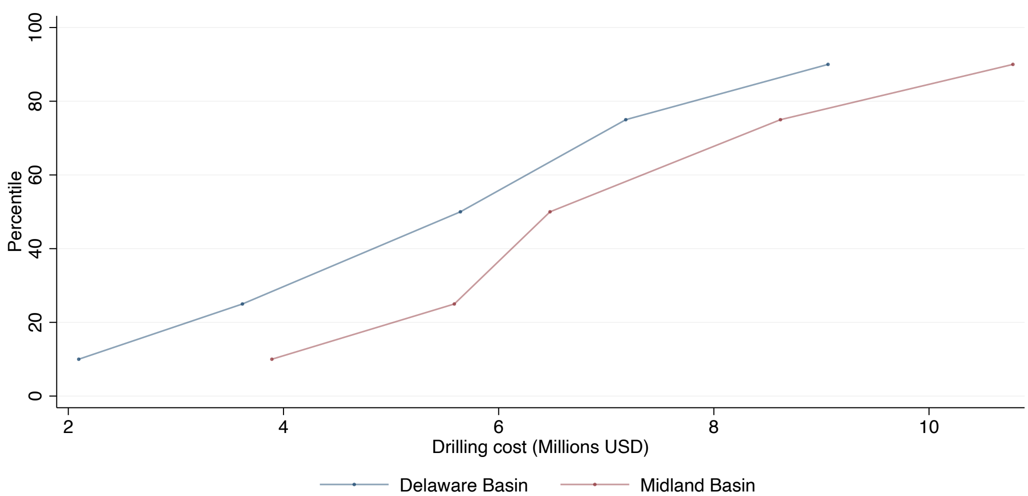


Figure B.3: Distributions of Drilling Costs in the Permian



Source: Energy Information Administration, "Trends in U.S. Oil and Natural Gas Upstream Costs," 2016.

D.2 International Application: Steel

This Appendix provides details on our calibration and shows additional results.

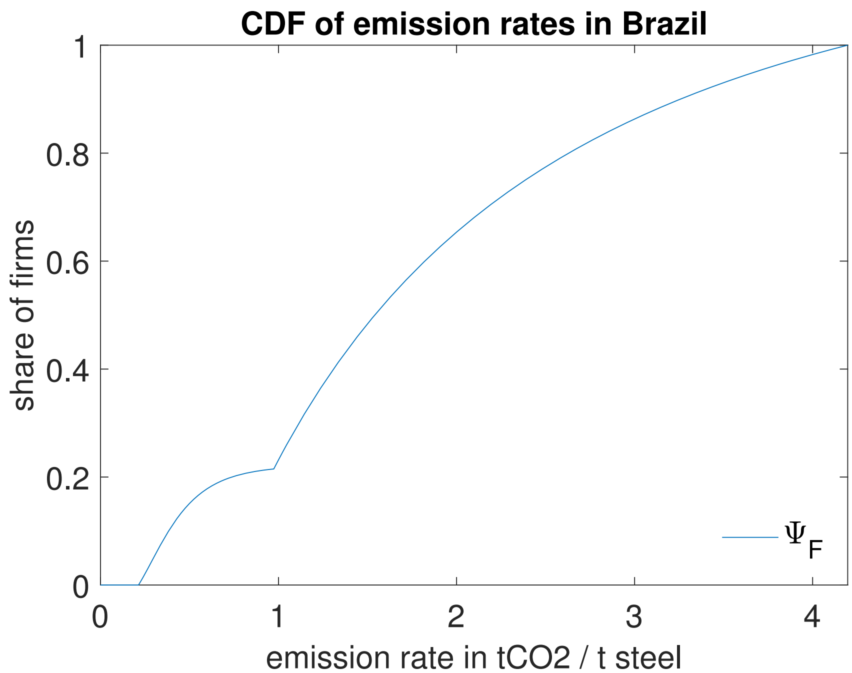
Details on the data. The main text describes how we calibrate production, trade flows, steel prices, transport costs, social cost of carbon and the elasticities. Table B.1 gives all parameters and their sources. In the following, we give additional information on the emission rates, the abatement technology and the certification costs.

As mentioned in the text, Hasanbeigi and Springer (2019) report the emission rates of the BF-BOF and EAF processes in Brazil in 2016. They report the same information for the following OECD countries: Canada, France, Germany, Italy, Japan, Mexico, Poland, South Korea, Spain, Turkey and the United States. We compute the mean emission rate for the OECD for each process by taking a weighted average of these numbers using the production according to each process in 2019 from World Steel (2020). From the same source we then get the total share of EAF versus BF-BOF in the OECD. Table B.1 reports the values.

To parameterize emission rate heterogeneity for a given process, we assume that for each process the distribution of emission rates is log-normal but bounded on each side. We then need to calibrate 4 parameters for each process (the two bounds, the mean μ and the standard-deviation σ of the unbounded log emission rates). We already have the average emission rate as 1 moment for each process. We then assume that the standard deviation of log emission rates for the joint distribution is equal to the standard deviation of log productivity in the basic metal products sector in Brazil (0.409 as reported in Schor (2004)). Collard-Wexler and De Loecker (2015) report a small productivity premium for the EAF process (of 0.074). We further assume that the standard deviation of log emission rate within each process is the same. This value is then given by $(0.409^2 - 0.222 \cdot (1 - 0.222) \cdot 0.074^2)^{1/2}$ where 0.222 is the share of EAF production in Brazil. This gives us one additional moment for each process. We then assume that the lower bound is equal to 2/3 of emission rate in the cleanest country as reported by Hasanbeigi and Springer (2019) (France for EAF and Canada for BOF) and the upper-bound is equal to 3/2 of the average rate in the dirtiest country (India in both cases). These moments can then be matched exactly and uniquely identify μ and σ for each parameter. Figure B.4 shows the c.d.f. $\Psi_F(e)$ for the resulting overall distribution of emission rates in Brazil.

Pinto et al. (2018) report a marginal abatement cost curve (MACC) for the steel sector in Brazil in 2010. We calibrate $b''(0)$ by matching the amount of abatement that occurs for our social cost of carbon according to our approximation $a = \tau/b''(0)$ and according to this curve for $\tau = \nu$ the social cost of carbon (51 USD). We assume that the ratio between the

Figure B.4: CDF of emission rates in Brazil



abatement cost curve and the price of steel in Brazil is constant (i.e. technology in abatement and in steel are "proportional" to each other). Using data from Instituto Aço Brasil (2016 and 2021) to compute the average price of for steel produced in Brazil in 2010 and 2019, we get that for a 51 USD tax, steel manufacturers would abate 0.169 t_{CO_2} per ton of steel. This directly gives a value for $b''(0)$ of 301.6 USD / $t^2_{CO_2}$ in 2019 (for simplicity, we ignore the distinction between EAF and BOF abatement technologies here as doing so needlessly complicates the computation of the equilibrium).

As described in the text, we use an EPA study of the iron and steel sector in the US (Gallaher and Depro (2002)) who found that the annualized cost of monitoring hazardous air pollutants (manganese, lead, benzene, etc. but not CO_2) for one plant were \$1.04M in 2001 (Table 3.5). These pollutants are generated by the BOF process and the monitoring costs are computed for 18 plants which produced with the BOF process. We estimate that these plants produced in 2001 53.4 Mt of steels with the BOF process.[41] We then get a cost of certification in 2001 USD per ton of steel of $1.04 \times 18/53.4 = 0.35$. To convert into 2019 values, we use the US GDP deflator and get a cost of certification of 0.49 USD per ton in

[41] We use information from their Table 2.1, remove the production of 2 additional plants which closed and are not included in the computation of monitoring costs and further adjust for a small amount of steel produced with EAF at the 18 plants.

Table B.1: Parameters and Sources for the Steel Numerical Example

Parameter	Value	Source
OECD production	480.5 Mt	World Steel (2020)
Brazil consumption	22.1 Mt	Instituo Aco Brazil (2021)
Brazil net exports to OECD	8.5 Mt	Instituo Aco Brazil (2021)
Price for Brazil exports to OECD	489 USD	USGS
Share of EAF in Brazil	0.222	World Steel (2020)
Share of EAF in OECD	0.454	World Steel (2020)
EAF av. emission rate (Brazil)	0.46 t CO_2 / t	Hasanbeigi and Springer (2019)
BOF av. emission rate (Brazil)	2.07 t CO_2 / t	HS 2019
EAF av. emission rate (OECD)	0.66 t CO_2 / t	HS 2019
BOF av. emission rate (OECD)	2.02 t CO_2 / t	HS 2019
Minimal EAF rate	2/3*0.32 t CO_2 / t	2/3 of France's rate (HS 2019)
Minimal BOF rate	2/3*1.46 t CO_2 / t	2/3 of Canada's rate (HS 2019)
Maximal EAF rate	3/2*1.62 t CO_2 / t	3/2 of India's rate (HS 2019)
Maximal EAF rate	3/2*2.80 t CO_2 / t	3/2 of India's rate (HS 2019)
St. dev. of ln(prod.)		
in metal sector in Brazil	0.409	Schor (2004)
Log prod. premium of EAF	0.074	Collard-Wexler and de Loecker (2015)
Social cost of carbon	51 USD per ton	US administration
Transport cost of carbon	50 USD per ton	
Slope of M.A.C curve, $b''(0)$	301.6 USD per t CO_2^2	Pinto et al.(2018) + own computation
certification cost (all economy)	16.1 M USD	EPA (2002) + own computation
OECD (US) demand elasticity	-0.306	Fernandez (2018)
Brazil demand elasticity	-0.414	Fernandez (2018)
Supply elasticity	3.5	EPA (2002)

2019. We assume that the same cost of certification per ton of steel in Brazil. Combining it with total Brazilian production, we get an estimate for F the cost of certifying the entire Brazilian industry.

Additional results. Table B.2 reproduces the first row of Table 2 for different parameterization and scenarios. As before, the Table reports welfare gains relative to a world where Home does not impose any trade tax on Foreign and Foreign is in laissez-faire. The Table adds one column for the welfare gains under the optimal output-based tariff which differs from a CBA for the reason discussed in Section 3.3 (at first order the optimal output tax is $t^* = s_F^{'}(p_0 - \kappa)/\left(s_F^{'}(p_0 - \kappa) - D_F^{'}(p_0 - \kappa)\right) vE_F(e)$. In all cases (except 7 for obvious reasons detailed below), the optimal output-based tariff is close to the CBA because the demand elasticity in Brazil is small relative to the supply elasticity.

Case 2 in Table B.2 is one where emission heterogeneity for the BOF process increases in Brazil (the standard deviation of log productivity increases by 50%, the lower bound of emission decreases by 50% and the upper bound increases by 50%). This does not change welfare calculations for an output tax which does not rely on heterogeneity. However, it increases the effect of voluntary certification with larger gains under the optimal certification tax but larger losses (relative to the CBA) without a certification tax.

A higher supply elasticity in Brazil (case 3, the elasticity doubles) makes welfare more sensitive to the policy in place. The welfare gains increase in all scenarios. In particular, it increases the reallocation effect which boosts welfare under voluntary certification but also the magnitude of the backfilling effect which reduces welfare in the absence of a certification tax.

A lower demand elasticity in Brazil (case 4, the elasticity is divided by 2) reduces the consumption leakage effect which increases welfare under the CBA. It does not change the effect of certification much: it decreases the magnitude of the consumption leakage effect, which is negative under the optimal certification tax but positive without certification tax.

A decrease in abatement costs (case 5, abatement costs are reduced by 25%) increases the benefits from certification. As a result, voluntary certification no longer leads to welfare losses in the absence of a certification tax but to small welfare gains instead (3 M).

Case 6 assumes that exports to the OECD increases by 50% though production stays constant. This makes Brazilian steel even more dependent on international markets. As a result, the voluntary certification program (with or without the optimal tax) leads to significantly larger welfare gains. As the share of firms certifying is higher, the reallocation effect and the abatement gains are significantly higher.

Case 7 assumes that the tax rate on taxed emission is lower than the social cost of carbon : the true social cost of carbon is $\nu = \$102$ but Home still uses the baseline tax rate $\tau_H = \tau_F = \$51$. In that case, voluntary certification brings large welfare gains since it reduces emissions. The welfare gains are actually larger without a tax on certification $f = 0$, than when the certification tax follows the formula $f^* = \tau_F \left(\frac{s_F'(p_0-\kappa)(1-\Psi_F(\hat{e}))}{s_F'(p_0-\kappa)(1-\Psi_F(\hat{e}))-D_F'(p_0-\kappa)} E_F(e|e>\hat{e}) - \hat{e} \right) s_F(p_0 - \kappa)$ since $\tau_F \neq \nu$ (implying that the certification tax is not set optimally). Correcting the social cost of carbon and implementing the true optimal CBA (with $\tau_F = \nu$) now leads to large welfare gains, and the first best which also corrects the home tax to even larger gains.

Finally, in case 8, we calibrate the Home country to the US only instead of the whole OECD. We treat trade with the rest of the world as exogenous as we did for non-OECD countries before. The welfare gains are a bit smaller since the US matters less for Brazilian steel than the whole OECD, but the relative welfare gains of the different policy program remain similar.

Table B.2: Parameters and Sources for the Steel Numerical Example

	Border carbon adjustment	Optimal output tax	Voluntary Certification $f = 0$	$f = f^*$	First best
1. Baseline	714	719	708	866	1212
2. More heterogeneity in Brazil BOF	714	719	561	884	1411
3. x2 supply elasticity in Brazil	1410	1413	1342	1706	2203
4. /2 demand elasticity in Brazil	750	751	723	914	1212
5. -25% abatement costs in Brazil	714	719	717	875	1245
6. 50% more trade	728	731	818	928	1212
7. Same τ_F but x2 SCC	998	2878	1483	1430	4849
8. Calibration to the US	516	520	497	635	979

Note: This Table reports welfare gains in M USD relative to the case where the only policy is a unilateral domestic carbon tax in the OECD without border adjustments. f is a tax on certification, f^* follows the formula for the optimum certification tax as a function of τ_F, it is therefore the optimal certification tax except in case 7.