Adverse Selection as a Policy Instrument: Unraveling Climate Change*

Steve Cicala
Tufts University and NBER

David Hémous
University of Zurich and CEPR

Morten Olsen
University of Copenhagen

Click here for the latest version.
January 21, 2026

Abstract

We propose a new policy instrument that leverages adverse selection when Pigouvian policies are infeasible or undesirable. Our policy gives firms the option to pay a tax on their voluntarily disclosed emissions, or an output tax based on the average emissions among undisclosed firms. We derive sufficient statistics formulas to calculate the welfare gains relative to an output tax, and an algorithm to implement the policy with minimal information. In an application to methane emissions from oil and gas fields, our policy generates significant welfare gains. Finally, we extend our analysis to the design of international carbon policy.

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, Florian Scheuer, Joseph Shapiro, Andrei Shleifer, Jim Stock, Allison Stashko, Bob Topel, Reed Walker, Siyu Zhang, and seminar participants at LMU Munich, UT Austin, Rice, the Public Economics meetings of the 2022 NBER Summer Institute, Indiana University, 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. Despite the simplicity of economists' advice—Pigouvian taxation—the obstacles to correcting such market failures are myriad: excessive implementation costs, political opposition, and jurisdictional limitations, among others. In this paper we propose a new policy tool to overcome some of these obstacles, based on voluntary certification combined with a rolling default. 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 firms pollute at heterogeneous rates, aggregate emissions are observable, and taxing output is feasible, but directly taxing emissions is either infeasible or costly. An output tax scales back production and emissions, but falls short of the first best since it does not differentiate between high- and low-pollution firms and it fails to incentivize abatement (Cropper and Oates (1992); Schmutzler and Goulder (1997); Fullerton et al. (2000)). We devise a tax mechanism that offers the *option* to certify one's emissions rate, upon which a Pigouvian tax will be levied, combined with an output tax that tracks the average emissions rate among those choosing not to participate in the certification program. This encourages low emissions firms 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 high emissions firms 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 directly tax externalities. Undesirability may arise from costly enforcement, as the regulator must balance the harm of the externality with that of ensuring compliance (Becker and Stigler (1974); Glaeser and Shleifer (2001); Millock et al. (2002)). We show how adverse selection can be employed to optimally separate high- and low-intensity polluters so as to economize on enforcement costs. Infeasibility may result from political or legal constraints. A gradual implementation of a certification program may decrease opposition to environmental taxation by partly shielding the most polluting firms while creating a constituency of cleaner firms that favor certification over an output tax. Lack of international jurisdiction prevents countries from directly taxing emissions abroad, so that environmental tariffs have often been thought of as output taxes. We show how a certification program that favors low-pollution foreign firms allows a government to better target externalities occurring outside of its borders.

In Section 2 we introduce the voluntary certification mechanism in a closed-economy model. Perfectly competitive firms are identical except for their emissions rate. We characterize the equilibrium and show that certification generates welfare gains relative to an output tax through two main channels: reallocation as low-emission firms expand and high-emission firms contract, and abatement. We then employ a sufficient statistics approach to approximate the emissions and welfare gains under certification relative to an output tax. These formulas highlight that the reallocation gains depend on the variance of emissions rates revealed through certification and the slope of the supply function, while the abatement gains depend on the share of certifying firms, their output, and the abatement technology.

When the monitoring and enforcement costs associated with emissions-level taxation are smaller than the abatement gains alone, the certification program achieves full unraveling, which is also the first best. When these costs are larger than the abatement gains, full unraveling is not optimal. Perhaps surprisingly, we show that the optimal level of certification requires a tax on certification itself. This is because firms are driven in part by a tax-reducing motive that the social planner does not weigh against the social cost of certification.

While direct implementation of the optimal certification mechanism requires knowing the distribution of emissions rates, we devise an algorithm that policymakers can use to encourage optimal certification when only the mean of the distribution is known.[1] These results speak to the practicality of our certification mechanism: i) welfare gains can be estimated with a few sufficient statistics, and ii) even the optimal level of certification can be implemented with limited information (including without prior knowledge of the optimal level of certification).

As an empirical application of the closed-economy model, Section 3 uses 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, and recent estimates put the cost of methane emissions in the Permian basin at around \$4B per year (Zhang et al. (2020)). This setting is particularly well-suited for such a mechanism since output taxes are already in place (in the form of mineral royalties—extraction fees paid to the government), and emissions rates across wells are highly heterogeneous (Robertson et al. (2020)).

We find that the marginal benefits from certification are increasing in the share of sites opting into the emissions tax, so that our certification program leads to complete unraveling. For the 2019 vintage of wells, this generates \$2B more in welfare benefits than an output tax reflecting average emissions. Relative to laissez-faire, the certification program reduces

[1] The updated output tax can be calculated because aggregate emissions are known, and subtracting the contribution of certified firms reveals the average contribution among those who remain uncertified.

emissions by 70% while an output tax only achieves a 9% reduction. Welfare gains and emissions reduction are driven by a combination of abatement incentives and a reallocation of production away from the dirtiest sites.

Finally, in Section 4, we 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)), our approach focuses on the direct interactions between a Home government and foreign firms whose emissions disclosures are voluntary. International sovereignty restricts what governments can mandate of foreign firms, but it 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 based on 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), 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 expand production and adopt incentives to abate emissions that would otherwise be lacking under a BCA regime.

The key complication in the international setting is that foreign firms serving the foreign market 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 when an unraveling mechanism is preferable to an output-based BCA on imports.[3]

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), 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 (Jin and Leslie (2003); Lewis (2011)).[4] This unraveling is incomplete when it is costly to

[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. Hsiao (2021) and Harstad (2022) engage with commitment issues in the context of BCA-type trade policy to address resource conservation. See also Elliott et al. (2010); Larch and Wanner (2017); Fowlie et al. (2021).

[3] To illustrate our certification mechanism in an international context, we present a calibration of our model to the case of trade in steel between the OECD and Brazil in Appendix D.

[4] It has also been used by firms to improve risk selection for credit (Einav et al. (2012)), improve safety

verify disclosures (Townsend (1979)). We apply these principles to overcome obstacles to the implementation of Pigouvian policies. 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. 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.

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)). In such settings the government creates hurdles so uptake is limited to those who value benefits more than the ordeal of enrollment (Nichols and Zeckhauser (1982)). As shown in González (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.

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)) 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.[5] 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. Recent work on voluntary environmental regulation notes the improved enforcement targeting for uncertified firms (Foster and Gutierrez (2013, 2016)), but does not consider changing audit probabilities as an instrument to encourage certification.

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. The "Building Back Better" package that passed the U.S. House of Representatives in 2021 included a "Methane Fee" drawn from this work as a voluntary methane certification program with a floating royalty-adder (Whitehouse (2021)).

(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)). See also Pram (2021), who shows in a broad class of adverse-selection environments that allowing verifiable disclosure can be Pareto-improving.

[5]See Segerson (1988); Xepapadeas (1991); Laffont (1994), and Xepapadeas (2011) for a review.

Internationally, the European Union is developing a "Carbon Border Adjustment Mechanism" (CBAM) that is based on the voluntary emissions certification of foreign firms and high default rates otherwise (Council of the European Union (2022)).[6] Given the potential challenges of certifying emissions in foreign jurisdictions, the concrete progress made by the EU suggests that such practical issues are not insurmountable. At the same time, our theoretical results that certification in that context need not bring welfare gains over an output-based tariff offer a cautionary note on the CBAM approach.

2 Unraveling in the Domestic Case

We examine problems characterized by four key features: (i) firms emit pollution at varying rates; ii) taxing output is feasible, iii) directly taxing emissions is either infeasible or costly, and iv) firms can voluntarily certify their emissions, making emissions-based taxation possible. We assume that certification is verifiable, albeit potentially costly (Townsend, 1979). For simplicity, we start with a closed economy under perfect competition, in which firms differ only in their emissions rates. Assumptions ii) and iii) are motivated both by the applications we study, as well as the broader empirical observation that output taxes (in the form of mineral royalties, industrial surcharges such as the Superfund, permitting fees, or bonding requirements) are ubiquitous in settings where Pigouvian taxes are challenging to implement. Voluntary certification offers two main advantages in such contexts. First, partial certification reduces monitoring costs, yielding welfare gains over universal emission taxation. Second, voluntary schemes can navigate legal and political barriers that might block direct emissions taxation. We develop these arguments in Sections 2.6 and 3 (for our methane leaks application).

We first present our certification mechanism and evaluate its welfare implications at different levels of certification. Second, we use a "sufficient statistics" approach and derive approximations for the welfare gains from certification, expressed in terms of easily estimable parameters, such as emission-rate variances and supply elasticities. Third, we derive the optimal certification level. Fourth, we relax informational assumptions, demonstrating that policymakers can implement the program effectively with knowledge only of the average emissions rate. Finally, we extend our baseline analysis to scenarios incorporating heterogeneous productivity and firm entry. These extensions yield expressions that share a common fundamental structure with the benchmark model.

[6] In particular, draft language states that default values shall equal mean emissions by country-good pair 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."

2.1 Baseline model

A representative agent has preferences represented by the 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 a non-polluting good with a price normalized to 1. G denotes total emissions from producing good C. The marginal social cost of emissions is v. We assume that the utility function is twice continuously differentiable, strictly increasing, strictly concave (with $u''(C) < 0$ for $C > 0$), and satisfies the Inada conditions ($u'(0) = \infty$ and $\lim_{C\to\infty} u'(C) = 0$). The outside good is produced with labor, which is inelastically supplied.[7]

The polluting good is produced by an (exogenous) mass 1 of firms who operate under perfect competition.[8] All firms share an identical cost function $c(q)$ with $c \in \mathcal{C}^2$, $c(0) = c'(0) = 0$, and $c''(q) > 0$ for $q > 0$. Firms differ, however, in their emissions rate. The pre-abatement emissions rate per unit produced is denoted by e and follows the cdf $\Psi(e)$, with pdf of $\psi(e) > 0$, on the domain $[\underline{e}, \bar{e}]$ where $\underline{e} \geq 0$ and $\bar{e}$ is finite. We impose throughout that Ψ satisfies the decreasing residual mean life property: denoting by E the expectation operator, the function $E[e|e > \hat{e}] - \hat{e}$ is decreasing in $\hat{e}$. Though not central to our mechanism, this assumption is sufficient to ensure a unique equilibrium, which simplifies the analysis.[9]

Firms can undertake costly abatement to reduce emissions: a firm can lower its per-unit emissions by an amount a at a cost $b(a)$ per unit of output, where the abatement cost function $b(a)$ is strictly increasing, strictly convex, and $b \in \mathcal{C}^2$ with $b(0) = 0$ and $b''(0) > 0$. To ensure that no firm fully eliminates its emissions, we assume $\lim_{a\to\underline{e}} b'(\underline{e}) = \infty$—alternatively, we could simply assume that $b'(\underline{e}) > \tau$ for all tax rates τ considered. Unless otherwise specified, we allow for the possibility that $b'(0) > \tau$ in which case abatement is not cost-effective. We discuss proportional abatement as an extension in Section 2.5.

While the overall distribution of emissions, Ψ, and the production of each firm are observable, the emissions of an individual firm and its abatement efforts are private information

[7] As usual in models with quasi-linear utility, we assume that the consumption of the non-polluting good C_0 is always positive. This is the case provided that its production is large enough given our assumption that $\lim_{C\to\infty} u'(C) = 0$.

[8] With imperfect competition, environmental policy would interact with pre-existing distortions; Pigouvian taxation need not be optimal and moving from an output tax to our certification mechanism would create additional welfare effects. See footnote 16 below.

[9] The decreasing residual mean life property is satisfied by distributions with an increasing hazard ratio, such as the uniform or a right-truncated exponential, the Gamma and Weibull distributions (for shape parameters higher than 1).

(unless the firm is certified as described below).

2.2 Equilibrium with certification

Policy and equilibrium. We introduce a *voluntary* certification mechanism, in which each firm can choose between two taxation schemes. A firm that voluntarily certifies and verifiably reports its emissions pays an emissions tax at rate τ per unit of emissions. If a firm chooses not to certify, it pays an output-based tax, t, equal to the average emissions rate of all uncertified firms: $t = \tau E\left[e|R\right]$, where R denotes the set of uncertified firms. This structure ensures that the policy reallocates tax burdens based on improved information about emissions rather than merely adjusting the overall level of taxation. Section 2.3 demonstrates that this choice of t is indeed optimal.

Let the market price be p. Certified firms solve the profit maximization problem:

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

which leads to a common abatement level $a^* = 0$ if $b'(0) > \tau$ and $a^* = b'^{-1}(\tau)$ otherwise. We denote by $A(\tau) \equiv \tau a^* - b(a^*)$, the per-unit cost reduction associated with optimal abatement. We then obtain the individual supply function $s\left(p - \tau e + A(\tau)\right)$ – where we define s as $s(\tilde{p}) \equiv c'^{-1}\left(\max\{\tilde{p}, 0\}\right)$ to accommodate for exit when high emissions firms face a prohibitive tax. Defining the profit function $\pi(p) \equiv ps(p) - c(s(p))$, we have profits of certified firms as $\pi\left(p - \tau e + A(\tau)\right)$. Given the convexity of the cost function, the supply curve slopes upward, making the profit function convex and increasing in p. Certified firms' profits are decreasing in the emissions rate e (provided that $\tau > 0$ and firms do not exit).

In contrast, uncertified firms do not abate. They all produce $s(p-t)$ and earn profits $\pi(p-t)$.

Certification involves a cost $F \geq 0$ (e.g., third-party monitoring or auditing expenses) and the government may tax or subsidize certification with $f \lesseqgtr 0$. Since the private gains from certification decrease with a firm's emission rate, there exists a threshold $\hat{e}$ such that all firms with $e < \hat{e}$ certify and firms with $e > \hat{e}$ do not. When total certification costs $F + f$ are sufficiently high, no firms certify ($\hat{e} = \underline{e}$). When they are sufficiently low, all do ($\hat{e} = \bar{e}$). In an equilibrium with partial certification, the threshold emissions rate $\hat{e}$ is implicitly defined by:

$$\pi\left(p - \tau\hat{e} + A(\tau)\right) - (F + f) = \pi(p - t), \tag{2}$$

with the output tax for uncertified firms defined as $t = \tau E[e|e > \hat{e}]$, which increases as more firms certify: $\partial t/\partial \hat{e} > 0$.[10] For a given tax rate τ, the certification tax f allows the policymaker to decentralize any $\hat{e}$, including corner solutions.[11]

To streamline the discussion, we introduce the effective emission rate ε, defined as the emission rate at which each firm is effectively taxed:

$$\varepsilon = \begin{cases} e & \text{if} \quad e \leq \hat{e} \\ E\left[e|e > \hat{e}\right] & \text{if} \quad e > \hat{e} \end{cases}, \text{ where } E\left[\varepsilon\right] = E\left[e\right]. \tag{3}$$

We can then write total production as

$$Q = \int_{\underline{e}}^{\hat{e}} s\left(p - \tau e + A\left(\tau\right)\right)\psi(e)de + (1-\Psi(\hat{e}))s\left(p - \tau E\left[e|e > \hat{e}\right]\right) = E\left[s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A\left(\tau\right)\right)\right]. \tag{4}$$

The indicator function $\mathbb{I}_{e\leq\hat{e}}$ equals 1 for certified firms ($e \leq \hat{e}$) and 0 for uncertified firms. This recovers the case of an emissions tax for $\hat{e} = \bar{e}$ and of an output tax for $\hat{e} = \underline{e}$.

The level of emissions is then obtained as:

$$G = E\left[\left(\varepsilon - \mathbb{I}_{e\leq\hat{e}}a^*\right)s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A\left(\tau\right)\right)\right]. \tag{5}$$

We will later exploit the following property: In the special case of linear supply ($s\left(p\right) = \tilde{s}p$), prohibitive abatement ($A\left(\tau\right) = 0$), and no exit, aggregate supply is given by $Q = \tilde{s}\left(p - \tau E\left[\varepsilon\right]\right) = \tilde{s}\left(p - \tau E\left[e\right]\right)$, which is independent of the certification threshold $\hat{e}$. Therefore, certification reshuffles production between firms of different emission rates while leaving aggregate supply unchanged.

The equilibrium price follows from utility maximization $u'(C) = p$, which leads to a demand function $C = D\left(p\right)$, and market clearing $C = Q$. In Appendix A.1.1, we show that:

Lemma 1. *The equilibrium exists and is unique.*

Changes in welfare and emissions. We now derive the consequences for welfare and emissions of implementing the voluntary certification policy. We compare outcomes relative to the output tax to focus on settings where it is the default policy option. For a given level of aggregate production, the voluntary certification mechanism increases welfare and reduces

[10] As long as $F+f \geq 0$ no firm would decide to certify and then exit, so all certified firms produce. However, uncertified firms may exit if $p - t \leq 0$. In that case $\hat{e}$ is defined through $\pi\left(p - \tau\hat{e} + A\left(\tau\right)\right) - \left(F + f\right) = 0$.

[11] Should the social planner only be able to set an output tax, they would choose $t = vE\left[e\right]$, while if they were able to directly tax emissions, they would set $\tau = v$.

emissions through two channels: the reallocation of production toward cleaner firms, and abatement. However, aggregate production generally adjusts and abating firms expand supply.

For any variable x, we let x^V denote its value with certification and x^U its value under the output tax without certification. Proposition 1 provides a decomposition of the welfare changes, characterizes conditions for which certification enhances welfare, and derives a simple formula for the linear case. To simplify exposition, we focus here on the case where emissions are taxed at the level of their social costs: $\tau = v$, and discuss the case $\tau \neq v$ further below.

Proposition 1. *Assume that $\tau = v$. The difference between social welfare under certification and under the output tax is given by:*

$$\begin{aligned} W^V - W^U &= \underbrace{E\left[\pi\left(p^V - \tau\varepsilon\right)\right] - \pi\left(p^V - \tau E\left[e\right]\right)}_{\text{Output Reallocation Effect}} \\ &\quad + \underbrace{\Psi\left(\widehat{e}\right) E\left[\pi\left(p^V - \tau e + A(\tau)\right) - \pi\left(p^V - \tau e\right) | e \leq \widehat{e}\right]}_{\text{Abatement Effect}} \\ &\quad + \underbrace{\int_{p^U}^{p^V} \left(s\left(p - \tau E\left[e\right]\right) - D\left(p\right)\right) dp}_{\text{Price Effect}} - F\Psi\left(\widehat{e}\right), \end{aligned} \tag{6}$$

where the "Output Reallocation Effect" is positive for $\hat{e} > \underline{e}$, the "Abatement Effect" is positive when abatement takes place and the "Price Effect" is weakly positive.

For linear supply and demand curves and no exit ($\tau\bar{e} < p$ is a sufficient condition)*, the welfare change can be written as:*

$$W^V - W^U = \frac{\tilde{s}\tau^2}{2} Var\left(\varepsilon\right) + \Psi\left(\widehat{e}\right)\tilde{s}A(\tau)\left(p^V - \tau E\left(e|e \leq \widehat{e}\right) + \frac{A(\tau)}{2}\right) + \frac{\left(\tilde{s}\Psi\left(\widehat{e}\right)A(\tau)\right)^2}{2\left(\tilde{s} + d_1\right)} - F\Psi\left(\widehat{e}\right). \tag{7}$$

Proof. See Appendix A.1.2. □

The *Output Reallocation Effect* in equation (6) is a direct effect of voluntary certification. Reallocating production from dirty to clean firms raises the aggregate producer surplus due to the convexity of the profit function. Second, the *Abatement Effect* captures the positive welfare effects of incentivizing abatement: certifying firms receive a higher net price after abatement and increase their profits. Third, the aggregate supply response of firms leads to a general equilibrium effect, *the Price Effect*, which is always positive.[12] Finally, the term $-F\Psi(\widehat{e})$ captures the social costs of certification and is negative. Thus, at the Pigovian tax level ($\tau = v$), voluntary certification always brings welfare gains *gross* of the certification costs. Note that with linear damages and a Pigovian tax rate, changes in emissions do not affect welfare: as emissions are always taxed at their social costs on average regardless of the certification threshold, the tax revenues collected by the government always exactly compensate for the social costs of emissions.

For further intuition, consider the special case of linear supply curves (with no exit). Then, the profit function π is quadratic, so that the reallocation of production raises the aggregate producer surplus in proportion of $Var(\varepsilon)$. The reallocation of production is proportional to $\tilde{s}$, the slope of the supply function, leading to the first term in (7). Without abatement, voluntary certification rearranges production among firms without altering aggregate supply, so the abatement and price effects disappear. As a result, the welfare change gross of certification solely depends on this variance and the slope of the supply function, $\tilde{s}\frac{\tau^2}{2}Var(\varepsilon)$. With abatement, there are additional welfare gains due to abatement directly and to the price effect. These gains scale with the cost saving engendered by abatement, $A(\tau)$, while the price change follows a classic incidence formula: $p^V - p^U = -\frac{\tilde{s}}{\tilde{s}+d_1}\Psi(\hat{e})A(\tau)$, which appears in equation (7).[13]

We now look at the effect on emissions. Using equation (5), we derive:

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

$$
\begin{aligned}
& G^V - G^U \\
= \; & \underbrace{Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right]}_{\textit{Reallocation Effect}} + \underbrace{E[e]\left\{E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E[e]\right)\right\}}_{\textit{Reallocation Rebound + Price effects}} \\
& + \underbrace{\Psi(\hat{e})\,E\left[-a^* s\left(p^V - \tau e + A(\tau)\right) + e\left(s\left(p^V - \tau e + A(\tau)\right) - s\left(p^V - \tau e\right)\right)|e < \hat{e}\right]}_{\textit{Abatement Effects (Direct + Rebound)}},
\end{aligned}
\tag{8}
$$

[12] This term is best understood moving from the certification equilibrium back to an output tax: When certification becomes unavailable, aggregate producer surplus contracts and – with no price adjustment – the aggregate supply curve is given by $s\left(p^V - \tau E(e)\right)$. But p^V is not the market clearing price for such a supply curve, and there is either excess supply or excess demand. In a closed economy, there is no external market that can absorb that excess supply or compensate for an excess demand, so moving toward the equilibrium price involves additional welfare losses.

[13] Technically a linear demand function requires a finite satiation point, which is inconsistent with our initial assumption that u is strictly concave and increasing. However, that assumption can be weakened to u strictly concave and increasing up to a satiation point.

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, but assuming no exit, emissions decline when i) s is weakly concave in p and there is no abatement or ii) s is weakly convex in p but $es\left(p^V - \tau e\right)$ is concave and increasing in e.

2) When the supply and demand functions are linear ($s(p) \equiv \tilde{s}p$ and $D(p) \equiv d_0 - d_1 p$) and there is no exit, the change in emissions is given by

$$G^V - G^U = -\tilde{s}\left(\begin{array}{c}\underbrace{\tau Var(\varepsilon)}_{Reallocation} + \underbrace{\Psi(\hat{e})\left[a^*\left(p^V - \tau E[e|e<\hat{e}] + A(\tau)\right) - E[e|e<\hat{e}]A(\tau)\right]}_{Abatement} \\ \underbrace{+E[e]\frac{\tilde{s}\Psi(\hat{e})A(\tau)}{\tilde{s}+d_1}}_{Price}\end{array}\right). \tag{9}$$

In particular, emissions decline if either i) there is no abatement, or ii) $p^V \geqslant 2\tau\hat{e}$.

Proof. See Appendix A.1.3. □

The Remark decomposes the impact of introducing voluntary certification on emissions into three distinct channels. First, there is a *reallocation* effect: emissions decrease as output is reallocated from dirty firms to clean firms. Second, there is an *abatement* effect, which has two components: a direct reduction in emissions as certified firms actively reduce their emissions, and an offsetting rebound effect since abatement lowers certified firms' costs, incentivizing greater output. Finally, the second term combines two effects: i) a *rebound* effect from the reallocation of production, which may be positive or negative, and ii) a counteracting *price effect* that mitigates both rebound effects. The net impact is generally ambiguous, but Remark 1 provides sufficient conditions under which certification decreases emissions. Emissions decline when either (i) the supply function s is weakly concave and there is no abatement or (ii) the supply function s is weakly convex, but the product $es(p-\tau e)$ is concave and increasing in emission intensity e (i.e. a firm's overall emissions are increasing in its emission rate, though less than linearly). [14]

[14] In case i), there is no rebound effect from abatement and the rebound effect from reallocation is weakly negative ensuring that emissions decrease. In case ii), the reallocation rebound effect is weakly positive, so that the equilibrium price weakly declines ($p^V \leq p^U$). When $es(p^V - \tau e)$ is concave in e, an application of Jensen's inequality ensures that $E\left[\varepsilon s(p^V - \tau\varepsilon)\right] < E[e]\, s\left(p^V - \tau E[e]\right)$, and with $es(p^V - \tau e)$ increasing in e, we get $E\left[\left(\varepsilon - \mathbb{I}_{e\leq\hat{e}}a^*\right) s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau)\right)\right] \leq E\left[\varepsilon s(p^V - \tau\varepsilon)\right]$. Putting the inequalities together, we obtain $E\left[\left(\varepsilon - \mathbb{I}_{e\leq\hat{e}}a^*\right) s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau)\right)\right] < E[e]\, s\left(p^U - \tau E[e]\right)$, so that emissions decrease with certification.

In the case of linear supply and demand functions (with no exit) and without abatement, aggregate supply is unchanged, and the change in emissions reduces neatly to $G^V - G^U = -\tilde{s}\tau Var(\varepsilon)$: emissions decline proportionally to the variance of taxed emission rates, $Var(\varepsilon)$, the slope of the supply function, and the tax rate. With abatement, a rebound effect emerges, as certified firms reduce their costs by the amount $A(\tau)$. The (sufficient) condition that $p^V \geqslant 2\tau\hat{e}$ corresponds to the assumption in ii) that $es(p - \tau e)$ be increasing in e (while linear supply guarantees the concavity).

Taylor expansions and formulas using "sufficient statistics." Evaluating the general expressions for welfare and emissions changes (equations (6) and (8)) requires detailed and extensive information. To derive practical approximations and focus the analysis on the important effects of certification, we assume that the emissions tax rate τ is small and take Taylor expansions around $\tau = 0$. This implies i) locally linear supply and demand curves, which is the usual public finance setting to obtain the Harberger formulas (e.g. Jacobsen et al. (2020)) and ii) a locally quadratic abatement curve. In Appendix A.2.2, we characterize conditions under which the approximation will be good and derive bounds to the error.

This approach yields straightforward formulas for welfare and emissions changes that depend only on a few key parameters commonly estimable from readily available data: the slope of the supply curve, the slope of the marginal abatement cost function, the variance of firms' effectively taxed emissions rates, the certification cost and the social cost of emissions. These approximate expressions serve as useful benchmarks even when τ is not particularly small. To avoid discussing multiple cases, we assume that $b'(0) = 0$. The approximations would still hold with $b'(0) > 0$, but without the abatement terms. We focus again on the Pigovian case: $\tau = v$. We then obtain (proof in Appendix A.2):

Corollary 1. *Assume that $b'(0) = 0$ and $\tau = v$. The expression $W^V - W^U$ in Proposition 1 can be written as:*

$$W^V - W^U = \frac{\tau^2}{2}\left(s'(p_0)Var(\varepsilon) + \frac{s(p_0)\Psi(\widehat{e})}{b''(0)}\right) - F\Psi(\widehat{e}) + o(\tau^2), \tag{10}$$

and the difference in emissions is:

$$G^V - G^U = -\tau s'(p_0) Var(\varepsilon) - \frac{\tau s(p_0)\Psi(\widehat{e})}{b''(0)} + o(\tau), \tag{11}$$

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 $\hat{e}$ (to a second order).

This corollary shows that the welfare and emissions changes induced by certification can be cleanly decomposed into a direct effect of reallocation and a direct effect of abatement – rebound effects are dominated. The size of the reallocation effect depends on the supply response, $s'(p_0)$, and grows with the variance of the taxed emissions rate, $Var(\varepsilon)$, which, naturally increases as more firms certify. Abatement gains scale with output, the inverse of the slope of the marginal abatement curve, $b''(0)$, and the share of certifying firms, $\Psi(\hat{e})$.[15]

With emissions taxed at their social costs, $\tau = v$, almost all welfare gains or losses from certification accrue to firms: as already discussed, changes in tax revenues and damages exactly offset each other, and with a small price response, the change in consumer surplus is negligible compared to that of the producer surplus—though this depends on the assumption of an exogenous mass of firms (see Appendix A.5.3).[16] The mechanism therefore creates producer surplus on net by encouraging cleaner firms to expand their production, while curbing excess production from dirtier firms.

Intuitively, the most critical aspect for the approximation is that τe should be small relative to p and the elasticities of the supply function and its derivative should not be too large.[17] The assumption that $\tau e/p$ is small is satisfied for many problems in environmental economics. For instance, in our methane leaks application, the external cost is 3% of the wholesale price per barrel of oil. In the case of steel discussed in Appendix D, the average social costs of emissions are 9% of the laissez-faire price in the US, knowing that iron and

[15] Certifying firms abate an amount $a^* = \tau/b''(0) + o(\tau)$. 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. Abatement allows certifying firms to save $A(\tau) = \tau^2/(2b''(0)) + o(\tau^2)$ per unit of production. These expressions are exact for a quadratic abatement cost curve.

[16] With imperfect competition, there would be additional welfare effects. For instance with monopolistic competition, reallocation across sectors would yield extra gains if the supply of the polluting good increases with certification and mark-ups are higher in the polluting good than in the homogeneous sector. With heterogeneous mark-ups within the polluting sector, there would be additional welfare effects from the reallocation of production within that sector.

[17] This relates to three roles played by a small τ: First, if τe is sufficiently small relative to p, the supply curve can be considered as locally linear, allowing us to ignore an aggregate rebound effect from reallocation. Second, the profit gains from abatement for individual firms can be computed at fixed supply because the supply response is comparatively small. Third, the aggregate rebound effect generated by abatement is small, so that price effects are negligible. Kleven (2021) makes an analogous point in the context of income taxes: a sufficient statistic approximation will be accurate either if the tax change is small or the relevant elasticities do not change too much over the range of the tax change.

steel is the second most CO_2-intensive industry after cement. The approximation for 'small' external costs therefore remains appropriate even when the externality itself is quite large.

In addition, the expressions in Corollary 1 approximate the level of abatement and the cost gains from abatement ($a(\tau)$ and $A(\tau)$). That approximation is good provided that the abatement curve does not deviate too much from quadratic up to the equilibrium level of abatement.[18]

Equation (10) highlights that the welfare benefits from certification gross of the costs are proportional to v^2: Intuitively, with a small tax, the allocations with and without voluntary certification only differ at first order in v, bringing second order welfare gains. As a result, certification can bring welfare benefits on net only if the certification costs F are at most second order in v (that is of the same magnitude as v^2 or smaller). Otherwise, the social costs from certification $F\Psi(\hat{e})$ would outweigh the social benefits as soon as a non-negligible mass of firms certify.

How far is welfare under the voluntary certification in Proposition 1 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:[19]

$$W^V = \frac{\left(Var(\varepsilon) + \frac{s(p_0)}{s'(p_0)b''(0)}\Psi(\widehat{e})\right)}{Var(e) + \frac{s(p_0)}{s'(p_0)b''(0)}} W^{FI} + \left(1 - \frac{\left(Var(\varepsilon) + \frac{s(p_0)}{s'(p_0)b''(0)}\Psi(\widehat{e})\right)}{Var(e) + \frac{s(p_0)}{s'(p_0)b''(0)}}\right) W^U - F\Psi(\hat{e}) + o(\tau^2). \tag{12}$$

By construction $Var(\varepsilon) \leq Var(e)$ and $\Psi(\widehat{e}) \leq 1$ 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 weights reflect the relative variance of the effectively-taxed emission rate, ε, and the actual emission rates, e, and the share of certified firms.

Welfare with $\tau \neq v$. While all expressions for emissions changes remain the same when $\tau \neq v$, welfare effects are different when emissions are not taxed at their Pigovian level. As taxes and damages do not offset each other, there is an additional term in the welfare change equation (6): $-(v-\tau)\left(G^V - G^U\right)$. This represents an *underinternalized emissions effect*, which is positive if emissions are undertaxed and decreases with certification. We

[18] In Appendix A.2.2, we show that for perfectly elastic demand, $W^V - W^U + F\Psi(\hat{e}) \approx \frac{1}{2}s'(p_0)\tau^2 Var(\varepsilon) + \Psi(\hat{e})s(p_0)A(\tau)$ for small $\tau e/p$. That expression does not rely on any approximation to the level of abatement. Analogous expressions can be derived for imperfectly elastic demand.

[19] To see this, note that the full information case corresponds to a special case of the certification policy with $\hat{e} = \bar{e}$ and $F = 0$. We can then use (10) to get $W^{FI} - W^U = s'(p_0)\frac{\tau^2}{2}\left(Var(e) + \frac{s(p_0)}{s'(p_0)b''(0)}\right) + o\left(\tau^2\right)$, which combined with (10) for a general $\hat{e}$ delivers (12).

can approximate the welfare change for $\tau \neq v$—noting that the expression for the emission change (11) remains valid. If the social cost v is either large or small relative to the tax rate τ, then the underinternalized emissions effect trivially dominates (e.g. a reduction in emissions necessarily enhances welfare if $v >> \tau$). When v and τ are of the same order of magnitude, the welfare change can be approximated as:

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right)\left(s'(p_0)Var(\varepsilon) + \frac{s(p_0)\Psi\left(\widehat{e}\right)}{b''(0)}\right) - F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \qquad (13)$$

reflecting a combination of the effects of reallocation and abatement on profits and on the underinternalized emissions. Thus the mechanism continues to deliver benefits when the externality is over-taxed, up to double the marginal social cost.

2.3 Optimal constrained policy

Having established that a certification program trades off welfare gains from the reallocation of production and abatement with the social costs of certification, we now examine the optimal level of certification. We consider a social planner who has access to three instruments: an emissions tax for certified firms, an output tax for uncertified firms and a tax on certification. That is, we relax the constraint that $t = \tau E\left[e|R\right]$, but maintain the assumption that certification is voluntary. [20] As in Section 2.2, there is a unique threshold $\widehat{e}$ such that firms with an emission rate below $\widehat{e}$ certify and firms with an emission rate above $\widehat{e}$ do not. We can derive (proof in Appendix A.3):

Proposition 2. *A social planner using an output tax, an emissions tax upon voluntary certification, and a certification tax will set an emission tax at the Pigouvian level $\tau = v$, a Pigouvian output tax using the average emission rate of the uncertified firms $t = vE\left[e|e > \widehat{e}\right]$, and a tax on certification given by*

$$f = vE\left(e - \widehat{e}|e > \widehat{e}\right) s\left(p - vE\left(e|e > \widehat{e}\right)\right) \geq 0. \qquad (14)$$

Further, if $b'\left(0\right) = 0$, then for a small social cost of emissions, v, there is full (or near full) unraveling if $F < \frac{v^2}{2}\frac{s(p_0)}{b''(0)}$, no (or near no-) certification if $F > \frac{v^2}{2}\left(s'\left(p_0\right)\left(E\left[e\right] - \underline{e}\right)^2 + \frac{s(p_0)}{b''(0)}\right)$,

[20] While the fully optimal mechanism in our environment would generally be nonlinear, we restrict attention to these instruments because they are institutionally realistic: they are transparent, easy to administer, and rely on few observables. This perspective is common in the contract theory literature, which emphasizes the practical appeal of simple, low-dimensional incentive schemes when institutions constrain what can be implemented (e.g. Laffont and Tirole (1993) Chapter 2).

and otherwise an interior optimal level of certification that satisfies:

$$\frac{v^2}{2} s'(p_0)\left(E\left[e|e>\widehat{e}\right]-\widehat{e}\right)^2+\frac{v^2}{2}\frac{s(p_0)}{b''(0)}=F+o\left(v^2\right). \tag{15}$$

Proposition 2 first establishes the standard Pigouvian result that emissions ought to be taxed at their (expected) social costs: $\tau = v$ and $t = vE(e|e > \hat{e})$—so that imposing the constraint that $t = \tau E[e|R]$ is in line with the optimal use of these instruments. Perhaps more surprisingly, equation (14) shows that certification is excessive when τ and t are set optimally, and should be *taxed*—i.e., on top of the social cost F, certifying firms need to pay f, which is positive unless full unraveling is optimal. The certification fee in (14) is equal to the reduced tax bill of the marginally-certified firm (holding quantity fixed). Intuitively, when an additional firm certifies, it reduces its tax burden and increases its scale at the expense of dirtier firms. This reallocation of production is a source of social gains, but the reduced tax burden is not and certification is a source of social cost. Thus the central planner levies a tax on certification to remove the tax-shifting incentive from the socially costly certification decision.

The optimal level of certification equates the marginal gain in allocative efficiency plus the marginal gain from abatement to the marginal social cost of certification, F. When there is no social cost of certification ($F = 0$), complete unraveling is optimal and certification is untaxed. For a small marginal social cost of emissions v, the welfare benefits from certification take a simple expression (equation (10) with $\tau = v$), and one can characterize the optimal $\widehat{e}$, implicitly given by (15) or a corner solution. In (15), the first term on the left-hand side represents the marginal gain in allocative efficiency from increasing $\widehat{e}$ and the second term the marginal gains from abatement.

This expression also clarifies conditions on the size of F under which certification is desirable. If the certification costs are large relative to the social gains (e.g. F is first order in v or larger), then certification is not socially desirable even if, absent a certification tax, the cleaner firms would like to certify. If the certification costs are smaller than the abatement gains alone (e.g. F is third order in v or smaller), then full unraveling is desirable: the social planner would like to mandate an emission tax but can use the certification mechanism if there are obstacles to direct implementation (see Section 2.6). Finally, if certification costs are commensurate with the social gains and larger than the abatement gains alone (F is second order in v but higher than $\frac{v^2}{2}\frac{s(p_0)}{b''(0)}$), then the optimal certification threshold may be interior, in which case the certification mechanism brings social gains over a mandated universal emissions tax. One can then derive the optimal certification level $\widehat{e}$ (and the optimal certification tax f) using (15) with data on the certification costs, the supply curve, the slope of the marginal abatement curve, and knowledge of the emission distribution.

2.4 An "Unraveling" Algorithm

Having solved for the decentralized equilibrium and the social planner's allocation, we now assess the informational requirements needed to implement such a policy. Equation (10) gives an intuitive result of the welfare gains based on statistics that are readily available. However, implementing the certification mechanism requires complete information on the distribution of emission rates e to determine the output tax t and (potentially) the certification tax f. Since this distribution may be unknown in practice, we introduce an iterative algorithm that allows policymakers to achieve the same outcome without initial complete information.

We assume that neither firms nor the government knows the distribution of emissions rates, but they do observe the average emissions rate (e.g., through aggregated accounts or ambient pollution levels). Initially, certification is not available and the government imposes an output tax $t_0 = \tau E(e)$. Subsequently, certification becomes available which allows firms to pay the emission tax τ alongside the certification cost F plus potentially a certification tax 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 $E\left(e|e > \widehat{e}_{n-1}\right)$.[21] 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, and firms decide whether to certify or not by comparing profits under certification $\pi\left(p - \tau e\right) - F - f$ 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 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 $\hat{e}_0 = \underline{e}$)

[21] At step $n-1$, aggregate emissions are given by $G_{n-1} = \int_{\underline{e}}^{\widehat{e}_{n-1}} (e-a)\, s\left(p - \tau e + A(\tau)\right) \psi(e)\, de + \left(1 - \Psi\left(\widehat{e}_{n-1}\right)\right) s\left(p - t_{n-1}\right) E\left(e|e > \widehat{e}_{n-1}\right)$. The government can observe the (post-abatement) emissions of certified firms (which corresponds to the integral in the previous expression), aggregate emissions G_{n-1} and the output of non-certified firms $\left(1 - \Psi\left(\widehat{e}_{n-1}\right)\right) s\left(p - t_{n-1}\right)$. This is sufficient to calculate the average emission rate of non-certified firms.

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

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

In the process described above, the certification tax is fixed (potentially set to 0) and the algorithm always converges to the unique equilibrium described in Section 2.2. The logic can be extended to include an iteratively adjusted certification tax given by $f_n = \tau\left(E\left(e|e > \widehat{e}_{n-1}\right) - \widehat{e}_{n-1}\right) s\left(p - \tau E\left(e|e > \widehat{e}_{n-1}\right)\right)$. This certification tax is computable with information that the planner would have at the beginning of period n. In that case, the process converges toward a solution to the first-order conditions of the social planner problem. One complication is that the first-order conditions may not uniquely idenfity the optimum. Proposition 3 formalizes these results and gives sufficient conditions under which the algorithm with the iteratively adjusted certification tax converges toward the optimum (see proof in Appendix A.4).

Proposition 3. *Assume that prices are exogenous. Then the procedure with a fixed certification tax converges monotonically toward the unique equilibrium level of certification $\widehat{e}$. The procedure with a certification tax f_n converges monotonically toward the smallest $\widehat{e}$ satisfying the first-order conditions of the social planner problem—this is the unique solution to the social planner problem if s is weakly convex or τ is small.*

Remarkably, the certification equilibrium and the social optimum can therefore be implemented even if the government has no information on the distribution of emission rates nor knows ex-ante what the optimal $\widehat{e}$ would be. 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.[22]

2.5 Extensions

This subsection considers three extensions to the baseline model: (1) proportional abatement, (2) productivity heterogeneity, and (3) free entry. The extensions fit easily in the framework presented thus far, and we analyze them with a focus on the added terms to the emissions and welfare expressions.

[22] 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.

Proportional abatement So far, we have considered an abatement technology where firms reduce their emission rates by a fixed absolute amount, a, independent of their initial emission rates. As an alternative, we assume that abatement reduces the emissions rate of firms by a share a—that is, by spending $b(a)$ per unit, a firm reduces its emission rate from e to $(1-a)\,e$. In this case, the optimal abatement level increases with the pre-abatement emission rate and is given by $a^*(e) = b'^{-1}(\tau e)$. At first order, a firm with pre-abatement emission rate e, abates a total amount $a^*(e)\,e = \tau e^2/b''(0) + o(\tau)$ per unit. The expressions (11) and (10) can then be directly applied 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 (see Appendix A.5.1). This alternative set-up is the one we use in our quantitative exercise in Section 3.

Productivity heterogeneity We now let firms have heterogeneous productivity levels. Specifically, firm i's production costs are $c(q)/\varphi_i$, 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 maintain the assumption that firms have the same abatement technology, which ensures that all certified firms still abate the same amount $a = b'^{-1}(\tau)$. Certifying firms face a price net of tax and abatement expenditures $p - \tau e + A(\tau)$ and remaining costs $c(q)/\varphi_i$. Non-certifying firms face the same output tax t.

Certified firms thus earn a profit of $\frac{1}{\varphi}\pi\left(\varphi\left(p - \tau e + A(\tau)\right)\right)$ (gross of certification costs) and uncertified firms a profit of $\frac{1}{\varphi}\pi\left(\varphi\left(p-t\right)\right)$ where π is the profit function previously defined. As firms differ both in productivity and emission rates, equation (2) becomes:

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

which defines a threshold $\widehat{e}_{\varphi}$. Firms with emission rates $e < \widehat{e}_{\varphi}$ certify while other firms do not. The cut-off function $\widehat{e}_{\varphi}$ depends positively on productivity because production increases with productivity whereas the certification cost, F, does not.

We assume that uncertified firms face an output tax equal to the average quantity-weighted emission rate of uncertified firms. This is the optimal output tax only when the supply function is iso-elastic, and we restrict attention to such supply functions.[23] In that case, the approximations for the changes in emissions and welfare of Corollary 1 still ap-

[23]More generally, the optimal output tax weights firms' emission rates by the slope of their supply curve which is equivalent to weighting by quantity in the isoelastic case. Our approach can also be used with other supply functions but the analysis of the welfare gains then includes an additional term that reflects that the introduction of certification modifies the average tax rate when weighting by the slopes of firms' supply curves.

ply with the variance computed over the (laissez-faire) quantity-weighted distribution and the abatement term reflecting the (laissez-faire) quantity produced by abating firms (see Appendix A.5.2 for the proof and the mathematical expressions).

Appendix A.5.2 also generalizes the discussion on optimality from section 2.3: the optimal constrained policy is implemented with a productivity-specific certification tax $f_\varphi = (t - v\widehat{e}_\varphi)\, s\,(\varphi\,(p - t))$ which generalizes (14). In the optimum, the threshold $\widehat{e}_\varphi$ is decreasing in productivity φ as the certification of larger firms brings larger welfare gains. Moreover, the algorithm from section 2.4 also generalizes to that case: provided that aggregate emissions are observable, the output tax and the set of certification taxes can be implemented iteratively even when the distribution of emissions is unknown to the social planner.

Free entry We allow for free-entry in Appendix A.5.3. The expressions from Corollary 1 still apply (pre-multiplied by the laissez-faire mass of firms). As a result, the welfare gains remain the same as in the baseline model. Nevertheless, profits net of entry are zero in that case so that the entire welfare benefits accrue to consumers.

2.6 Advantages of the certification mechanism in the domestic context

Political Economy: Environmental and carbon taxes have been met with strong opposition, which has often succeeded in denying their implementation. In Appendix A.6, we formalize an argument to explain why our certification mechanism offers a "gradualist" approach which may encounter less resistance.[24] We first model why output taxes appear to garner less opposition than emissions taxes. Consider an economy in laissez-faire and a government that wishes to introduce either an output tax or an emission tax. Firms lobby and potentially block policy reforms. While both output and emissions taxes reduce firms' profits, high emissions firms lose more under an emission tax. If lobbying efforts are sufficiently convex, then aggregate lobbying opposition will be higher for an emission tax than for an output tax, so that the output tax is more likely to be implemented.

With an output tax as the status quo, the government can use a "salami tactic" approach (Schelling (1965)) by offering either the certification mechanism or again directly moving to an emissions tax. Less polluting firms gain from the lower tax burden relative to the output tax, and lobby in favor of reform. This is one of the advantages of a gradualist approach which uses an output tax as an intermediary step: it builds a constituency for further reform. At

[24] The virtues of "gradualist" approaches have been analyzed in the context of transition economies, for instance by Dewatripont and Roland (1992, 1995). They credit gradualist approaches for building constituencies in favor of reforms and for the use of "divide-and-rule" tactics–even though, in principle a gradualist approach may be less efficient than a more immediate implementation of the optimal policy.

the same time, the most polluting firms will be partly shielded by incomplete unraveling and therefore lobby less against certification than against an emission tax. The same logic applies every time that the government raises the certification threshold: opposition to the reform becomes increasingly concentrated among a small group of (so far uncertified) firms and the constituency that stands to benefit from an emissions tax grows. Finally, note that this dynamic holds at the outset for forward-looking firms who anticipate future unraveling: the certification mechanism spreads out implementation over time, reducing ex ante opposition through interest rate and survival probability discounting.

At a more practical level, a voluntary certification mechanism exploits asymmetries in how much political consensus is required for various actions: the opt-in to reduce one's tax burden can be enacted once with a law, while output taxes (royalty fees in particular) can be updated administratively without new legislation (30 USC 226).

Enforcement Costs: A voluntary certification mechanism also delivers benefits when it is costly to monitor emissions. In such settings the social planner would like to balance the social cost of mis-taxation with that of program implementation (Millock et al., 2002). As shown in Section 2.3, this can be optimally achieved with a voluntary certification program whose participation is taxed. Relative to a mandatory Pigouvian tax that entails bearing enforcement costs for the entire regulated community, a voluntary certification program delivers net benefits by economizing on wasteful enforcement. This is especially valuable in diverse industries where the regulations that make sense for large, fixed sources of pollution would be unduly burdensome for numerous, small producers (see for instance the model with heterogeneous productivity of section 2.5).

We have thus far considered enforcement costs to be an object upon which to economize, rather than a constraint that must be respected (i.e. only a fixed share of plants can be inspected). This latter approach to enforcement costs is applicable in cases of limited state capacity. When regulatory capacity is constrained, a voluntary certification program allows inspectors to focus their efforts on uncertified firms, thereby extending the reach of their oversight. The motive to certify grows as uncertified firms are subject to more intensive scrutiny, thus using adverse selection to expand state capacity. Voluntary certification imposes relatively light information requirements on the regulator in each of these settings, as the decision of whether to certify or not comes from firms, and tax options can be constructed from average emissions rates alone.

3 Domestic Empirical Application: Methane Emissions in the Permian Basin

We now demonstrate the effectiveness of voluntary certification by examining methane leaks from oil and gas production. This section also illustrates how one can use our formulas to estimate the emissions and welfare gains from the certification program from readily available data. Subsection 3.1 details the context and explains how our baseline model must be amended to this case. Subsection 3.2 presents the empirical results.

3.1 Context and Extensive-margin Model

Context. Methane (CH_4) is the main component of natural gas. It is extracted both deliberately when drilling in gas-rich geological formations, as well as incidentally when it is co-produced with oil. When released into the atmosphere, it 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 with \$51/metric ton for carbon dioxide. Emissions occur during production through faulty equipment, pressure-relief valves, or waste disposal in oil extraction.[25]

Despite these substantial environmental costs, the state of methane regulation is in flux. A combination of tax and regulatory approaches were enacted during the Biden administration, but these have been or are in the process of being repealed during the Trump administration. Future administrations will therefore face an essentially unregulated pollutant that causes large social damages.

The oil & gas sector presents several features that make it amenable to the certification program we propose: emissions rates are highly heterogeneous, abatement strategies are economical, and 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.[26] Our certification program can therefore be implemented on top of the existing policy: firms are offered the option to subject their wells to emissions taxation or pay a royalty rate that is administratively adjusted to reflect the average emissions of uncertified wells. This suggests a fruitful path forward for methane emissions.

[25] Best practices for disposing of waste methane entail flaring 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)).

[26] 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)).

Extensive-margin model. In the oil and gas context, the reallocation of production across wells occurs at the extensive margin when a firm decides whether or not to drill a well (Anderson et al. (2018)). Our empirical analysis is therefore based on a model of the driller's decision, weighing the lifetime revenues and emissions of a well against the up-front cost of drilling. Though the microfoundations differ when all economic decisions are made on the extensive margin, the results and insights are similar to the intensive margin model of section 2.1. We therefore summarize the model here and provide complete details in Appendix B.

A unit mass of firms consider drilling in specific locations, and these potential wells differ in their drilling costs (c), output (q), prices (p), and emissions rates (e). Emissions rates depend on different factors, some known before drilling, such as distance to gas collection infrastructure, and others that are uncertain before drilling such as the exact underground structure. Accordingly, we distinguish between the expected emissions rate before drilling, $\mathfrak{e}$, and the realized emissions rate, $e = \varkappa\mathfrak{e}$, with $\varkappa \sim \Psi_\varkappa$ representing a mean-one multiplicative emissions shock. Cost, price, and output heterogeneity are known ex ante.[27] Further, firms can abate a fraction of their emissions a at cost $b(a)$ per unit of output with $b'(0) = 0$ as in (2.5), which is consistent with how abatement rates in the industry are calculated, i.e. Plant et al. (2022). We denote by u the ratio of drilling costs to revenues such that $c = upq$, and each firm is thus characterized by a tuple $(u, q, p, \mathfrak{e}, \varkappa)$ distributed according to some c.d.f Ψ. We allow q and $\mathfrak{e}$ to be dependent on each other (as empirically observed), but assume that u, p, and $\varkappa$ are independent of $(q, \mathfrak{e})$ and of each other.[28]

A firm faces four sequential decisions: i) whether to drill, given $(u, q, p, \mathfrak{e})$; (ii) whether to certify after observing the actual emissions rate e $(= \varkappa\mathfrak{e})$; (iii) whether to produce q or shut down; and (iv) how much to abate. A firm with expected emissions rate $\mathfrak{e}$ decides to drill or not based on the (pre-abatement) emissions rate at which it expects to be taxed. We denote by ε the realized pre-abatement emissions rate at which a firm ends up being

[27]Price heterogeneity reflects heterogeneity in the oil share. As oil and gas prices are internationally determined, we hold prices fixed, so that p is a firm characteristic and does not adjust to the taxation mechanism. We abstract from price and output uncertainty, which would further complicate the model without additional insight.

[28]These independence assumptions are not necessary to solve the model but serve two roles. First, the independence assumption on p allows us to identify key moments of the (unobserved) distribution of potential entrants from the (observed) distribution of actual producers. Second, the independence assumption on u implies that drilling costs scale with revenues (e.g. they scale with the number of horizontal wells, which raises production). This ensures that the optimal output tax is equal to the quantity-weighted average emissions rate times the social cost of emissions and that the optimal tax on uncertified firms similarly depends on the quantity-weighted average emissions rate of uncertified firms (see footnote 23 for a similar discussion in the case of heterogeneous productivity in the intensive margin case).

taxed under the certification program: Firms are taxed at their true emissions rate when their rate is below the certification threshold ($\varepsilon = e$ if $e \leq \hat{e}$) and at the average rate of uncertified firms otherwise ($\varepsilon = E_{\tilde{\psi}_e}[e|e > \hat{e}]$, where $\tilde{}$ denotes quantity-weights, such that $\tilde{\psi}_e$ is the quantity-weighted distribution over e).

By construction, $E_{\tilde{\psi}_e}[\varepsilon] = E_{\tilde{\psi}_e}[e] = E_{\tilde{\psi}_{\mathfrak{e}}}[\mathfrak{e}]$, and before drilling, a firm expects to be taxed at a pre-abatement emissions rate given by $m(\mathfrak{e}) \equiv E_{\psi_{\varkappa}}[\varepsilon(\mathfrak{e})]$. Importantly, generally it does not hold that $m(\mathfrak{e}) = \mathfrak{e}$: uncertified firms with a high expected rate $\mathfrak{e}$ tend to have a higher emissions rate than uncertified firms with a low $\mathfrak{e}$, but pay the same tax when not certified, $E_{\tilde{\psi}_e}[e|e > \hat{e}]$, which is independent of $\mathfrak{e}$. Certified firms abate at a rate that depends on the post-drilling emissions rate: $a(e) = b'^{-1}(\tau e)$.

We denote by S^{LF}, the aggregate supply in laissez-faire, by $\dot{S}^{LF}$ its slope with respect to a common additive price shock, and by N^{LF} the mass of firms that enter in laissez-faire. As in the intensive margin case, we derive approximations to the change in emissions and welfare assuming that v is small. We then obtain the following Proposition (proof in Appendix B):

Proposition 4. *For $\tau = v$, and assuming that F is at most second order in v, the welfare difference between certification and the output tax is given by:*

$$W^V - W^U = \underbrace{\dot{S}^{LF}\frac{v^2}{2}\left(Var_{\tilde{\psi}_e}(m(\mathfrak{e})) + 2Cov_{\tilde{\psi}_e}(\mathfrak{e} - m(\mathfrak{e}), m(\mathfrak{e}))\right)}_{\textit{Reallocation Effect}} + \underbrace{S^{LF}\frac{v^2}{2}\frac{E_{\tilde{\psi}_e}(e^2|e \leq \hat{e})\,\tilde{\Psi}_e(\hat{e})}{b''(0)}}_{\textit{Abatement Effect}} - N^{LF}\Psi_e(\widehat{e})\,F + o(\nu^2), \tag{16}$$

while the emissions difference is given by:

$$G^V - G^U = -v\dot{S}^{LF}\left[Var_{\tilde{\psi}_e}(m(\mathfrak{e})) + Cov_{\tilde{\psi}_{\mathfrak{e}}}(\mathfrak{e} - m(\mathfrak{e}), m(\mathfrak{e}))\right] - v\frac{S^{LF}}{b''(0)}E_{\tilde{\psi}_e}\left[e^2|e \leq \hat{e}\right]\tilde{\Psi}_e(\hat{e}) + o(\nu), \tag{17}$$

where $m(\mathfrak{e}) = E_{\psi_{\varkappa}}[\varepsilon(\mathfrak{e})]$ is the rate at which a firm with expected rate $\mathfrak{e}$ anticipates being taxed.

The welfare gains from certification still depend on reallocation gains and abatement and Corollary 1 is preserved with three main adjustments.[29] First, all terms become quantity-

[29] The assumption that F is second order in v simplifies the analysis by ensuring that the potential payment

weighted due to output heterogeneity. Second, reallocation is driven by the variance of the *anticipated, tax-relevant* emissions rate. In other words, firm take expectations over the emissions shock and anticipate their certification decision given $\hat{e}$. Third, there is a correction term to account for a wedge between anticipated emissions rates and the tax treatment of that anticipated emissions rate (which depends on certification). This wedge disappears for full unraveling (as $m(\mathfrak{e}) = \mathfrak{e}$), in which case the reallocation effect directly scales with the variance of expected emissions rates, and when there is no ex-post uncertainty (i.e. $\varkappa = 1$, so that $m(\mathfrak{e}) = \varepsilon$). Otherwise, the covariance term, $Cov_{\tilde{\psi}_{\mathfrak{e}}}(\mathfrak{e} - m(\mathfrak{e}), m(\mathfrak{e}))$, is positive. This is because the variance of estimated tax rates underestimates the gains from reallocation as the firms that are discouraged from drilling have higher expected emissions rates than expected tax rates ($\mathfrak{e} > m(\mathfrak{e})$ for high $\mathfrak{e}$). Abatement gains depend on the post-drilling emissions rate e and follow the formula given in Section 2.5.

Data To estimate the potential impact of this mechanism, we focus our analysis on the Permian Basin of west Texas and southeast New Mexico. The Permian's oil-rich shales account for nearly one-third of U.S. oil and 10% of U.S. natural gas production (Energy Information Administration (2019)). We draw upon the complete monthly production data from New Mexico and Texas for the 2012-2024 vintages of wells. These data track oil and gas production at the lease level through 2024, and we use them to estimate the lifetime (i.e. 8 year) production from each site based on a standard exponential decline curve analysis (Jacobs (2020), see Appendix B.4 for details).

We draw upon two sources to estimate site-level emissions rates (i.e. for the 8-year life of the well). First, some venting and flaring activities are reported directly to the state regulators alongside production volumes. As this is only one component of total emissions, we draw upon the estimates of Omara et al. (2024), which provide emissions rates as a share of gas production (i.e. loss rates) for a 0.1° x 0.1° grid of the Permian. A site's total emissions is the sum of site-level venting and flaring reports and cell-level loss rates applied to site-level gas production. A site's lifetime emissions rate is then this total mass divided by its lifetime production in BOE. See Appendix B.4 for complete details.

The final task before applying our mechanism to this setting is to separate anticipated emissions rates from those realized ex post. The key predictors we use are location (i.e. from the grid partition of the Permian), vintage, and output. We regress lease-level emissions rates from the 2012-2019 vintages on these predictors using a Poisson regression to account for wells with zero emissions rates. The predicted values of this regression form the driller's

of F has little effect on firms' drilling decision.

expected emissions rate, $\mathfrak{e}_l$ when drilling in grid cell l. The residuals from this regression form the mean one distribution of emissions shocks, $\Psi_\varkappa$, that drillers use to take expectations when deciding whether or not to drill. For outcomes that depend on the overall distribution of emissions rates, $\Psi(e)$, we construct a complete sample space of e by taking the Kronecker product of each well's ex ante emissions rate with the distribution of emissions rates shocks, $\Psi_\varkappa$, described above. This process is detailed in Appendix B.5.

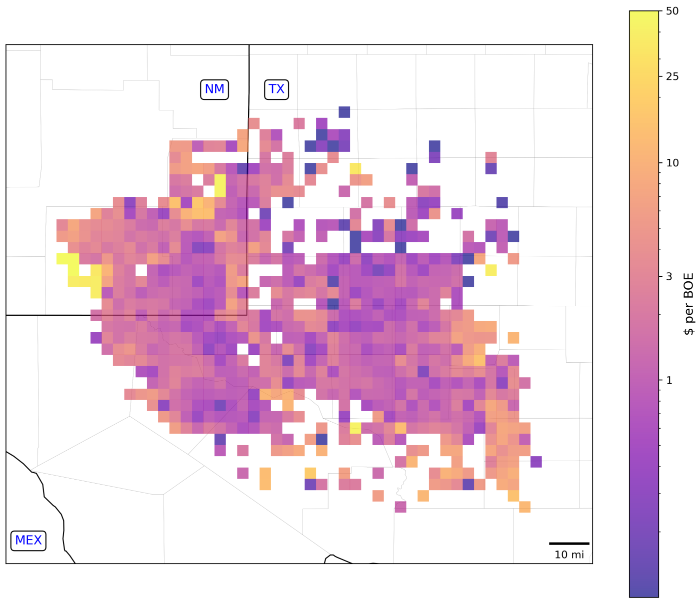
This approach yields a predicted distribution of emissions whose variance is 67% of that of the realized emissions rate. There is also a spatial pattern for the expected emissions. We plot the mean expected emissions as the external cost per barrel of oil-equivalent, $\$/BOE$, on the 0.1° x 0.1° grid of the Permian Basin in Figure 1. The map is presented on a log scale, demonstrating the fact that many areas can anticipate paying less than $1/BOE, while the social cost of methane emissions of the most polluting wells rivals their output price. In particular, the areas around the edges of the Permian's two sub-basins tend to be associated with greater emissions. Peripheral areas are less likely to have dense infrastructure networks to bring gas to market, and therefore rely on flaring. This variability in anticipated emissions rates is what drives the reallocation of well drilling in response to emissions taxation.

With the variance of anticipated and realized 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 Proposition 4. For this last number we use an elasticity of supply of 1.63 from Newell and Prest (2019).[30] Output prices per BOE are constructed by weighting 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. Throughout our analysis we assume that world output prices are unaffected, and therefore do not calculate impacts on consumer surplus.

Because lifetime production and emissions rates (both ex ante and realized) are estimated, we account for estimation error by bootstrapping the entire process 1,000 times and present the means and 95% confidence intervals across bootstrap samples. Exact formulas for each calculation are provided in Appendix B.5.

[30] This is the drilling elasticity for shale wells 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)).

Figure 1: Mean Social Cost of Methane per BOE



Notes: This figure plots the mean external social cost of methane emissions per BOE on a 0.1° x 0.1° grid of the Permian Basin. It is based on the 2012-2019 vintages of wells, site-level venting and flaring data, and gas loss rates from Omara et al. (2024).

3.2 Empirical Analysis

We use the sufficient statistics formulas from Proposition 4 to derive the welfare gains and emissions reductions from the voluntary certification mechanism on the 2019 vintage of wells. These formulas correspond to a long-run equilibrium in which firms make drilling and abatement decisions based on the expected lifetime production of sites.

To understand the boundaries of the outcome space, Table 1 presents the results of this analysis for an output tax and a universal emissions tax with no certification costs. The first column summarizes expected eight-year production and emissions for wells drilled in 2019.[31] We estimate that wells drilled in 2019 will produce 3.5B barrels of oil-equivalent (BOE) over the subsequent eight years, and release methane emissions causing $5.7B in external costs.

The second column of Table 1 presents estimates for approximate changes under an

[31] Our approximations for the impact of tax instruments are invariant to any underlying cost distribution beyond the supply elasticity, so to begin our analysis we need not take a stand on prevailing producer surplus.

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 (Uniform Cost Distribution)
Quantities				
Production (Billions BOE)	3.50	-0.31	-0.31	-0.27
		[-0.34,-0.28]	[-0.34,-0.28]	[-0.30, -0.25]
Methane Emissions (Tg)	3.80	-0.34	-2.68	-2.07
	[3.46, 4.14]	[-0.39, -0.29]	[-2.98, -2.38]	[-2.28, -1.86]
Welfare				
Producer Surplus (Billion USD)		-5.45	-3.69	-3.64
		[-5.93, -4.98]	[-4.00, -3.38]	[-3.95, -3.34]
Tax Revenue (Billion USD)		5.19	1.69	2.60
		[4.76, 5.64]	[1.51, 1.85]	[2.39, 2.81]
External Cost (Billion USD)	5.70	-0.51	-4.01	-3.10
	[5.18, 6.21]	[-0.58, -0.43]	[-4.47, -3.57]	[-3.42, -2.80]
Welfare: Total (Billion USD)		0.25	2.01	2.06
		[0.21, 0.29]	[1.79, 2.24]	[1.85, 2.27]
Welfare: Abatement (Billion USD)			1.46	
			[1.33, 1.59]	

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.

output tax that reflects the average external cost of methane per BOE in the Permian, which is \$1.62. This amounts to a 2.9% tax on oil and an 11% tax on natural gas at 2019 prices. The impact on production quantities is small, reflecting the modest size of the tax and near unit elasticity of supply. The poor emissions targeting of an output tax in this setting and lack of abatement incentives mean that it is not particularly effective at reducing pollution. Instead, the main effect of an output tax is to raise revenue 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 pollution by about 70%. 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 \$2B for the 2019 vintage of wells. Reflecting the efficacy of abatement opportunities in this setting, we estimate that nearly three quarters of the net welfare benefits come from emissions reductions in response to the Pigouvian tax.

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. We assume a uniform distribution for drilling cost shocks calibrated to observed prices, quantities, and an elasticity of supply of 1.63. 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). All details of the model and formulas for each outcome are presented in Appendix B.6.

We find that the estimates when assuming a uniform underlying cost distribution in the driller's decision follow those derived from the distribution-free approximation formulas reasonably well. There is a slightly smaller output response and a modestly smaller emissions response in the uniform cost case. These offset each other so that the aggregate change in producer surplus is very close to that of the approximation formula. Higher emissions under the uniform model yield a smaller reduction in external costs and greater tax revenues. Overall, both the approximation and uniform cost distribution yield similar net social benefits of Pigouvian taxation.

Because complete output and emissions taxes form the endpoints of an unraveling mechanism, the welfare estimates in Table 1 are the starting points for our analysis 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. 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.

The shaded grey area at the bottom of the figure represents the welfare gains from the output tax. We note this amount across all levels of certification to help decompose the development of reallocation and abatement gains. The green shaded area represents the welfare gains from abatement. Consistent with a large number of wells having extremely low emissions rates, there are minimal abatement benefits at even 50% certification. Instead, the majority of gains earlier in the unraveling process come from reallocating output. This is represented with the gap between the upper blue line and the abatement gains. The gains from abatement accelerate as more heavily-polluting wells are induced to certify.

The figure also shows that ample abatement opportunities yield rising marginal benefits of certification. In this particular setting, there is no level of certification cost for which the social planner would like to implement an interior solution. 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.

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. With about 4,000 sites in the 2019 vintage, an upper bound on aggregate certification costs would be \$0.4B, significantly lower than the estimated benefits of certification. We therefore estimate that the social planner would like to achieve full unraveling.

As in the baseline model, the certification mechanism delivers full unraveling when there is no certification tax and the aggregate certification costs are commensurate with the welfare gains (i.e. second order in v). Therefore, the certification policy delivers the first best for realistic certification costs values.[32]

We further illustrate how unraveling would develop when implemented by an information-constrained regulator using the algorithm in Section 2.4. There is a subtle difference in the welfare achieved under the algorithm that merits discussion. Implemented iteratively, 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. This difference is noticeable in the first round, and quickly disappears as the increments in conditional expectations become smaller.

In the first round, we estimate that about two thirds of sites (denoted $\Psi(\hat{e}_1)$) would certify, reflecting the skewness of the emissions distribution. In the first round there is

[32] Technically, we obtain that there is (near-) full unraveling in equilibrium when F is second order in v. This amounts to F being small relative to the social costs of emissions. This is true for the average well as the social cost of emissions is \$1.4M. However, this may not hold for particularly small wells. We do check numerically that the equilibrium must feature certification by at least all wells that are in the top 99.6% of the quantity-weighted output distribution with the upper bound on the value of F. Similarly, our previous claim that the social planner prefers either full or no unraveling is only approximately true, as the social planner might prefer not to certify particularly small wells for which F/q becomes large.

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. Thus a program with voluntary certification that does not update the default rate is particularly unattractive: it incurs the social cost of certifying the majority of wells with little to show for it.

The first update of the default output tax to reflect the mean emissions of uncertified firms would yield about $500M in gains per vintage, 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 by the lower tax burden under abatement induces the final sites to certify. A voluntary certification program for methane emissions therefore delivers the first-best outcome so long as the benefits of emissions taxation exceed certification costs. This holds even if it is implemented iteratively by an information-constrained regulator.

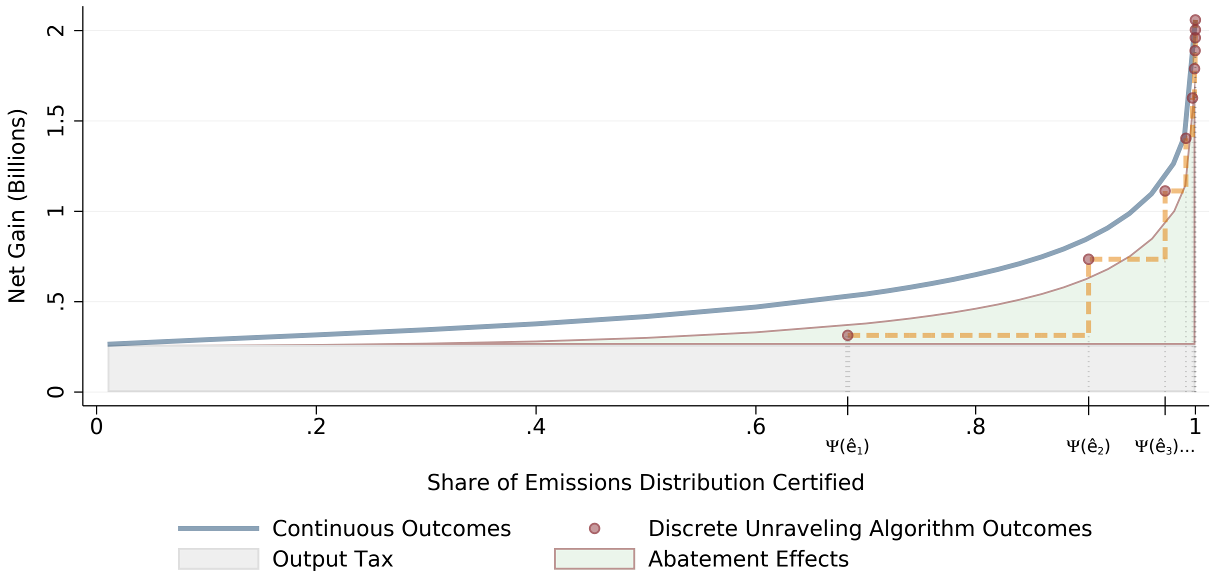
4 Unraveling in the International Case

Our voluntary certification mechanism is particularly relevant in the international setting. A country without jurisdiction abroad, but aiming to tax emissions embedded in its imports would need the cooperation of the foreign exporter to tax the actual carbon content of the goods. As a result, the literature has often focused on output-based carbon tariffs where imports are taxed according to the average carbon content of imports from a given industry-country pair. Given an output tariff, clean foreign firms would have an incentive to credibly demonstrate that they are clean, creating demand for certification. 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 is aligned with our proposed certification-based approach: Exporters are required to report their emissions rate or face a default tariff that can be updated using information revealed by certifying firms.

Setting the default value as the average emission rate of uncertified firms as in our mechanism is feasible when average emission rates and emission rates of certified firms are observable. In addition, such a default value is more likely to be WTO-compatible than an arbitrarily high default—a crucial consideration given ongoing debate about international trade rules and carbon tariffs.

In this section, we extend our model of voluntary certification to an international setting.

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, ignoring certification costs. 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. 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.

We consider a Home policy maker who values welfare both in Home and Foreign but can implement policies only in Home. 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.[33] There is no Foreign policy maker.

As in the domestic setting, certification leads to a reallocation of production among Foreign producers and incentivizes certified Foreign firms to abate. However, two additional complications emerge due to untaxed consumption abroad, making the international case more nuanced. First, when prices decline in Foreign, consumption there increases. Since Foreign consumption is untaxed and consequently inefficiently high, this additional consumption lowers overall welfare. Second, 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. In this section, we formally derive these two new effects and compare

[33]The social cost of carbon used by the U.S. government is calculated to reflect global damages.

them with the now familiar reallocation and abatement effects.

We build on the the model in Section 2. Demand for the polluting good is characterized by the demand functions $D_H(p)$ in Home and $D_F(p)$ in Foreign. The polluting good is supplied competitively by a continuum of mass 1 of firms in Foreign – for simplicity, we abstract from Home production in the body of the paper, but include it in Appendix C. Foreign firms are identical except for their emissions rates: we denote by s_F their common supply function and by $\Psi_F(e)$ the c.d.f. of Foreign emissions rates. Firms can abate with the same technology as in the domestic model. Shipping the polluting good between countries costs κ. Consumers in both countries experience the same disutility, v, from global emissions, G. See Appendix C for a microfoundation of this set-up.

We focus on 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 discuss alternatives at the end of this section and the optimal output tax in Appendix C.3.

4.1 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 sell only in the Foreign market, their domestic 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:

$$\rho \equiv p_F - (p_H - \tau_F E_F[e|e > \hat{e}] - \kappa). \tag{18}$$

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

Certified Foreign exporters pay the emission tax τ_F, abate at rate $a_F = b'^{-1}(\tau_F)$, leading to a net gain in price $A_F \equiv \tau_F a_F(\tau_F) - b_F(a_F(\tau_F))$. Accounting for shipping costs, their production is given by $s_F(p_H - \tau_F e + A_F - \kappa)$. The world market clearing condition for the polluting good is then given by:

$$D_H(p_H) + D_F(p_F) = \Psi_F(\hat{e}) E_F \left[s_F(p_H - \tau_F e - \kappa + A_F) | e < \hat{e} \right] + (1 - \Psi_F(\hat{e})) s_F(p_F). \quad (19)$$

The threshold $\hat{e}$ is determined by an indifference condition similar to equation (2):

$$\pi_F(p_H - \tau_F \hat{e} + A_F - \kappa) - (F + f) = \pi_F(p_F), \quad (20)$$

which again reflects that some uncertified firms serve the Foreign market. We can take $\hat{e}$ as a policy parameter determined by the certification tax f, or set exogenously by the government only permitting firms below some predetermined level to certify.

For a given $\hat{e}$, the equilibrium is characterized by the two endogenous variables (p_H, p_F), and equation (19) provides one equilibrium equation through global market clearing. In the pooled equilibrium, the second equilibrium equation is $\rho = 0$. In that case, the difference in consumer prices, $p_H - p_F$, is increasing in $\hat{e}$ (see 18): 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. This case mirrors the domestic setting, where the price gap between a certified firm with a given emission rate and uncertified firms increases with $\hat{e}$. 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).

In the separating equilibrium, the second equilibrium equation is instead provided by market clearing in Foreign:

$$D_F(p_F) = s_F(p_F)(1 - \Psi_F(\hat{e})). \quad (21)$$

This equation pins down p_F as an increasing function of the certification cutoff $\hat{e}$, since fewer firms supply the Foreign market with higher certification. The Home price p_H is then given by equation (19) and decreases in $\hat{e}$, so that the price difference $p_H - p_F$ decreases in $\hat{e}$. In contrast with the domestic setting, an increase in certification narrows the gap between the price faced by a given certified firm and uncertified firms.

Whether greater certification raises or lowers the price difference $p_H - p_F$ therefore depends on which equilibrium is active. At low certification levels, certified supply is insufficient to meet Home demand, so that the economy must be in the pooling equilibrium. As certification expands, certified supply becomes sufficient to match $D_H(p_H)$, which activates the separating equilibrium: uncertified exports cease and the price gap $p_H - p_F$ becomes independent of the output tariff (which no one pays).

4.2 Welfare

We now analyze the effects of certification on world welfare W, which can be written as:

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

CS_H and CS_F refer to consumer surplus at Home and Foreign and PS_F to producer surplus in Foreign. These expressions are standard and the details are presented in Section C in the appendix. The novel term is $-vG_{F,F}$ where $G_{F,F} \equiv E_F [e|e > \hat{e}] D_F(p_F)$ captures the Foreign emissions from production for Foreign consumption: only uncertified firms supply the Foreign domestic market so that the average emission rate for Foreign domestic consumption is $E_F [e|e > \hat{e}]$. These emissions are never taxed. In contrast, emissions from Foreign exports to Home, $G_F - G_{F,F}$ are taxed under Home jurisdiction and can be taxed so as to internalize the externality ($\tau_F = v$).

The following Proposition parallels Corollary 1 and gives the difference in welfare and emissions between an equilibrium with voluntary certification and one without—we do not impose $\tau_F = v$ here. As before, we assume that $b'(0) = 0$ to avoid a dichotomy of cases. We again take approximations in τ_F and assume that v and κ are of the same order as τ_F (which is satisfied in the steel case of Appendix D):

Proposition 5. *To a second-order approximation in (τ_F, v, κ), 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:*

$$W^V - W^U = \underbrace{s'_F \tau_F \left(v - \frac{\tau_F}{2}\right) Var_F(\varepsilon)}_{\text{Reallocation Effect}} + \underbrace{\tau_F \left(v - \frac{\tau_F}{2}\right) \frac{s_F \Psi_F(\hat{e})}{b''(0)}}_{\text{Abatement Effect}} \tag{22}$$

$$\underbrace{-(v - \tau_F)E_F [e]\, s'_F \Delta p_H}_{\text{Price Effect on Untaxed Emissions}} \underbrace{- \frac{D'_F}{2} \left(\tau_F \left(E_F [e] + E_F [e|e > \hat{e}]\right) - \rho\right) \Delta p_F}_{\text{Consumption Leakage Effect}}$$

$$\underbrace{-\rho s'_F \left(1 - \Psi^F (\hat{e})\right) \left(\frac{\Delta p_H + \rho}{2} + (v - \tau_F)E_F [e|e > \hat{e}]\right)}_{\text{Backfilling Effect}} \underbrace{-F\Psi_F(\hat{e})}_{\text{Cost of Certification}} + o(\tau^2),$$

where $\rho \geq 0$ – defined by equation (18) – 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 price changes due to certification in each country.[34] *It holds that*

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

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

- The Consumption Leakage Effect is always negative in the pooling equilibrium,

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

- The total net welfare change is ambiguous.

To a first-order approximation, the emissions change from moving to certification obeys:

$$G_F^V - G_F^U = s_F^{'} \left(E_F\left[e\right]\Delta p_H - \tau_F Var_F\left(\varepsilon\right) + \rho E_F\left[e|e>\hat{e}\right]\left(1-\Psi_F\left(\hat{e}\right)\right)\right) - \Psi_F\left(\hat{e}\right) a_F s_F + o(\tau). \tag{27}$$

The Proposition mirrors 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. This term arises because prices are no longer constant and an increase in home prices p_H encourages global production. This is harmful to welfare only if the tax rate on Foreign certified firm is less than Pigouvian ($\tau_F < v$). More interestingly, we have the two terms with which we opened this section:

[34] We can use first and second order approximations to get expressions for the price changes. We denote ϵ_F^S the Foreign supply elasticity and use similar notations for local demand elasticities. We denote ϵ^D the world demand elasticity which is the quantity-weighted average elasticity: $\epsilon^D = \epsilon_F^D\theta_F^D + \epsilon_H^D\theta_H^D$ with $\theta_H^D = D_H/(D_F + D_H)$. In Appendix C.2.2, we show that the voluntary certification program will have the following effects on the price at Home and Foreign:

$$\Delta p_H = \underbrace{\frac{-\epsilon_F^D\theta_F^D}{\epsilon_F^S - \epsilon^D}\tau_F\left(E_F\left[e|e>\hat{e}\right] - E_F\left[e\right]\right)}_{>0} \underbrace{-\rho\frac{\left(1-\Psi_F(\hat{e})\right)\epsilon_F^S - \epsilon_F^D\theta_F^D}{\epsilon_F^S - \epsilon^D}}_{\leq 0} + o\left(\tau\right), \tag{23}$$

$$\Delta p_F = \underbrace{\frac{\epsilon_H^D\theta_H^D - \epsilon_F^S}{\epsilon_F^S - \epsilon^D}\tau_F\left(E_F\left[e|e>\hat{e}\right] - E_F\left[e\right]\right)}_{<0} \underbrace{+\rho\frac{\epsilon_F^S\Psi_F - \epsilon_H^D\theta_D^H}{\epsilon_F^S - \epsilon^D}}_{\geq 0} + o(\tau), \tag{24}$$

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

$$\Delta p_H - \Delta p_F = \tau_F\left[E_F\left[e|e>\hat{e}\right] - E_F\left[e\right]\right] - \rho. \tag{25}$$

In the pooling equilibrium $\rho = 0$ and in the separating equilibrium $\rho \geq 0$. In the latter case, ρ is given by combining equations (18) and (20)

$$\rho = \tau_F\left(E_F\left[e|e>\hat{e}\right] - \hat{e}\right) - \frac{F+f}{s^F(p_0-\kappa)} + o(\tau). \tag{26}$$

First, the *Consumption Leakage Effect*: When Home imposes an output-based tariff without certification (equal to $\tau_F E_F[e]$), the Foreign price decreases which encourages untaxed Foreign consumption. 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 $(\tau_F(E_F[e] + E_F[e|e > \hat{e}]) - \rho)$, the average of the price gap under no certification and certification, respectively. This effect disappears when Foreign demand is inelastic.

Second, the *Backfilling Effect*:[35] Whereas the Consumption Leakage Effect focuses on consumer prices that are 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 than the average firm under certification (their price is reduced by $\tau_F(E_F[e|e > \hat{e}] - E_F[e])$) and they correspondingly reduce their relative 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, and receive a price that is too high by ρ. As the relatively clean firms are exporters, those left to serve demand in Foreign are particularly polluting. With certification, 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.[36]

Changes in emissions 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

[35] 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)).

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

reallocation effect and abatement, but increase because of a backfilling effect if $\rho > 0$.

Consequently, Proposition 5 shows how Foreign demand alters the conclusion from the domestic model of Section 2. Though this policy continues to reallocate production from the more polluting to the less polluting firms and to incentivize abatement, the presence of Foreign consumption poses limits to the taxing ability of the Home policy maker. As a result, the welfare effects of the certification mechanism are generally ambiguous. However, it is possible to derive sufficient conditions under which they are positive: for instance, in the pooling case ($\rho = 0$) with Pigouvian taxation ($\tau_F = v$), certification raises welfare when abatement effects are large or when the Foreign supply elasticity is large relative to the demand elasticity (or s'_F large relative to D'_F). Given that the EU is implementing a similar two-tiered mechanism, our theoretical result acts as a word of caution and suggests that the policy should not be implemented blindly.

Our certification approach is particularly applicable to carbon-intensive commodity sectors, such as iron and steel, cement, and aluminum (essentially the sectors targeted by the EU-ETS). How feasible would it be for a regulator to obtain the sufficient statistics required to evaluate the prospective welfare impacts of certification? Estimating demand and supply elasticities is a standard exercise, engineering studies of marginal abatement curves are generally available and a regulator would know the certification costs. While estimating average emissions intensity by sector-country pairs can feasibly be achieved through input-output databases, determining the full distribution of firm-level emission rates could be more challenging. Policymakers could address this by implementing certification progressively, gradually increasing $\hat{e}$. This would allow them both to learn the emission distribution and to avoid the case where certification may lead to welfare losses instead of welfare gains.

In Appendix D, we illustrate how to proceed in the case of Brazilian steel exports to the OECD, calibrating the model with publicly available data. Though our numbers should be taken as a proof-of-concept instead of a precise answer, we find that the welfare gains can be substantial when the mechanism is implemented judiciously: In our baseline calibration, the certification mechanism recoups 71.5% of the gains of the First best with Pigouvian taxation in Brazil relative to laissez-faire, while an output-based tariff recoups 58.9% of the gains. In addition, we show that most of the gains already materialize for low levels of certification.

4.3 Alternative policy environments

The previous analysis naturally raises the question of the second-best policy, that is the optimal program when Home has access to an emissions tax on certified firms, an output-based tariff, and a tax/subsidy on certification, while Foreign does not impose taxes. We find that the social planner sets the emission tax for certified Foreign firms at the Pigouvian level: $\tau_F = v$ (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:

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

The social planner only taxes the uncertified firms at the Pigouvian 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 tax/subsidy $f = (t^* - v\hat{e})s_F(p_F)$ which mirrors the expression from the domestic setting. Additional certification creates convergence (divergence) between Home and Foreign prices for the separating (pooling) equilibrium. A certification tax allows the government to balance the benefits of more targeted taxation with the distortions in the Foreign market.[37]

Interestingly, with Home production of the polluting good, the policy maker can potentially set the certification tax/subsidy and an export subsidy such that Home exports the polluting good to Foreign, thereby creating bilateral trade in a homogenous product.[38] 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 (covered under the voluntary program), which reduces the Backfilling effect. This is not optimal when the transport cost κ is of the same order as the social cost v (as in the steel example), but it may be so in contexts with very low trade costs such as electricity sold on a grid that spans jurisdictional borders.

[37] 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 emission taxes on Home production ($\tau_H = v$).

[38] 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.

5 Conclusion

Adverse selection is widely considered an undesirable market attribute that requires government intervention to prevent (Einav and Finkelstein (2011)). In this paper we show how it can be used to policymakers' advantage to overcome obstacles to internalizing externalities. Facing enforcement costs, or lacking jurisdiction or political will, the government presents heterogeneous firms with the option to certify their emissions for Pigouvian taxation, or pay an output tax that reflects the average emissions of uncertified firms. The reduced burden of an emissions tax induces relatively low-emissions firms to certify. 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. This mechanism generates welfare gains by reallocating production to low emissions firms and incentivizing abatement. If unraveling is complete, such a voluntary program achieves the same outcome as the otherwise-infeasible mandate to tax emissions directly.

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. This would yield welfare gains of $2B per vintage from the U.S. Permian basin alone.

In the international setting, the voluntary emissions tax mechanism extends the incentive to reduce emissions beyond the borders of the country adopting a carbon tax. This delivers gains from reallocation and abatement beyond the policymaker's jurisdiction. However, the possibility for Foreign firms to avoid regulation by serving the Foreign market erodes, and potentially reverses the welfare benefits of the certification program. We derive conditions that determine whether such a program would further increase welfare and reduce emissions beyond those achievable with an output-based border carbon adjustment.

Since this paper's first circulation there have been important examples demonstrating real-world demand for the kind of voluntary certification program developed here (Whitehouse (2021); Council of the European Union (2022)). Though our focus has been on settings with a pollution externality, the general structure of the policy design has broad applicability to help the government separate heterogeneous types. These include implementing vehicle miles travelled taxes as gas tax revenues dwindle, and concentrating enforcement efforts on high-risk offenders. The common thread throughout is that uniform policies misallocate regulatory burden in the presence of heterogeneity, and adverse selection can serve as a lever to undo such misallocation.

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.

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. 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.

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 2016.

—, "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.

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.

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.

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.

Dewatripont, Mathias and Gerard Roland, "Economic Reform and Dynamic Political Constraints," *Review of Economic Studies*, 1992, *59* (4), 703–730. ISBN: 00346527.

— **and** — , "The Design of Reform Packages under Uncertainty," *The American Economic Review*, 1995, *85* (5), 1207–1223.

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 and Amy Finkelstein, "Selection in Insurance Markets: Theory and Empirics in Pictures," *Journal of Economic Perspectives*, 2011, *25* (1), 115–138.

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

Elliott, Joshua, Ian Foster, Samuel Kortum, 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.

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

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.

Harstad, Bård, "Trade and Trees," 2022.

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.

Hsiao, Allan, "Coordination and Commitment in International Climate Action: Evidence from Palm Oil," Technical Report, Massachusetts Institute of Technology 2021. Publication Title: Mimeo.

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 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 J, "Sufficient statistics revisited," *Annual Review of Economics*, 2021, *13* (1), 515–538.

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.

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.

Millock, Katrin, David Sunding, and David Zilberman, "Regulating Pollution with Endogenous Monitoring," *Journal of Environmental Economics and Management*, 2002, *44* (2), 221–241.

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

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.

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

Omara, Mark, Anthony Himmelberger, Katlyn MacKay, James P. Williams, Joshua Benmergui, Maryann Sargent, Steven C. Wofsy, and Ritesh Gautam, "Constructing a measurement-based spatially explicit inventory of US oil and gas methane emissions," April 2024.

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.

Plant, Genevieve, Eric A. Kort, Adam R. Brandt, Yuanlei Chen, Graham Fordice, Alan M. Gorchov Negron, Stefan Schwietzke, Mackenzie Smith, and Daniel Zavala-Araiza, "Inefficient and unlit natural gas flares both emit large quantities of methane," *Science*, September 2022, *377* (6614), 1566–1571.

Pram, Kym, "Disclosure, welfare and adverse selection," *Journal of Economic Theory*, October 2021, *197*, 105327.

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.

Schelling, Thomas C., *Arms and Influence*, New Haven, CT: Yale University Press, 1965.

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.

Townsend, Robert M., "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 1979, *21*, 265–293.

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.

—, "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, Ruixiong 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.

A Theory Appendix: Domestic case

This Appendix contains details and proofs for the results of the domestic model of Section 2.

A.1 Setup and proofs of uniqueness, Proposition 1, and Remark 1.

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 $\hat{e} = \underline{e}$.

The maximization problem of certified firms (1) can be decomposed into two parts, where firms first choose their level of abatement and then their level of production. The level of abatement is chosen by maximizing $\tau a - b(a)$. With b convex, we get that there is a single, potential interior, solution given by $b'^{-1}(\tau)$. With $b'(\underline{e}) > \tau$, the interior solution features $a < \underline{e}$ so that no firms ever want to abate all emissions. If $b'(0) \geq \tau$, then there is no abatement, while with $b'(0) < \tau$, all certified firms abate the same amount $b'^{-1}(\tau)$. We then get that $A(\tau) = 0$ if $b'(0) \geq \tau$ and $A(\tau) = \tau b'^{-1}(\tau) - b(b'^{-1}(\tau))$ otherwise.

The equilibrium is determined by two equations: i) the indifference condition on certification (2) (together with the conditions for corner values for $\hat{e}$) with t defined by $t = \tau E[e|e > \hat{e}]$ and ii) the market clearing equation. Using (4), the market clearing equation can be written as:

$$D(p) = E\left[s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}} A(\tau)\right)\right]. \tag{1}$$

For a given $\hat{e}$, equation (1) uniquely defines the equilibrium price p: the function $D(p)$ is strictly decreasing in p from a positive value to 0 because of the Inada condition imposed on u, while s is non-decreasing in p with $s(0) = 0$, and s positive for p large enough. Since D and s are continuously differentiable, then, using the implicit function theorem, equation (1) defines p as a differentiable function of $\widehat{e}$ on $(\underline{e}, \overline{e})$.[1]

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:

$$W = I - pC + u(D(p)) - vG + E\left[\pi\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}} A(\tau)\right)\right] + \tau G - F\Psi(\hat{e}), \tag{2}$$

where the representative agent has exogenous labor income I and emissions are given by equation (5).

[1]Technically, $E\left[s\left(p - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}} A(\tau)\right)\right]$ may not be differentiable for $\hat{e}$ such that uncertified firms just exit: $p = \tau E\left[e|e > \hat{e}\right]$, however, the rest of the proofs work even with this single point of non-differentiability.

A.1.1 Proof of Lemma 1

In this subsection, we show that the equilibrium is unique. To do so, we let the certification threshold vary freely and define

$$g\left(\widehat{e}\right) \equiv \pi\left(p\left(\widehat{e}\right) - \tau\widehat{e} + A\left(\tau\right)\right) - \pi\left(p\left(\widehat{e}\right) - \tau E\left[e|e > \hat{e}\right]\right), \tag{3}$$

where $p\left(\widehat{e}\right)$ is implicitly defined by the market clearing condition (1) and where we set $\pi\left(p\right) = 0$ if $p < 0$. An equilibrium is characterized by $g\left(\widehat{e}\right) = F + f$ if $\widehat{e}$ is interior, $g\left(\overline{e}\right) \geq F + f$or $g\left(\underline{e}\right) \leq F + f$. To start we establish:

Lemma 2. *Assume that $g\left(\widehat{e}\right) \geq F + f$, then the function g is strictly decreasing in $\widehat{e}$ unless the firm with emission rate $\widehat{e}$ exits in which case g is constant and equal to zero.*

Proof. Note that if $g\left(\widehat{e}\right) \geq F + f$, then $\pi\left(p - \tau\widehat{e} + A\left(\tau\right)\right) - \left(F + f\right) \geq \pi\left(p - \tau E\left[e|e > \hat{e}\right]\right)$, so that if uncertified firms (i.e. firms with $e \geq \widehat{e}$) produce (i.e. $p > \tau E\left[e|e > \hat{e}\right]$), then so do certified firms (firms with $e \leq \widehat{e}$).

Case with no exit. Consider first the case where $p > \tau E\left[e|e > \hat{e}\right]$, so that there is no exit, and set $\hat{e} < \overline{e}$. Then, p is a differentiable function of $\widehat{e}$ with

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

where

$$\frac{dE\left[e|e > \hat{e}\right]}{d\widehat{e}} = \left(E\left[e|e > \hat{e}\right] - \widehat{e}\right)\frac{\psi\left(\widehat{e}\right)}{1 - \Psi\left(\widehat{e}\right)}. \tag{4}$$

Therefore, we get

$$\frac{dp}{d\widehat{e}} = -\psi\left(\widehat{e}\right)\frac{s\left(p - \tau\widehat{e} + A\left(\tau\right)\right) - s\left(p - \tau E\left[e|e > \hat{e}\right]\right) - \tau E\left[e - \widehat{e}|e > \widehat{e}\right]s'\left(p - \tau E\left[e|e > \hat{e}\right]\right)}{\int_0^{\widehat{e}} s'\left(p - \tau e + A\left(\tau\right)\right)\psi\left(e\right)de + \left(1 - \Psi\left(\widehat{e}\right)\right)s'\left(p - \tau E\left[e|e > \hat{e}\right]\right) - D'\left(p\right)}. \tag{5}$$

Moreover, g is also differentiable with

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

In the case where $\frac{dp}{d\widehat{e}}\leq 0$, we then get $g'(\widehat{e})<0$.

Alternatively consider the case where $\frac{dp}{d\widehat{e}}>0$, then we can rewrite:

$$\begin{aligned}
g'(\widehat{e}) &= \left(\frac{dp}{d\widehat{e}}-\tau\right)\left(s\left(p-\tau\widehat{e}+A(\tau)\right)-s\left(p-\tau E\left[e|e>\hat{e}\right]\right)\right)+\tau\left(\frac{dE\left[e|e>\hat{e}\right]}{d\widehat{e}}-1\right)s\left(p-\tau E\left[e|e>\hat{e}\right]\right) \\
&= \left(\psi(\widehat{e})\frac{\tau E\left(e-\widehat{e}|e>\widehat{e}\right)s'\left(p-\tau E\left[e|e>\hat{e}\right]\right)-\left(s\left(p-\tau\widehat{e}+A(\tau)\right)-s\left(p-\tau E\left[e|e>\hat{e}\right]\right)\right)}{\int_0^{\widehat{e}}s'\left(p-\tau e+A(\tau)\right)\psi(e)\,de+\left(1-\Psi(\widehat{e})\right)s'\left(p-\tau E\left[e|e>\hat{e}\right]\right)-D'(p)}-\tau\right) \\
&\quad \times\left(s\left(p-\tau\widehat{e}+A(\tau)\right)-s\left(p-\tau E\left[e|e>\hat{e}\right]\right)\right) \\
&\quad +\tau\left(\frac{dE\left[e|e>\hat{e}\right]}{d\widehat{e}}-1\right)s\left(p-\tau E\left[e|e>\hat{e}\right]\right)
\end{aligned} \tag{6}$$

where the second line uses (5). We further note that if $\frac{dp}{d\widehat{e}}>0$, then (5) implies that $\tau E\left(e-\widehat{e}|e>\widehat{e}\right)s'\left(p-\tau E\left[e|e>\hat{e}\right]\right)-\left(s\left(p-\tau\widehat{e}+A(\tau)\right)-s\left(p-\tau E\left[e|e>\hat{e}\right]\right)\right)>0$. Moreover, with $E\left[e|e>\hat{e}\right]-\hat{e}$ decreasing in $\hat{e}$, we must have (using (4)):

$$\frac{dE\left[e|e>\hat{e}\right]}{d\widehat{e}}<1\iff\psi(\widehat{e})<\frac{1-\Psi(\widehat{e})}{E\left[e|e>\hat{e}\right]-\widehat{e}}. \tag{7}$$

As $\psi(\widehat{e})$ is multiplied by positive terms in (6), we can bound $g'(\widehat{e})$ in that equation using

(7) and obtain:

$$\begin{aligned}
& g'(\widehat{e}) \\
< \; & \left(\frac{(1-\Psi(\widehat{e}))\left[\tau E(e-\widehat{e}|e>\widehat{e})\, s'(p-\tau E[e|e>\hat{e}]) - (s(p-\tau\widehat{e}+A(\tau)) - s(p-\tau E[e|e>\hat{e}]))\right]}{(E[e|e>\hat{e}]-\widehat{e})\left[\int_0^{\widehat{e}} s'(p-\tau e+A(\tau))\,\psi(e)\,de + (1-\Psi(\widehat{e}))\, s'(p-\tau E[e|e>\hat{e}]) - D'(p)\right]} - \tau \right) \\
& \times (s(p-\tau\widehat{e}+A(\tau)) - s(p-\tau E[e|e>\hat{e}])) \\
& + \tau \left(\frac{dE[e|e>\hat{e}]}{d\widehat{e}} - 1 \right) s(p-\tau E[e|e>\hat{e}]) \\
< \; & -\frac{1-\Psi(\widehat{e})}{E[e|e>\hat{e}]-\widehat{e}} \frac{(s(p-\tau\widehat{e}+A(\tau)) - s(p-\tau E(e|e>\widehat{e})))^2}{\int_0^{\widehat{e}} s'(p-\tau e+A(\tau))\,\psi(e)\,de + (1-\Psi(\widehat{e}))\, s'(p-\tau E[e|e>\hat{e}]) - D'(p)} \\
& -\tau \frac{\left(\int_0^{\widehat{e}} s'(p-\tau e+A(\tau))\,\psi(e)\,de - D'(p)\right)(s(p-\tau\widehat{e}+A(\tau)) - s(p-\tau E[e|e>\hat{e}]))}{\int_0^{\widehat{e}} s'(p-\tau e+A(\tau))\,\psi(e)\,de + (1-\Psi(\widehat{e}))\, s'(p-\tau E[e|e>\hat{e}]) - D'(p)} \\
& + \tau \left(\frac{dE[e|e>\hat{e}]}{d\widehat{e}} - 1 \right) s(p-\tau E[e|e>\hat{e}]).
\end{aligned}$$

In the last line all the terms are negative, which ensures that $g'(\widehat{e}) < 0$.

As a result, g is a strictly decreasing function of $\widehat{e}$ as long as there is no exit and $g(\widehat{e}) \geq F + f$.

Case with exit of uncertified firms only. Consider now the case where $\tau\widehat{e} - A(\tau) + \pi^{-1}(\max(F+f,0)) < p \leq \tau E[e|e>\hat{e}]$ so that uncertified firms exit but all certified firms still produce (as $p - \tau\widehat{e} + A(\tau) > 0$ and $\pi(p-\tau\widehat{e}+A(\tau)) > F + f$). The market clearing equation (1) becomes:

$$D(p) = \int_0^{\widehat{e}} s(p-\tau e+A(\tau))\,\psi(e)\,de. \tag{8}$$

We can still differentiate that equation (for $p < \tau E[e|e>\hat{e}]$) and obtain that:

$$\frac{dp}{d\widehat{e}} = \frac{s(p-\tau\widehat{e}+A(\tau))\,\psi(\widehat{e})}{D'(p) - \int_0^{\widehat{e}} s'(p-\tau e+A(\tau))\,\psi(e)\,de} < 0,$$

since increasing $\widehat{e}$ in this context increases the set of firms that produce. Differentiating g, we then get

$$g'(\widehat{e}) = s(p-\tau\widehat{e}+A(\tau)) \left(\frac{dp}{d\widehat{e}} - \tau \right) < 0.$$

Therefore g is still decreasing in $\widehat{e}$, as $p - \tau\widehat{e}$ decreases in $\widehat{e}$.

Case with exit of certified firms. Finally, for $\widehat{e}$ large enough, we may have $\tau\widehat{e} - A(\tau) + \pi^{-1}(\max(F+f,0)) \geq p$: the firm with emission rate $\widehat{e}$ does not produce. In that case, g is a constant equal to 0. The equilibrium is still uniquely defined: there is an exit threshold $\overleftrightarrow{e}$ such that

$$D(p) = \int_0^{\overleftrightarrow{e}} s\left(p - \tau e + A\left(\tau\right)\right)\psi\left(e\right)de \text{ and } p = \tau\overleftrightarrow{e} - A\left(\tau\right) + \pi^{-1}\left(\max\left(F+f,0\right)\right), \quad (9)$$

all firms with $e < \overleftrightarrow{e}$ produce and certify, all firms with $e \geq \overleftrightarrow{e}$ exit.

Summing up. Considering all cases together establishes lemma 2. □

Next, we note that $g\left(\underline{e}\right) > 0$ for a non-degenerate distribution Ψ and that $g\left(\overline{e}\right) \geq 0$– whether exit has occured or not (and $g\left(\overline{e}\right)$ can be strictly positive only if there is abatement).

Assume that $g\left(\underline{e}\right) \leq F + f$, since g must be weakly decreasing whenever $g\left(e\right) \geq F + f$, then g is never above $F+f$ and we have that $g\left(e\right) < F+f$ for all e. Not even the cleanest firm ever has an incentive to certify, there is a unique equilibrium that features no certification.

Assume that $g\left(\underline{e}\right) > F + f > g\left(\overline{e}\right) \geq 0$. Given that g is strictly decreasing whenever it is weakly above $F + f > 0$, then there is a unique equilibrium characterized by $g\left(\widehat{e}\right) = F + f$ (i.e. equation (2) in the paper).

Assume now $F + f \leq g\left(\overline{e}\right)$, then all firms that produce certify (and non producing firms are indifferent between certifying or not). Again, the equilibrium is uniquely defined, with either all firms producing if $\pi\left(p - \tau\overline{e} + A\left(\tau\right)\right) \geq F + f$ or a common exit-certification threshold $\overleftrightarrow{e}$ defined through (9).

Therefore in all cases, there exists an unique equilibrium, which establishes Lemma 1.

A.1.2 Proof of Proposition 1

General case. We solve for the welfare change in the more general case where τ may not equal v. Using equation (2), we can write the welfare change following certification

$$
\begin{aligned}
& W^V - W^U && (10)\\
= \;& u(D(p^V)) - u(D(p^U)) + \left[\pi\left(p^V - \tau E\left[e\right]\right) - \pi\left(p^U - \tau E\left[e\right]\right)\right] - \left[p^V C^V - p^U C^U\right] \\
& -(v-\tau)\left[G^V - G^U\right] + E\left[\pi\left(p^V - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}} A\left(\tau\right)\right)\right] - E\left[\pi\left(p^V - \tau\varepsilon\right)\right] \\
& + E\left[\pi\left(p^V - \tau\varepsilon\right)\right] - \pi\left(p^V - \tau E\left[e\right]\right) - F\Psi(\hat{e}) \\
= \;& \underbrace{\int_{p^U}^{p^V} \left[s(p - \tau E(e)) - D(p)\right] dp}_{\text{Price Effect}} \underbrace{-(v-\tau)(G^V - G^U)}_{\text{Untaxed emissions effect}} + \underbrace{E\left[\pi\left(p^V - \tau\varepsilon\right)\right] - \pi\left(p^V - \tau E\left[e\right]\right)}_{\text{Reallocation Effect}} \\
& \underbrace{\Psi\left(\widehat{e}\right) E\left[\pi\left(p^V - \tau e + A(\tau)\right) - \pi\left(p^V - \tau e\right) | e \leq \widehat{e}\right]}_{\text{Abatement Effect}} - F\Psi(\hat{e})
\end{aligned}
$$

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.

Abatement effect. The profit function is increasing in price so the abatement effect is positive as soon as there is abatement.

Untaxed emissions effect. This effect is positive when emissions decrease and are under-taxed ($\tau < v$), it is zero when $v = \tau$ as in Proposition 1.

Linear case. Finally, we consider the linear case, assuming no exit throughout. With a linear supply curve, we get $\pi(p) = \tilde{s}p^2/2$. Therefore:

$$
\begin{aligned}
& W^V - W^U && (11)\\
= \;& \int_{p^U}^{p^V} \left[\tilde{s}(p - \tau E(e)) - D(p)\right] dp - (v-\tau)(G^V - G^U) + \frac{\tau^2}{2}\tilde{s}Var\left(\varepsilon\right) \\
& + \Psi\left(\widehat{e}\right)\tilde{s}A(\tau)\left(p^V - \tau E\left(e | e \leq \widehat{e}\right) + \frac{A(\tau)}{2}\right) - F\Psi(\hat{e})
\end{aligned}
$$

Using (4), aggregate supply in the voluntary case can be written as

$$\begin{aligned} Q^V &= \Psi\left(\widehat{e}\right)\tilde{s}\left(p^V - \tau E\left[e|e \leq \widehat{e}\right] + A(\tau)\right) + \left(1 - \Psi\left(\widehat{e}\right)\right)\left(p^V - \tau E\left[e|e \geq \widehat{e}\right]\right) \\ &= \tilde{s}\left(p^V - \tau E\left[e\right] + \Psi\left(\widehat{e}\right) A(\tau)\right). \end{aligned} \tag{12}$$

Without abatement, the aggregate supply curve is independent of $\widehat{e}$. In that case, we get $Q^V = Q^U$, leading to $p^V = p^U$ and to $W^V - W^U = \tilde{s}\frac{\tau^2}{2} Var\left(\varepsilon\right) - F\Psi(\hat{e})$, as written in the text.

Allow now for abatement but assume linear demand: $D\left(p\right) \equiv d_0 - d_1 p$. Market clearing then implies that

$$\tilde{s}\left(p^V - \tau E\left[e\right] + \Psi\left(\widehat{e}\right) A(\tau)\right) = d_0 - d_1 p^V \text{ and } \tilde{s}\left(p^U - \tau E\left[e\right]\right) = d_0 - d_1 p^U.$$

Taking the difference between the two equations gives the price change as

$$p^V - p^U = -\frac{\tilde{s}\Psi\left(\widehat{e}\right) A(\tau)}{\tilde{s} + d_1}. \tag{13}$$

We can then express the price effect as

$$\begin{aligned} \int_{p^U}^{p^V} \left[\tilde{s}(p - \tau E(e)) - D(p)\right] dp &= \int_{p^U}^{p^V} \left[\tilde{s}(p - \tau E(e)) - d_0 + d_1 p\right] dp \qquad (14) \\ &= \frac{\left[\tilde{s}(p^V - \tau E(e)) - d_0 + d_1 p^V\right]^2}{2\left(\tilde{s} + d_1\right)} \\ &= \frac{\left(\tilde{s}\Psi\left(\widehat{e}\right) A(\tau)\right)^2}{2\left(\tilde{s} + d_1\right)} \end{aligned}$$

Combining (11) with (14) delivers equation (7) when $v = \tau$.

A.1.3 Proof of Remark 1

Taking the difference between equations (5) evaluated for any $\widehat{e}$ and for $\widehat{e} = \underline{e}$ respectively gives:

$$\begin{aligned}
& G^V - G^U \\
= & E\left[\left(e - \mathbb{I}_{e\leq\hat{e}}a^*\right) s\left(p^V - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A\left(\tau\right)\right)\right] - E\left[\varepsilon s\left(p^V - \tau\varepsilon\right)\right] \\
& + E\left[\varepsilon s\left(p^V - \tau\varepsilon\right)\right] - E\left[e\right] E\left[s\left(p^V - \tau\varepsilon\right)\right] + E\left[e\right] E\left[s\left(p^V - \tau\varepsilon\right)\right] - E\left[e\right] s\left(p^U - \tau E\left[e\right]\right) \\
= & \Psi\left(\hat{e}\right) E\left[\left(e - a^*\right) s\left(p^V - \tau e + A\left(\tau\right)\right) - es\left(p^V - \tau e\right) | e < \hat{e}\right] + Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right] \\
& + E\left[e\right]\left(E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E\left[e\right]\right)\right),
\end{aligned}$$

which corresponds to equation (8).

Linear case. Consider first the linear case ($s\left(p\right) = \tilde{s}p$), and assume that there is no exit. Then, equation (8) implies

$$\begin{aligned}
& G^V - G^U && (15)\\
= & -\left(\Psi\left(\hat{e}\right)\tilde{s}E\left[a^*\left(p^V - \tau e\right) - \left(e - a^*\right) A\left(\tau\right) | e < \hat{e}\right] + \tilde{s}\tau Var\left(\varepsilon\right) + \tilde{s}E\left[e\right]\left(p^U - p^V\right)\right).
\end{aligned}$$

As argued above, without abatement, $Q^V = Q^U$, and $p^V = p^U$, so that $G^V - G^U = -\tilde{s}\tau Var\left(\varepsilon\right)$, and emissions decline with certification.

More generally, using (12), we note that through abatement, certification shifts the supply curve upward so that $Q^V \geqslant Q^U$, and $p^U \geqslant p^V$. In addition,

$$a^*\left(p^V - \tau e\right) - \left(e - a^*\right) A\left(\tau\right) = a^*\left(p^V - 2\tau e\right) + \left(e - a^*\right) b\left(a^*\right),$$

which shows that a sufficient condition for emissions to decline is that $p^V \geqslant 2\tau\hat{e}$.

Finally, with linear demand, using the price change given by (13) and plugging it in (8) delivers (9).

Conditions for the general case. Case i) Assume that s is weakly concave and that there is no abatement. Then, we get that $E\left[s\left(p - \tau\varepsilon\right)\right] \leq s\left(p - \tau E\left[e\right]\right)$: at given prices, production decreases and the rebound effect is negative. Since the price change only mitigates the rebound effect but cannot overturn it, then $E\left[s\left(p^V - \tau\varepsilon\right)\right] \leq s\left(p^U - \tau E\left[e\right]\right)$, ensuring that emissions decrease.

Case ii) Consider now the case where s is weakly convex and there may be abatement. Then, the rebound effect increases emissions but the price change is negative: $p^V \leq p^U$.

Defining $h(e) \equiv es(p^V - \tau e)$, we note that

$$(e-a)\, s\left(p - \tau e + A\left(\tau\right)\right) = h\left(e - \frac{A\left(\tau\right)}{\tau}\right) - \frac{b\left(a\right)}{\tau} s\left(p - \tau e + A\left(\tau\right)\right) \leq h\left(e\right),$$

where the second line uses that h is increasing. We then obtain

$$G^V = E\left[\left(e - \mathbb{I}_{e\leq\hat{e}}a^*\right) s\left(p^V - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A\left(\tau\right)\right)\right] \leq E\left[\varepsilon s\left(p^V - \tau\varepsilon\right)\right].$$

Using that h is concave, we further have $E\left[\varepsilon s\left(p^V - \tau\varepsilon\right)\right] \leq E\left[e\right] s\left(p^V - \tau E\left(e\right)\right)$. Then, using that $p^V \leq p^U$, we get that $G^V \leq G^U$ as stated in Remark 1.

A.2 Taylor expansions

In this Appendix, we first derive the Taylor approximations for the change in emissions and in welfare (equations (11), (10), and (13)) and prove Corollary 1. We then argue that the approximations will be good when $\tau e/p$ is small as long as the supply curve and its derivative have "reasonable" elasticities.

A.2.1 Proof of Corollary 1

We derive equations (13) and (11) using first order approximations in τ around $\tau = 0$. To avoid repetition, we look at the general case where $\tau = v$ may not hold first and then derive how the welfare change expression simplifies further for $\tau = v$ (equation (10)). We assume that v is of the same order as τ (that is τ/v is neither large nor small, and terms in v^2, τ^2, and $v\tau$ are small relative to terms in v or in τ). We denote by p_0 the equilibrium price in laissez-faire (i.e. for $\tau = 0$).

Abatement level. Taking a Taylor expansion of the first order condition for abatement $b'\left(a^*\right) = \tau$, we get $b'\left(0\right) + a^* b''(0) + o(\tau) = \tau$. Using that $b'\left(0\right) = 0$, we then obtain that the optimal abatement level is given by

$$a^* = \frac{\tau}{b''(0)} + o(\tau). \tag{16}$$

This implies that:

$$A(\tau) = \tau a^* - b(a^*) = \frac{1}{2}\frac{1}{b''(0)}\tau^2 + o(\tau^2), \tag{17}$$

so that $A(\tau)$ is second order in τ. Therefore, the amount produced by certified firms does not change at first order, i.e. $s\left(p - \tau e + A(\tau)\right) = s\left(p - \tau e\right) + o\left(\tau\right)$.

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

$$s\left(p^U - \tau E(e)\right) = D\left(p^U\right) \text{ and } s(p_0) = D(p_0). \tag{18}$$

Taking a Taylor expansion of $s\left(p^U - \tau E(e)\right)$ and $D\left(p^U\right)$ for a small τ (so that both p^U and $p^U - \tau E(e)$ are close to p_0), we get:

$$s\left(p^U - \tau E(e)\right) = s(p_0) + s'(p_0)\left(p^U - \tau E(e) - p_0\right) + o(\tau), \tag{19}$$

$$D\left(p^U\right) = D(p_0) + D'(p_0)\left(p^U - p_0\right) + o(\tau). \tag{20}$$

Combining these two expressions with (18) gives that the equilibrium price change satisfies:

$$p^U - p_0 = \frac{E(e)\, s'(p_0)}{s'(p_0) - D'(p_0)}\tau + o(\tau), \tag{21}$$

where we note that the assumptions that $c''(q) > 0$ and $u''(q) < 0$ for $q > 0$ ensure that $s'(p_0) > 0$ and $D'(p_0) < 0$. We take a similar approach with p^V using equation (1) which we write explicitly as:

$$\int_{\underline{e}}^{\widehat{e}} s\left(p^V - \tau e + A(\tau)\right)\psi(e)\, de + \left(1 - \Psi(\widehat{e})\right) s\left(p^V - \tau E(e|e > \widehat{e})\right) = D\left(p^V\right), \tag{22}$$

We note that for τ small and a bounded distribution of emission rates, τe is small as well and no firm exits. A first order Taylor expansion of $s\left(p^V - \tau\varepsilon + 1_{e<\widehat{e}}A(\tau)\right)$ gives:

$$s\left(p^V - \tau\varepsilon + 1_{e<\widehat{e}}A(\tau)\right) = s(p_0) + s'(p_0)\left(p^V - \tau\varepsilon - p_0\right) + o(\tau), \tag{23}$$

where we used that $A(\tau)$ is second order. We can then take a Taylor expansion of (22) for $p^V - \tau\varepsilon + 1_{e<\widehat{e}}A(\tau)$ around p_0 and we obtain that:

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

$$\Leftrightarrow s'(p_0)\left(p^V - \tau E(e) - p_0\right) = D'(p_0)\left(p^V - p_0\right) + o(\tau), \tag{24}$$

which, when reordered returns equation (21) (with p^Vreplacing p^U). This establishes that, to a first order, prices are the same under certification and no certification:

$$p^V = p^U + o(\tau) \tag{25}$$

Taylor expansion of the welfare change $W^V - W^U$. We take Taylor expansions of each element of equation (10) 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. We use (19), (20) and (25) to get:

$$\begin{aligned}
& \int_{p^U}^{p^V} \left[s(p - \tau E(e)) - D(p)\right] dp \\
= & \int_{p^U}^{p^V} \left(\left(s'\left(p_0\right)\left(p - \tau E\left(e\right) - p_0\right) - D'\left(p_0\right)\left(p - p_0\right)\right) + o\left(\tau\right)\right) dp + o(\tau^2) \\
= & \left(s'\left(p_0\right) - D'\left(p_0\right)\right) \frac{\left(p^V - p_0\right)^2 - \left(p^U - p_0\right)^2}{2} - \tau E\left(e\right) s'\left(p_0\right)\left(p^V - p^U\right) + o\left(\tau^2\right) \\
= & \; o\left(\tau^2\right), \qquad (26)
\end{aligned}$$

such that the price effect is zero at second order.

We proceed with the reallocation effect. Note that a second order Taylor expansion of profits allows us to write

$$\pi\left(p^V - \tau\varepsilon\right) = \pi\left(p_0\right) + s\left(p_0\right)\left(p^V - \tau\varepsilon - p_0\right) + s'\left(p_0\right)\frac{\left(p^V - \tau\varepsilon - p_0\right)^2}{2} + o\left(\tau^2\right).$$

Using this expression and the equivalent one for $\pi\left(p^V - \tau E\left(e\right)\right)$, we can then express the change in producer surplus as:

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

Using that $\int_{\underline{e}}^{\widehat{e}} e\psi(e)de + E\left(e|e>\widehat{e}\right) = E(e)$, we then obtain:

$$\begin{aligned}
& E\left(\pi\left(p^V-\tau\varepsilon\right)\right)-\pi\left(p^V-\tau E\left(e\right)\right) \\
= & \left(\int_{\underline{e}}^{\widehat{e}}\left(p^V-\tau e-p_0\right)^2\psi\left(e\right)de+\left(1-\Psi\left(\widehat{e}\right)\right)\left(p^V-\tau E\left(e|e>\widehat{e}\right)-p_0\right)^2-\left(p^V-\tau E\left(e\right)-p_0\right)^2\right) \\
& \times\frac{s'\left(p_0\right)}{2}+o\left(\tau^2\right) \\
= & \left(\int_{\underline{e}}^{\widehat{e}}\tau^2e^2\psi\left(e\right)de+\left(1-\Psi\left(\widehat{e}\right)\right)\tau^2E\left(e|e>\widehat{e}\right)^2-\tau^2E\left(e\right)^2\right)\frac{s'\left(p_0\right)}{2}+o\left(\tau^2\right) \\
= & \frac{s'\left(p_0\right)}{2}\tau^2Var\left(\varepsilon\right)+o\left(\tau^2\right)>0. \qquad (27)
\end{aligned}$$

We follow the same steps with the abatement effect and we write:

$$\begin{aligned}
& \Psi\left(\widehat{e}\right)\left(E\left(\pi\left(p^V-\tau e+A(\tau)\right)-\pi\left(p^V-\tau e\right)|e<\widehat{e}\right)\right) \qquad (28) \\
= & \int_{\underline{e}}^{\widehat{e}}\left(\begin{array}{c} s\left(p_0\right)\left(\left(p^V-\tau e+A(\tau)-p_0\right)-\left(p^V-\tau e-p_0\right)\right) \\ +\frac{s'(p_0)}{2}\left(\left(p^V-\tau e+A(\tau)-p_0\right)^2-\left(p^V-\tau e-p_0\right)^2\right)+o\left(\tau^2\right)\end{array}\right)de \\
= & s\left(p_0\right)\frac{\tau^2}{2b''\left(0\right)}\Psi\left(\widehat{e}\right)+o\left(\tau^2\right),
\end{aligned}$$

where terms multiplied by $s'\left(p_0\right)$ disappear in the last line because they are of a higher order.

We proceed with the change in emissions. Using (19), (23) and (25), we can write that at first order the chance in emissions is given by:

$$\begin{aligned}
G^V-G^U = & Cov\left(\varepsilon,s\left(p_0\right)+s'\left(p_0\right)\left(p^V-\tau\varepsilon-p_0\right)\right) \qquad (29) \\
& +\Psi\left(\widehat{e}\right)E\left[-a^*\left(s\left(p_0\right)+s'\left(p_0\right)\left(p^V-\tau\varepsilon-p_0\right)\right)+es'\left(p_0\right)A\left(\tau\right)|e<\widehat{e}\right] \\
& +E\left[e\right]\left(E\left[s\left(p^V-\tau\varepsilon\right)\right]-s\left(p^U-\tau E\left[e\right]\right)\right)+o\left(\tau\right) \\
= & -s'\left(p_0\right)\tau Var\left(\varepsilon\right)-\Psi\left(\widehat{e}\right)\frac{\tau}{b''(0)}s\left(p_0\right)+o\left(\tau\right),
\end{aligned}$$

which is equation (11).

Since v is of the same order as τ, then to obtain a second order approximation in τ of $\left(v-\tau\right)\left(G^V-G^U\right)$, we only to need to take a first order approximation of the change in emissions G^V-G^U (if v were to be of order 0 - that is large relative to τ - then we would need to develop G^V-G^U at second order to get a second order approximation of

$(v-\tau)\left(G^V - G^U\right)$). We can then write the untaxed emissions effect as:

$$-(v-\tau)\left(G^V - G^U\right) = (v-\tau)\, s'(p_0)\, \tau Var(\varepsilon) + (v-\tau)\, \Psi(\widehat{e})\, \frac{\tau}{b''(0)} s(p_0) + o\left(\tau^2\right). \quad (30)$$

We combine (26), (27), (28) and (30), and add the certification costs to get equation (13), which establishes that the expression $W^V - W^U + F\Psi(\hat{e})$ is indeed second order. Then we set $\tau = v$ to derive (10).

Proof of statements a)-b) in Corollary 1: Parts a) and b) directly follow from the fact that $Var(\varepsilon)$ and $\Psi(\widehat{e})$ are increasing in $\widehat{e}$ (as long as distribution of ε is not degenerative).

A.2.2 Under what conditions is τ small?

The previous Taylor expansions are derived for τ small, but what does "small" mean? In this Appendix, we argue that the approximation will be good for τe small relative to p provided that the elasticity of the supply function and that of its derivative are not too large. To fix ideas, we focus on the case where $v = \tau$ and demand is perfectly elastic so that the price is fixed: $p^V = p^U = p_0$ – though our analysis can be extended to the general case. We also ignore the certification costs since those are exact in the approximation. Finally, we assume that $u, c, b \in \mathcal{C}^3$.

We proceed in two steps: First, we do no approximate the cost gains from abatement $A(\tau)$, we (re-)derive the relevant approximation and bound it. Second, we approximate $A(\tau)$ as in Corollary 1.

Approximation for general $A(\tau)$. With fixed price, $\tau = v$, the change in welfare gross of certification costs can be written (exactly) as:

$$W^V - W^U + F\Psi(\hat{e}) = E\left[\pi\left(p_0 - \tau\varepsilon + \mathbb{I}_{e \le \hat{e}} A(\tau)\right)\right] - \pi\left(p_0 - \tau E[e]\right). \quad (31)$$

If $\tau e / p_0$ is small, there is no exit and π is continuously differentiable. Applying Taylor's theorem with remainder, we (exactly) get:

$$\begin{aligned}\pi\left(p_0 - \tau\varepsilon + \mathbb{I}_{e \le \hat{e}} A(\tau)\right) &= \pi(p_0) + s(p_0)\left(\mathbb{I}_{e \le \hat{e}} A(\tau) - \tau\varepsilon\right) + \frac{s'(p_0)}{2}\left(-\tau\varepsilon + \mathbb{I}_{e \le \hat{e}} A(\tau)\right)^2 \quad (32)\\ &+ \frac{s''(\widehat{p}_\varepsilon)}{6}\left(\mathbb{I}_{e \le \hat{e}} A(\tau) - \tau\varepsilon\right)^3,\end{aligned}$$

where $\widehat{p}_{\varepsilon}$ is an unknown price in the interval $(p_0 - \tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau), p_0)$. Similarly, we have

$$\pi(p_0 - \tau E[e]) = \pi(p_0) + s(p_0)(-\tau E[e]) + \frac{1}{2}s'(p_0)(-\tau E[e])^2 + \frac{1}{6}s''\left(\widehat{p}_{E(e)}\right)(-\tau E[e])^3, \quad (33)$$

where $\widehat{p}_{E(e)}$ is an unknown price in the interval $(p_0 - \tau E[e], p_0)$. Plugging (32) and (33) in (31), we obtain:

$$\begin{aligned}
& W^V - W^U + F\Psi(\hat{e}) && (34)\\
= \; & E\left[s(p_0)(-\tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau)) + \frac{1}{2}s'(p_0)(-\tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau))^2\right] \\
& -\left(s(p_0)(-\tau E[e]) + \frac{1}{2}s'(p_0)(-\tau E[e])^2\right) \\
& +\frac{1}{6}\left(E\left[s''(\widehat{p}_{\varepsilon})(-\tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau))^3\right] - s''\left(\widehat{p}_{E(e)}\right)(-\tau E[e])^3\right) \\
= \; & \underbrace{\underbrace{\frac{1}{2}s'(p_0)\tau^2 Var(\varepsilon) + \Psi(\hat{e})s(p_0)A(\tau)}_{\text{terms of the Taylor approximation}} + \underbrace{\Psi(\hat{e})s'(p_0)\left(\frac{1}{2}A(\tau)^2 - \tau E[e|e<\hat{e}]A(\tau)\right)}_{\text{adjustment to abatement gains from changing supply}}}_{\text{linear term}} \\
& \underbrace{+\frac{1}{6}\left(E\left[s''(\widehat{p}_{\varepsilon})(-\tau\varepsilon + \mathbb{I}_{e\leq\hat{e}}A(\tau))^3\right] - s''\left(\widehat{p}_{E(e)}\right)(-\tau E[e])^3\right)}_{\text{error term}}.
\end{aligned}$$

Therefore, we can exactly decompose the welfare gains into a term that corresponds to the welfare gains for linear supply (with $\tau = v$ and a constant price) and an error term. The approximation itself corresponds to the first two terms in the linear term: the reallocation effect $\left(\frac{1}{2}s'(p_0)Var(\varepsilon)\right)$ and the gains from abatement at fixed supply $(s(p_0)\Psi(\hat{e})A(\tau))$. The term $\Psi(\hat{e})s'(p_0)\left(\frac{1}{2}A(\tau)^2 - \tau E[e|e<\hat{e}]A(\tau)\right)$ adjusts the gain from abatement by taking into account that production of certifying firm adjusts because of taxation (by $-s'(p_0)\tau E[e|e<\hat{e}]$) and because of abatement itself (by $s'(p_0)A(\tau)$ – with abatement gains obtained along the way hence the fraction $\frac{1}{2}$).

We seek to establish that the "adjustment to abatement gains from changing supply" is small relative to the "terms of Taylor approximation". To do so, we note that $A(\tau) \leq \tau a \leq \tau \underline{e}$ and therefore get:

$$\frac{\left|\Psi(\hat{e})s'(p_0)\left(\frac{1}{2}A(\tau)^2 - \tau E[e|e<\hat{e}]A(\tau)\right)\right|}{\Psi(\hat{e})s(p_0)A(\tau)} \leq \frac{s'(p_0)\tau\hat{e}}{s(p_0)} = \sigma_s(p_0)\frac{\tau\hat{e}}{p_0}. \quad (35)$$

where $\sigma_s(p_0)$ is the elasticity of the supply function at p_0. Then, for $\tau\hat{e}/p$ small enough and as long as $\sigma_s(p_0)$ is not too large, the expression on the left <<1. Therefore it holds that

$$\frac{\left|\Psi(\hat{e})\, s'(p_0)\left(\frac{1}{2}A(\tau)^2 - \tau E[e|e<\hat{e}]\, A(\tau)\right)\right|}{\frac{1}{2}s'(p_0)\,\tau^2 Var(\varepsilon) + \Psi(\hat{e})\, s(p_0)\, A(\tau)} << 1.$$

In other words the term "adjustment to abatement gains from changing supply" is small relative to the "Taylor approximation".

Furthermore, one can write

$$\frac{s''(\widehat{p}_\varepsilon)(\tau\varepsilon)^3}{s'(p_0)\,\tau^2 Var(\varepsilon)} = \sigma_{s'}(\widehat{p}_\varepsilon)\frac{s'(\widehat{p}_\varepsilon)\, p_0}{\widehat{p}_\varepsilon s'(p_0)}\frac{\tau\varepsilon^3}{p_0 Var(\varepsilon)},$$

where $\sigma_{s'}(\widehat{p}_\varepsilon)$ is the elasticity of the derivative of the supply curve evaluated at $\widehat{p}_\varepsilon$. We then get

$$\frac{s''(\widehat{p}_\varepsilon)(\tau\varepsilon)^3}{s'(p_0)\,\tau^2 Var(\varepsilon)} = \sigma_{s'}(\widehat{p}_\varepsilon)\exp\left(\int_{\widehat{p}_\varepsilon}^{p_0}\frac{-\sigma_{s'}(\rho)}{\rho}d\rho\right)\frac{p_0}{\widehat{p}_\varepsilon}\frac{\tau\varepsilon^3}{p_0 Var(\varepsilon)}.$$

We denote by $|\sigma_{s'}|^{\max} \equiv \max_{(p_0-\tau E[e|e>\hat{e}],p_0)}|\sigma_{s'}|$, we then get

$$\left|\frac{s''(\widehat{p}_\varepsilon)(\tau\varepsilon)^3}{s'(p_0)\,\tau^2 Var(\varepsilon)}\right| \leq |\sigma_{s'}|^{\max}\left(\frac{p_0}{\widehat{p}_\varepsilon}\right)^{1+|\sigma_{s'}|^{\max}}\frac{\tau\varepsilon^3}{p_0 Var(\varepsilon)}.$$

We can then write

$$\begin{aligned}
&\frac{\left|\frac{1}{6}\left(E\left[-s''(\widehat{p}_\varepsilon)(\tau\varepsilon - \mathbb{I}_{e\leq\hat{e}}A(\tau))^3\right] + s''\left(-es(\varphi\widehat{(p^V-\tau e)})\, p_{E(e)}\right)(\tau E[e])^3\right)\right|}{\frac{1}{2}s'(p_0)\,\tau^2 Var(\varepsilon) + \Psi(\hat{e})\, s(p_0)\, A(\tau)} \qquad (36)\\
\leq\ & \frac{1}{3}\left(E\left(\left|\frac{s''(\widehat{p}_\varepsilon)(\tau\varepsilon)^3}{s'(p_0)\,\tau^2 Var(\varepsilon)}\right|\right) + \left|\frac{s''(\widehat{p}_{E(e)})(\tau E(e))^3}{s'(p_0)\,\tau^2 Var(\varepsilon)}\right|\right)\\
\leq\ & \frac{2}{3}|\sigma_{s'}|^{\max}\left(\frac{p_0}{p_0 - \tau E(e|e>\widehat{e})}\right)^{1+|\sigma_{s'}|^{\max}}\frac{E(e|e>\widehat{e})^2}{Var(\varepsilon)}\frac{\tau E(e|e>\widehat{e})}{p_0}.
\end{aligned}$$

In general, the right-hand side will be a lax bound: we have systematically replaced the ratio $\frac{p_0}{\widehat{p}_\varepsilon}$ by its largest possible value, $\sigma_{s'}$ by its largest value in magnitude, and ignored that the left-hand-side features the difference between two terms (which have the same sign if s'' does not flip signs). For a generic distribution of emission rates, $E(e|e>\widehat{e})^2$ and $Var(\varepsilon)$ are commensurate to each other. Therefore, for $\tau E(e|e>\widehat{e})/p_0$ small and as long as $|\sigma_{s'}|^{\max}$ is not too large, we get:

$$\frac{\left|\frac{1}{6}\left(E\left[-s''\left(\widehat{p}_{\varepsilon}\right)\left(\tau\varepsilon-\mathbb{I}_{e\leq\hat{e}}A\left(\tau\right)\right)^{3}\right]+s''\left(\widehat{p}_{E(e)}\right)\left(\tau E\left[e\right]\right)^{3}\right)\right|}{\frac{1}{2}s'\left(p_{0}\right)\tau^{2}Var\left(\varepsilon\right)+\Psi\left(\hat{e}\right)s\left(p_{0}\right)A\left(\tau\right)}<<1,$$

in other words the error term is small relative to the approximation.

Using (34), (35) and (36), we can then bound the relative error between the change in welfare and the approximation as a function of $\frac{\tau e}{p}$:

$$\begin{aligned}&\frac{\left|W^{V}-W^{U}+F\Psi(\hat{e})-\left(\frac{1}{2}s'\left(p_{0}\right)\tau^{2}Var\left(\varepsilon\right)+\Psi\left(\hat{e}\right)s\left(p_{0}\right)A\left(\tau\right)\right)\right|}{\frac{1}{2}s'\left(p_{0}\right)\tau^{2}Var\left(\varepsilon\right)+\Psi\left(\hat{e}\right)s\left(p_{0}\right)A\left(\tau\right)} \qquad (37)\\ \leq\ &\left[\sigma_{s}\left(p_{0}\right)+\frac{2}{3}|\sigma_{s'}|^{\max}\left(\frac{p_{0}}{p_{0}-\tau E\left(e|e>\widehat{e}\right)}\right)^{1+|\sigma_{s'}|^{\max}}\frac{E\left(e|e>\widehat{e}\right)^{2}}{Var\left(\varepsilon\right)}\right]\frac{\tau E\left(e|e>\widehat{e}\right)}{p_{0}},\end{aligned}$$

where for reasonable elasticities, the bound is small when $\tau E\left(e|e>\widehat{e}\right)/p_{0}$ is small (recall that this will generally be a lax bound). That is, we get:

$$W^{V}-W^{U}+F\Psi(\hat{e})=\left(\frac{1}{2}s'\left(p_{0}\right)\tau^{2}Var\left(\varepsilon\right)+\Psi\left(\hat{e}\right)s\left(p_{0}\right)A\left(\tau\right)\right)\left(1+o\left(\frac{\tau E\left(e|e>\widehat{e}\right)}{p_{0}}\right)\right).$$

Approximating $A\left(\tau\right)$. In Corollary 1, we go further since we also do a Taylor expansion of $A\left(\tau\right)$. We do this in a second step here to recognize that additional assumptions are necessary to ensure that $A\left(\tau\right)$ can be well approximated. Recall that $A\left(\tau\right)=\tau a\left(\tau\right)-b\left(a\left(\tau\right)\right)$ where $a\left(\tau\right)=b'^{-1}\left(\tau\right)$, so that $A'\left(\tau\right)=a\left(\tau\right)$ and $A''\left(\tau\right)=1/b''\left(b'^{-1}\left(\tau\right)\right)$. With $b\left(0\right)=b'\left(0\right)=0$, we get that $A\left(0\right)=A'\left(0\right)=0$. Applying again Taylor's theorem with remainder, we get:

$$A\left(\tau\right)=\frac{1}{b''\left(0\right)}\frac{\tau^{2}}{2}+A'''\left(\widetilde{\tau}\right)\frac{\tau^{3}}{6},$$

for some $\widetilde{\tau}\in\left(0,\tau\right)$ and $A'''\left(\tau\right)=-\frac{b'''\left(b'^{-1}\left(\tau\right)\right)}{\left(b''\left(b'^{-1}\left(\tau\right)\right)\right)^{3}}$. We then get a bound on the approximation of $A\left(\tau\right)$ by its second order Taylor expansion $\frac{1}{b''\left(0\right)}\frac{\tau^{2}}{2}$, namely

$$\left|\frac{A\left(\tau\right)-\frac{1}{b''\left(0\right)}\frac{\tau^{2}}{2}}{\frac{1}{b''\left(0\right)}\frac{\tau^{2}}{2}}\right|\leq\frac{\tau}{3}B\text{ where }B=b''\left(0\right)\max_{\widetilde{a}\in\left(0,b'^{-1}\left(\tau\right)\right)}\left|\frac{b'''\left(\widetilde{a}\right)}{\left(b''\left(\widetilde{a}\right)\right)^{3}}\right|. \qquad (38)$$

Naturally, the bound tends to 0 as τ tends to 0. In addition, the approximation will be better if the abatement curve does not deviate too much from a quadratic curve (i.e. when

b''' is small compared to $(b'')^2$). Combined (37) and (38), we obtain an overall bound for the approximation as:

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

$$\leq \left|\frac{W^V - W^U + F\Psi(\hat{e}) - \left(\frac{1}{2}s'(p_0)\,\tau^2 Var(\varepsilon) + \Psi(\hat{e})\,s(p_0)\,A(\tau)\right)}{\frac{1}{2}s'(p_0)\,\tau^2 Var(\varepsilon) + \Psi(\hat{e})\,s(p_0)\,A(\tau)}\right|\left(1+\frac{\tau}{3}B\right)+\frac{\tau}{3}B$$

$$\leq \left[\sigma_s(p_0) + \frac{2}{3}|\sigma_{s'}|^{\max}\left(\frac{p_0}{p_0 - \tau E(e|e>\widehat{e})}\right)^{1+|\sigma_{s'}|^{\max}}\frac{E(e|e>\widehat{e})^2}{Var(\varepsilon)}\right]\frac{\tau E(e|e>\widehat{e})}{p_0}\left(1+\frac{\tau}{3}B\right)+\frac{\tau}{3}B.$$

This bound depends on the supply elasticity, the elasticity of the supply slope, the distribution of emission rates, and a measure of how much the abatement function deviates from quadratic.

A.3 Optimal policy: Proof of Proposition 2

In this Appendix, we study the constrained optimal policy of Section 2.3. We start by establishing the first part of Proposition 2 (on the general case), then we derive conditions under which the set of first order conditions identifies the optimum uniquely, and finally we show the second part of Proposition 2 (focused on small v).

A.3.1 Optimum in the general case

Consider a social planner who can only use an emission tax, τ, an output tax, t, and a certification tax/subsidy, f, as policy instruments. Then non-certifying firms earn profits $\pi(p-t)$, independent of their emission rate e, while certifying firms earn profits net of certification costs, $\pi(p-\tau e+A(\tau))-(F+f)$, decreasing in e. Therefore, the set of certifying firms must be empty or an interval of the type $[\underline{e},\widehat{e}]$.

To solve for the constrained optimum, we start by solving for the more general problem of the optimal allocation conditional on the constraint that firms certify if and only if they are on an interval $[\underline{e},\widehat{e}]$. Non-certified firms must then produce the same amount denoted q^U

while certified firms produce $q^V(e)$. That problem can be written as:

$$\begin{aligned}\max_{\widehat{e},a(e),q^V(e),q^U} W &= u\left(\int_{\underline{e}}^{\widehat{e}} q^V(e)\,\psi(e)\,de + (1-\Psi(\widehat{e}))\,q^U\right) \\ &\quad -v\left(\int_{\underline{e}}^{\widehat{e}} (e-a(e))\,q^V(e)\,\psi(e)\,de + (1-\Psi(\widehat{e}))\,E[e|e>\widehat{e}]\,q^U\right) \\ &\quad -\int_{\underline{e}}^{\widehat{e}} \left(c\left(q^V(e)\right) + b(a(e))\,q^V(e)\right)\psi(e)\,de + (1-\Psi(\widehat{e}))\,c\left(q^U\right) - F\Psi(\widehat{e})\end{aligned}$$

The first-order condition with respect to $q^V(e)$ is given by:

$$u'(C) - \left(c'\left(q^V(e)\right) + b(a(e))\right) - (e-a(e))\,v = 0 \tag{39}$$

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 $a(e)$ leads to

$$a(e) = a^* = b'^{-1}(v),$$

which is how much certified firms abate in equilibrium if they face an emission tax $\tau = v$. Denoting $p = u'(C)$ the shadow price, we get

$$q^V(e) = s(p - v(e-a) + b(a)).$$

The first-order condition with respect to q^U is given by:[2]

$$u'(C) - c'\left(q^U\right) - vE[e|e>\widehat{e}] = 0 \Leftrightarrow q^U = s(p - vE[e|e>\widehat{e}]). \tag{40}$$

Therefore, for a given $\widehat{e}$, the optimal allocation can be decentralized by setting an emission tax $\tau = v$ and an output tax $t = vE[e|e>\hat{e}]$.

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

$$u'(C)\left(q^V(\widehat{e}) - q^U\right) - \left(c\left(q^V(\widehat{e})\right) + b(a(\widehat{e}))\,q^V(\widehat{e}) - c\left(q^U\right)\right) - v\left((\widehat{e} - a(\widehat{e}))\,q^V(\widehat{e}) - \widehat{e}q^U\right) - F = 0, \tag{41}$$

which we can rewrite as

$$\pi(p - v\widehat{e} + A(v)) - \pi(p - vE[e|e>\widehat{e}]) = F + v(E[e|e>\hat{e}] - \widehat{e})\,s(p - vE[e|e>\widehat{e}]). \tag{42}$$

[2]We ignore here the possibility that the optimum features exit for uncertified firms. The analysis could be generalized to that case, but it will not occur when v is small which is our leading example.

Therefore, the optimum can be decentralized with a tax on certification given by equation (14).

A corner solution for the certification threshold is characterized by

$$\pi\left(p-v\underline{e}+A\left(v\right)\right)-\pi\left(p-vE(e)\right)<F+v\left(E(e)-\underline{e}\right)s\left(p-vE\left[e\right]\right)\ \text{ for } \widehat{e}=\underline{e}, \tag{43}$$

$$\pi\left(p-v\overline{e}+A\left(v\right)\right)-\pi\left(p-v\overline{e}\right)>F \text{ for } \widehat{e}=\overline{e}, \tag{44}$$

which can both be decentralized by a certification tax that follows the same formula given by equation (14).

A.3.2 Uniqueness

In general, there could be multiple solutions to the set of equations (42), (43), and (44) that identifies $\widehat{e}$. We now derive sufficient conditions under which this set of equations has a unique solution (namely the optimal $\widehat{e}$). Specifically, we show that the solution is unique if s is weakly convex or $\tau(=v)$ is small.

Whether there is a unique solution to the first order conditions depends on whether the function g^* is monotonic or not, where

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

We get

$$\begin{aligned}
& g^{*\prime}\left(\widetilde{e}\right)\\
= & -v\left(s\left(p-\tau\widetilde{e}+A\left(\tau\right)\right)-s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)-\tau\frac{dE\left(e|e>\widetilde{e}\right)}{d\widetilde{e}}\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)\right)\\
& +\frac{dp}{d\widetilde{e}}\left(s\left(p\left(\widetilde{e}\right)-\tau\widetilde{e}+A\left(\tau\right)\right)-s\left(p\left(\widetilde{e}\right)-\tau E\left(e|e>\widetilde{e}\right)\right)-\tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p\left(\widetilde{e}\right)-\tau E\left(e|e>\widetilde{e}\right)\right)\right)
\end{aligned}$$

Using (5), we get that the term in $\frac{dp}{d\widetilde{e}}$ must be weakly negative. Assume that s is weakly convex then

$$\begin{aligned}
s\left(p-\tau\widetilde{e}+A\left(\tau\right)\right)-s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right) &\geq s\left(p-\tau\widetilde{e}\right)-s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)\\
&\geq \tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s'\left(p-\tau E\left(e|e>\widetilde{e}\right)\right),
\end{aligned}$$

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 $g^{*\prime}\left(\widetilde{e}\right)<0$.

For small τ, we get:

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

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

A.3.3 Optimum in the small v case

We now derive an approximate solution to the optimal $\widehat{e}$ for small v. Taking a second order approximation of (42), we obtain that for an interior solution:

$$\begin{aligned} & s(p_0)\left(p^V - v\widehat{e} + A(v) - p_0\right) + \frac{s'(p_0)}{2}\left(p^V - v\widehat{e} + A(v) - p_0\right)^2 \\ & - s(p_0)\left(p^V - vE(e|e > \hat{e}) - p_0\right) - \frac{s'(p_0)}{2}\left(p^V - vE(e|e > \hat{e}) - p_0\right)^2 \\ = \; & F + v\left(E(e|e > \hat{e}) - \widehat{e}\right)\left(s(p_0) + s'(p_0)\left(p^V - vE(e|e > \hat{e}) - p_0\right)\right) + o\left(v^2\right). \end{aligned}$$

Rearranging terms, this gives equation (15). Similarly, we obtain that in corner cases:

$$\frac{v^2}{2} s'(p_0)\left(E([e] - \underline{e}\right)^2 + \frac{v^2}{2}\frac{s(p_0)}{b''(0)} < F + o\left(v^2\right) \text{ for } \widehat{e} = \underline{e}, \tag{45}$$

$$\frac{v^2}{2}\frac{s(p_0)}{b''(0)} > F + o\left(v^2\right) \text{ for } \widehat{e} = \overline{e}. \tag{46}$$

As derived above, the system (15), (45), and (46) uniquely charecterizes the optimal $\widehat{e}$. If F is large relative to v^2 (e.g. if F is first order in v), then (45) holds and there is no certification: $\widehat{e} = \underline{e}$. If F is small relative to v^2 (for instance F is third order in v), then (46) holds and there is full certification.[3] If F is second order in v, then all three cases are possible.

A.4 Proof of Proposition 3

We prove Proposition 3, first with a fixed and then an iteratively adjusted certification tax.

A.4.1 Case with a fixed (or no) certification tax.

We prove the claim by induction. For the first step, we have $\widehat{e}_0 = \underline{e} < \hat{e}$ and a higher value in the following round, $\widehat{e}_1 > \widehat{e}_0$, unless no firm certifies. If $\pi(p - \tau\overline{e} + A(\tau)) - F - f \geq \pi(p - \tau E(e))$, then all firms certify in the first round. In that case, we also have

[3] Without abatement, we would not have full certification in that case, but nearly full certification: $\widehat{e} = \overline{e} + o(v)$.

that $\pi(p-\tau\overline{e}+A(\tau))-F-f\geq\pi(p-\tau E(e))>\pi(p-\tau\overline{e})$, so that full certification does correspond to the equilibrium.

Assume instead that $\pi(p-\tau\underline{e}+A(\tau))-F-f\leq\pi(p-\tau E(e))$, then no firm certifies and this is indeed the equilibrium.

Finally, if $\pi(p-\tau\overline{e}+A(\tau))-F-f>\pi(p-\tau E(e))>\pi(p-\tau\underline{e}+A(\tau))-F$, then there exists an interior solution $\widehat{e}_1$, which satisfies:

$$\begin{aligned}\pi(p-\tau\widehat{e}_1+A(\tau))-F-f &= \pi(p-\tau E[e|e>\widehat{e}_0]) \\ &> \pi(p-\tau E[e|e>\widehat{e}_1]).\end{aligned}$$

The last inequality implies that $g(\hat{e}_1)>F+f$, with g defined as in (3). Since g is decreasing when $g(e)\geq F+f$ (see Lemma 2), we must have $\widehat{e}_1<\widehat{e}$.

Assume now that we have $\widehat{e}_0<\widehat{e}_1<...<\widehat{e}_{n-1}<\widehat{e}$ and consider step n. Again, if $\pi(p-\tau\overline{e}+A(\tau))-F-f\geq\pi(p-\tau E[e|e>\widehat{e}_{n-1}])$, then the algorithm immediately converges toward full unraveling, which corresponds to the equilibrium.

Assume instead that $\pi(p-\tau\widehat{e}_{n-1}+A(\tau))-F-f\leq\pi(p-\tau E[e|e>\widehat{e}_{n-1}])$. Since $\widehat{e}_{n-1}$ is an interior solution, we must have that $\pi(p-\tau\widehat{e}_{n-1}+A(\tau))-F-f=\pi(p-\tau E[e|e>\widehat{e}_{n-2}])>\pi(p-\tau E[e|e>\widehat{e}_{n-1}])$, as we have assumed that $\widehat{e}_{n-1}>\widehat{e}_{n-2}$. Therefore, this case is impossible.

We must then have $\pi(p-\tau\overline{e}+A(\tau))-F-f>\pi(p-\tau E[e|e>\widehat{e}_{n-1}])>\pi(p-\tau\widehat{e}_{n-1}+A(\tau))-F$, which ensures that there is a unique interior solution $\widehat{e}_n>\widehat{e}_{n-1}$, defined according to

$$\pi(p-\tau\widehat{e}_n+A(\tau))-F-f=\pi(p-\tau E[e|e>\widehat{e}_{n-1}]). \tag{47}$$

Furthermore, using that $\widehat{e}_{n-1}<\widehat{e}_n$, we must have $g(\hat{e}_n)>F+f$, which ensures that $\widehat{e}_n<\widehat{e}$,

By induction, we get that $\widehat{e}_n$ must converge monotonically towards the equilibrium $\widehat{e}$ which is uniquely defined as a bound or the fixed point of (47).

A.4.2 Case with an iteratively adjusted certification tax.

With the optimal certification tax, the optimal threshold e^* may be $e^*=\overline{e}$ if $\pi(p-\tau\overline{e}+A(\tau))-F>\pi(p-\tau\overline{e})$; $e^*=\underline{e}$ if

$$\pi(p-\tau\underline{e}+A(\tau))-F-\tau(E[e]-\underline{e})\,s(p-\tau E[e])<\pi(p-\tau E[e]); \tag{48}$$

or a solution to:

$$\begin{aligned} & \pi\left(p-\tau e^{*}+A\left(\tau\right)\right)-F-\tau\left(E\left[e|e>e^{*}\right]-e^{*}\right)s\left(p-\tau E\left[e|e>e^{*}\right]\right) \\ = \;& \pi\left(p-\tau E\left[e|e>e^{*}\right]\right). \end{aligned} \tag{49}$$

We define the function

$$k\left(\widetilde{e}\right)\equiv\pi\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)+\tau\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right),$$

which is decreasing in e as

$$k'\left(\widetilde{e}\right)=-\tau s\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)-\tau^{2}\left(E\left(e|e>\widetilde{e}\right)-\widetilde{e}\right)\frac{dE\left(e|e>\widetilde{e}\right)}{d\widetilde{e}}s'\left(p-\tau E\left(e|e>\widetilde{e}\right)\right)<0.$$

The optimal threshold e^{*} must then satisfy $\pi\left(p-\tau e^{*}+A\left(\tau\right)\right)-F=k\left(e^{*}\right)$ for an interior solution ($>$ for $e^{*}=\overline{e}$ and $<$ $\underline{e}$).

We proceed in two steps: First, we show that the algorithm converges toward e_1^{*} which is $\underline{e}$ if (48) holds, otherwise the smallest solution to (49) if there is one, and $\overline{e}$ otherwise. Second, we derive conditions under which the set of first-order conditions uniquely identifies the optimum, so that $e_1^{*}=e^{*}$.

Algorithm convergence. We follow a similar logic to the previous case and prove the result by induction. We first note that at each step, $\widehat{e}_n$ will be uniquely defined as a bound or as the solution to:

$$\pi\left(p-\tau\widehat{e}_n+A\left(\tau\right)\right)-F=k\left(\widehat{e}_{n-1}\right). \tag{50}$$

For a given $\widehat{e}_{n-1}$, this problem is uniquely defined because the LHS is decreasing in $\widehat{e}_n$.

Step $n=1$. If no firm certifies, then $\widehat{e}_0=\underline{e}=e_1^{*}$ and the process stops, otherwise, $\widehat{e}_1>\underline{e}$. Consider then the case where $\pi\left(p-\tau\overline{e}+A\left(\tau\right)\right)-F\geq k\left(\underline{e}\right)>k\left(\overline{e}\right)$. Since k is decreasing and $\pi\left(p-\tau e+A\left(\tau\right)\right)$ is increasing, we have that for any $e\in\left(\underline{e},\overline{e}\right)$, $\pi\left(p-\tau e+A\left(\tau\right)\right)-F>k\left(e\right)$, which ensures that there is no interior solution to (49). Instead, we have that $e_1^{*}=\overline{e}$ and the algorithm immediately converges toward $\overline{e}$.

Assume now that $\pi\left(p-\tau\overline{e}+A\left(\tau\right)\right)-F<k\left(\underline{e}\right)$, then there exists an $\widehat{e}_1\in\left(\widehat{e}_0,\overline{e}\right)$ such that

$$\pi\left(p-\tau\widehat{e}_1+A\left(\tau\right)\right)-F=k\left(\underline{e}\right)>k\left(\widehat{e}_1\right).$$

For all $e \in (\widehat{e}_0, \widehat{e}_1)$, we have that

$$\pi\left(p - \tau e + A\left(\tau\right)\right) - F > k\left(\underline{e}\right) > k\left(e\right),$$

which guarantees that $\widehat{e}_1 < e_1^*$, as e_1^* is the smallest solution to $\pi\left(p - \tau e + A\left(\tau\right)\right) - F = k\left(e\right)$. Therefore, we get $\widehat{e}_0 < \widehat{e}_1 < e_1^*$.

Step n. We assume that $\widehat{e}_0 < \widehat{e}_1 < ... < \widehat{e}_{n-1} < e_1^*$. If

$$\pi\left(p - \tau \overline{e} + A\left(\tau\right)\right) - F > k\left(\widehat{e}_{n-1}\right), \tag{51}$$

then we must have that for any $e \in (\widehat{e}_{n-1}, \overline{e}]$, we have $\pi\left(p - \tau e + A\left(\tau\right)\right) - F > k\left(e\right)$. Given that the process did not converge before, this ensures that $e_1^* = \overline{e}$ and $\widehat{e}_n = e_1^*$.

Assume now instead that

$$\pi\left(p - \tau \widehat{e}_{n-1} + A\left(\tau\right)\right) - F < k\left(\widehat{e}_{n-1}\right). \tag{52}$$

As $\widehat{e}_{n-1}$ is an interior solution of the previous step, we must have

$$\pi\left(p - \tau \widehat{e}_{n-1} + A\left(\tau\right)\right) - F = k\left(\widehat{e}_{n-2}\right).$$

Since $\widehat{e}_{n-1} > \widehat{e}_{n-2}$ and k is decreasing, then (52) cannot hold.

Therefore, as long as (51) is violated, there must exist an $\widehat{e}_n \in (\widehat{e}_{n-1}, \overline{e})$, such that

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

In addition for all $e \in (\widehat{e}_{n-1}, \widehat{e}_n]$, we get that $\pi\left(p - \tau e + A\left(\tau\right)\right) - F \geq k\left(\widehat{e}_{n-1}\right) > k\left(e\right)$. This establishes that there is no solution to (49) in $(\widehat{e}_{n-1}, \widehat{e}_n)$, and we get that $\widehat{e}_0 < \widehat{e}_1 < ... < \widehat{e}_{n-1} < \widehat{e}_n < e_1^*$.

Conclusion. By induction, the algorithm converges toward e_1^*, which is the smallest solution to a bound or a fixed point of (49). As demonstrated in Section A.3.2, that solution is unique if s is weakly convex or τ is small.

A.5 Extensions

We consider in turn three extensions: proportional abatement, free entry and heterogeneity in productivity. These extensions are discussed in the text in Section 2.5.

A.5.1 Proportional Abatement

We now assume that by spending $b(a)$ per unit of output a firm can reduce its emission rate by a factor a (i.e. from e to $e(1-a)$). 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: $a^*(e) = b'^{-1}(\tau e)$. The effective price per unit received by a certifying firm, $p(e) = p - \tau e(1-a) - b(a)$, 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 $A(\tau, e) \equiv \tau ea - b(a) = (\tau e)^2/(2b''(0)) + o(\tau^2)$. Equation (8) becomes

$$\begin{aligned} G^V - G^U &= Cov\left[\varepsilon, s\left(p^V - \tau\varepsilon\right)\right] + \Psi(\hat{e})E\left\{e(1-a^*(e))\, s\left(p^V - \tau e + A(\tau, e)\right) - es\left(p^V - \tau e\right)|e \leq \hat{e}\right\} \\ &\quad + E[e]\left\{E\left[s\left(p^V - \tau\varepsilon\right)\right] - s\left(p^U - \tau E[e]\right)\right\}, \end{aligned}$$

while equation (6) is still valid provided that $A(\tau)$ is replaced by $A(\tau, e)$. Taking Taylor expansions as before, we then obtain

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

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

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

A.5.2 Heterogeneous Productivity

In this Appendix, we consider the extension with heterogeneous productivity presented in Section 2.5. As mentioned in the text, we assume that the supply function s is isoelastic: $s(x) \equiv s_0 x^\alpha$ with $\alpha > 0$. As a result, certified firms produce $q_i = s(\varphi_i(p - \tau e + A(\tau)))$ and uncertified firms $q^u = s(\varphi_i(p - t))$. The quantity-weighted uniform output tax is given by $t = \tau\tilde{E}(e)$, where

$$\tilde{E}(e) = \int_\varphi\int_e e\tilde{\psi}(\varphi, e)ded\varphi = \frac{G^U}{S^U},$$

where $\tilde{\psi}(\varphi, e)$ is the density distribution rescaled by output, which is proportional to φ^α under both laissez-faire and uniform output taxation:

$$\tilde{\psi}(\varphi, e) = \varphi^\alpha \psi(\varphi, e) / \left(\int_\varphi \int_e \varphi^\alpha \psi(\varphi, e) de d\varphi \right).$$

Similarly, under the certification mechanism, uncertified firms are taxed at the rate $t = \tau \frac{G^V_{uncertified}}{S^V_{uncertified}}$ where $G^V_{uncertified}$ is the aggregate emissions of uncertified firms and $S^V_{uncertified}$ their aggregate supply. The ratio $\frac{G^V_{uncertified}}{S^V_{uncertified}}$ is equal to the conditional expectation of emission rates computed using the density $\tilde{\psi}$:

$$\begin{aligned} \frac{G^V_{uncertified}}{S^V_{uncertified}} &= \frac{\int_{\varphi,e} es\left(\varphi\left(p-t\right)\right)\psi\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi}{\int_{\varphi,e} s\left(\varphi\left(p-t\right)\right)\psi\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi} \\ &= \frac{\int_{\varphi,e} e\varphi^\alpha\psi\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi}{\int_{\varphi,e} \varphi^\alpha\psi\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi} \\ &= \frac{\int_{\varphi,e} e\tilde{\psi}\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi}{\int_{\varphi,e} \tilde{\psi}\left(\varphi,e\right)1_{e>\widehat{e}_\varphi} de d\varphi} = \tilde{E}\left(e|e>\widehat{e}_\varphi\right). \end{aligned}$$

As before, we can define the pre-abatement taxed emission rate as $\varepsilon = e$ if $e < \widehat{e}_\varphi$ and $\varepsilon = \tilde{E}\left(e|e > \widehat{e}_\varphi\right)$ otherwise. In this expression, the expectation is taken unconditionally on φ, so that all uncertified firms pay the same tax regardless of their productivity (but whether they certify or not depends on their productivity). We we still have $\tilde{E}(\varepsilon) = \tilde{E}(e)$, the average (quantity weighted) pre-abatement taxed emission rate is independent of the threshold function $\widehat{e}_\varphi$. We then establish:

Corollary 2. *Assume that $b'(0) = 0$. The welfare and emission changes from introducing certification are given by:*

$$W^V - W^U = \tau\left(v - \frac{\tau}{2}\right)S(p_0)\left(\frac{\alpha}{p_0}V\tilde{ar}(\varepsilon) + \frac{\tilde{\Psi}\left(e < \widehat{e}_\varphi\right)}{b''(0)}\right) - F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \text{ and} \tag{53}$$

$$G^V - G^U = -\tau S\left(p_0\right)\left(\frac{\alpha}{p_0}V\tilde{ar}(\varepsilon) + \frac{\tilde{\Psi}\left(e < \widehat{e}_\varphi\right)}{b''(0)}\right) + o(\tau), \tag{54}$$

where $S\left(p_0\right)$ is aggregate supply in laissez-faire, $V\tilde{ar}(\varepsilon)$ is the output-weighted variance of pre-abatement taxed emission rates: $V\tilde{ar}(\varepsilon) = \int_\varphi \int_e (\varepsilon - \tilde{E}(\varepsilon))^2 \tilde{\psi}(\varphi, e) de d\varphi$, and $\tilde{\Psi}\left(e < \widehat{e}_\varphi\right)$ is the output weighted share of firms which certify $\tilde{\Psi}\left(e < \widehat{e}_\varphi\right) = \int_{\varphi,e} \tilde{\psi}\left(\varphi, e\right) 1_{e<\widehat{e}_\varphi} de d\varphi$.

We note that in the computation of the variance, we can still weigh outputs at the laissez-faire (or uniform output tax) level of production, since doing it at the equilibrium level under certification only adds terms of higher order). Further, the term $\alpha S(p_0)/p_0$ corresponds to the slope of the aggregate supply curve $S'(p_0)$ so that these expressions exactly mirror the formulas in the baseline case given in Corollary 1

Proof. In the certification equilibrium, we can write aggregate output as:

$$S^V(p) = \int_\varphi \left(\int_{\underline{e}}^{\widehat{e}_\varphi} s(\varphi(p - \tau e + A(\tau)))\,\psi(e|\varphi)\,de + s\left(\varphi\left(p - t^V\right)\right)(1 - \Psi(\widehat{e}_\varphi)) \right) \psi_\varphi(\varphi)d\varphi, \tag{55}$$

where $t^V = \tau\tilde{E}(e|e > \widehat{e}_\varphi)$ denotes the output tax faced by uncertified firms, $\psi_\varphi(\varphi)$ is the (unconditional) distribution of productivity and $\psi_e(e|\varphi)$ the distribution of emissions conditional on productivity. We let t^U denotes the tax when no certification is in place ($\hat{e}(\varphi) = \underline{e}$ for all φ). We similarly get that aggregate emissions are given by

$$\begin{aligned} & G^V(p) \\ = & \int_\varphi \begin{pmatrix} \int_{\underline{e}}^{\widehat{e}_\varphi} (e - a^*)\, s(\varphi(p - \tau e + A(\tau)))\,\psi(e|\varphi)\,de \\ + (1 - \Psi(\widehat{e}_\varphi))\, E(e|e > \widehat{e}_\varphi, \varphi)\, s\left(\varphi\left(p - t^V\right)\right) \end{pmatrix} \psi_\varphi(\varphi)d\varphi. \end{aligned} \tag{56}$$

Welfare is still given by (2), so that we can follow the same steps as in Appendix (A.1.2) and write the welfare change as:

$$\begin{aligned} & W^V - W^U \\ = & \underbrace{\int_{p^U}^{p^V} \left[\left(\int_\varphi s\left(\varphi\left(p - t^U\right)\right) \psi_\varphi(\varphi) \right) d\varphi - D(p) \right] dp}_{\text{price effects}} \\ & \underbrace{+ \int_\varphi \frac{1}{\varphi} \begin{pmatrix} \int_{\underline{e}}^{\widehat{e}_\varphi} \pi\left(\varphi\left(p^V - \tau e\right)\right) \psi_e(e|\varphi)\,de \\ + (1 - \Psi_e(\widehat{e}_\varphi|\varphi))\, \pi\left(\varphi\left(p^V - t^V\right)\right) - \pi\left(\varphi\left(p^V - t^U\right)\right) \end{pmatrix} \psi_\varphi(\varphi)\,d\varphi}_{\text{reallocation gains}} \\ & \underbrace{+ \int_\varphi \frac{1}{\varphi} \int_{\underline{e}}^{\widehat{e}_\varphi} \left(\pi\left(\varphi\left(p^V - \tau e + A(\tau)\right)\right) - \pi\left(\varphi\left(p^V - \tau e\right)\right) \right) \psi_e(e|\varphi)\,de\psi_\varphi(\varphi)\,d\varphi}_{\text{abatement effect}} \\ & - \underbrace{(v - \tau)\left(G^V - G^U\right)}_{\text{untaxed emissions}} - F \int_\varphi \psi_\varphi(\varphi)\, \Psi_\varphi(\widehat{e}_\varphi|\varphi)\,d\varphi. \end{aligned} \tag{57}$$

Using (56), the change in emissions itself is given by:

$$
\begin{aligned}
& G^V - G^U \\
= & \underbrace{\int_\varphi \psi_\varphi(\varphi)\left(s\left(\varphi\left(p^V - t^U\right)\right) - s\left(\varphi\left(p^U - t^U\right)\right)\right) E\left(e|\varphi\right) d\varphi}_{\text{price effect}} \\
& \underbrace{+\int_\varphi \psi_\varphi(\varphi)\, d\varphi \left[\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_\varphi} s\left(\varphi\left(p^V - \tau e\right)\right) e\psi_e\left(e|\varphi\right) de \\ +\left(1 - \Psi_e\left(\widehat{e}_\varphi|\varphi\right)\right) s\left(\varphi\left(p^V - t^V\right)\right) E\left[e|e > \widehat{e}_\varphi, \varphi\right] \\ -E\left(e|\varphi\right) s\left(\varphi\left(p^V - t^U\right)\right) \end{array}\right]}_{\text{reallocation}} \\
& \underbrace{+\int_\varphi \int_{\underline{e}}^{\widehat{e}_\varphi} \left(\left(e - a^*\right) s\left(\varphi\left(p^V - \tau e + A\left(\tau\right)\right)\right) - es\left(\varphi\left(p^V - \tau e\right)\right)\right) \psi_e\left(e|\varphi\right) de \psi_\varphi\left(\varphi\right) d\varphi.}_{\text{abatement}}
\end{aligned}
\tag{58}
$$

We now use Taylor expansions to simplify these expressions. We differentiate (55) to get:

$$
\begin{aligned}
S^V(p) &= S(p) \\
& \quad + \frac{\alpha s_0 p^\alpha}{p} \int_\varphi \left(\int_{\underline{e}}^{\widehat{e}_\varphi} \varphi^\alpha\left(-\tau e + A\left(\tau\right)\right) \psi\left(e|\varphi\right) de + \varphi^\alpha\left(-t^V\right)\left(1 - \Psi\left(\widehat{e}_\varphi\right)\right)\right) \psi_\varphi(\varphi) d\varphi + o(\tau) \\
&= S(p)\left(1 - \tau \frac{\alpha}{p} \tilde{E}(\varepsilon)\right) + o(\tau),
\end{aligned}
$$

where $S(p) = s_0 p^\alpha \int_\varphi \varphi_i^\alpha \psi_\varphi(\varphi) d\varphi$ is total production at price p in laissez-faire and where we used that $A(\tau)$ is second order in τ. As $\tilde{E}(\varepsilon) = \tilde{E}(e)$, we get 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 the price effect in (57) is still 0 at second order.

Further, a second order Taylor expansion of the reallocation gains gives:

$$
\begin{aligned}
&\text{Reallocation gains} \\
= &\int_{\varphi} \frac{\psi_{\varphi}(\varphi)}{\varphi} \left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_{\varphi}} \left(\varphi s(\varphi p_0)\left(t^U - \tau e\right) + \frac{\varphi^2 s'(\varphi p_0)\left(\left(p^V - \tau e - p_0\right)^2 - \left(p^V - t^U - p_0\right)^2\right)}{2} \right) \psi_e(e|\varphi)\, de \\ + \left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right) \left[\varphi s(\varphi p_0)\left(t^U - t^V\right) + \frac{\varphi^2 s'(\varphi p_0)\left(\left(p^V - t^V - p_0\right)^2 - \left(p^V - t^U - p_0\right)^2\right)}{2} \right] \end{array} \right) d\varphi \\
&+ o\left(\tau^2\right) \\
= &\int_{\varphi} \psi_{\varphi}(\varphi)\, \varphi s'(\varphi p_0) \left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_{\varphi}} \left(\left(t^U - \tau e\right)\left(p^V - p_0 - \frac{\tau e + t^U}{2}\right)\right) \psi_e(e|\varphi)\, de \\ + \left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right)\left(t^U - t^V\right)\left(p^V - p_0 - \frac{t^V + t^U}{2}\right) \end{array} \right) d\varphi + o\left(\tau^2\right) \\
= &\tau \int_{\varphi, e} \psi(e, \varphi)\, \alpha s_0 \varphi^{\alpha} p_0^{\alpha - 1} \left(\left(\tilde{E}(e) - \varepsilon\right) \left(p^V - p_0 - \frac{\tau\varepsilon + \tau\tilde{E}(e)}{2}\right)\right) d\varphi + o\left(\tau^2\right) \\
= &\tau \frac{\alpha S(p_0)}{p_0} \int_{\varphi, e} \tilde{\psi}(e, \varphi) \left(\left(\tilde{E}(e) - \varepsilon\right)\left(p^V - p_0 - \tau\tilde{E}(e)\right) + \frac{\tau}{2}\left(\varepsilon - \tilde{E}(e)\right)^2 \right) d\varphi + o\left(\tau^2\right) \\
= &\tau^2 S(p_0)\ \frac{1}{2}\frac{\alpha}{p_0} V a\tilde{r}(\varepsilon)\ + o(\tau^2), \qquad (59)
\end{aligned}
$$

where we used that $\tilde{E}(\varepsilon) = \tilde{E}(e)$ and where $Va\tilde{r}(\varepsilon)$ is the output-weighted variance of the pre-abatement taxed emission rate (i.e. the variance of ε using $\tilde{\psi}$ as a pdf).

Next, a second order Taylor approximation of the abatement effect gives:

$$
\begin{aligned}
\text{Abatement gains} &= A(\tau) \int_{e, \varphi} s(\varphi p_0)\, \psi(\varphi, e)\, de d\varphi + o\left(\tau^2\right) \qquad (60) \\
&= \frac{\tau^2 S(p_0) \tilde{\Psi}\left(e < \hat{e}(\varphi)\right)}{2b''(0)} + o\left(\tau^2\right),
\end{aligned}
$$

where we used that $A(\tau) = \frac{\tau^2}{2b''(0)} + o\left(\tau^2\right)$.

Further, in a first order expansion of (58), we obtain:

$$
\begin{aligned}
& G^V - G^U \\
= & \int_\varphi s'(\varphi p_0)\varphi\psi_\varphi(\varphi)\,d\varphi \left[\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_\varphi}\left(p^V-\tau e-p_0\right)e\psi_e(e|\varphi)\,de \\ +\left(1-\Psi_e(\widehat{e}_\varphi|\varphi)\right)\left(p^V-t^V-p_0\right)E\left(e|e>\widehat{e}_\varphi,\varphi\right) \\ -E(e|\varphi)\left(p^V-t^U-p_0\right) \end{array}\right] \\
& -a^*\int_\varphi\int_{\underline{e}}^{\widehat{e}_\varphi} s(\varphi p_0)\psi_e(e|\varphi)\,de\psi_\varphi(\varphi)\,d\varphi + o(\tau) \\
= & -\frac{\tau\alpha S(p_0)}{p_0}\left[\begin{array}{c} \int_\varphi\tilde{\psi}_\varphi(\varphi)\Psi_e(\widehat{e}_\varphi|\varphi)E\left(e^2|e<\widehat{e}_\varphi,\varphi\right)d\varphi \\ +\tilde{E}\left(e|e>\widehat{e}_\varphi\right)\int_\varphi\tilde{\psi}_\varphi(\varphi)\,d\varphi\left(1-\Psi_e(\widehat{e}_\varphi|\varphi)\right)E\left(e|e>\widehat{e}_\varphi,\varphi\right) \\ -\tilde{E}(e)\int_\varphi\tilde{\psi}_\varphi(\varphi)\,d\varphi E(e|\varphi) \end{array}\right] \\
& -\frac{\tau S(p_0)\tilde{\Psi}\left(e<\widehat{e}_\varphi\right)}{b''(0)} + o(\tau) \\
= & -\frac{\tau\alpha S(p_0)}{p_0}\left(\tilde{\Psi}\left(e<\widehat{e}_\varphi\right)\tilde{E}\left(e^2|e<\widehat{e}_\varphi\right)+\left(1-\Psi_e\left(e<\widehat{e}_\varphi\right)\right)\tilde{E}\left(e|e>\widehat{e}_\varphi\right)^2-\tilde{E}(e)^2\right) \\
& -\frac{\tau S(p_0)\tilde{\Psi}\left(e<\widehat{e}_\varphi\right)}{b''(0)} + o(\tau).
\end{aligned}
$$

We then directly get that the change in emissions is given by (54). Using that expression together with (59) and (60) in (57) directly implies that the welfare change is given by (53). □

Optimum We now analyze optimal certification with heterogeneous productivity. As in Section 2.3, the government can implement an output tax, an emission tax for certified firms and a certification tax. We allow the certification tax f_φ to vary by firm productivity φ but not the output tax.[4] We then obtain:

Proposition 6. *The optimal policy is characterized by an emission tax $\tau = v$, an output tax given by the social cost of emissions times the average emissions of uncertified firms, where uncertified firms are weighted by their supply slope,*

$$
t = v\frac{\int_\varphi\left(1-\Psi_e(\widehat{e}_\varphi|\varphi)\right)E\left[e|e>\widehat{e}_\varphi,\varphi\right]\varphi s'(\varphi(p-t))\psi_\varphi(\varphi)\,d\varphi}{\int_\varphi\left(1-\Psi_e(\widehat{e}_\varphi|\varphi)\right)\varphi s'(\varphi(p-t))\psi_\varphi(\varphi)\,d\varphi}, \tag{61}
$$

[4]If the output tax is also allowed to vary with the productivity φ, the social planner can solve for optimal certification independently for firms of different productivity levels.

and a certification tax

$$f_{\varphi} = (t - v\widehat{e}_{\varphi})\, s\left(\varphi\left(p - t\right)\right). \tag{62}$$

In the isoelastic case, the output tax is equal to the social cost of emissions times the quantity weighted average of uncertified firms emission rates

$$t = v\widetilde{E}\left[e|e > \widehat{e}_{\varphi}\right],$$

and the optimal certification threshold $\widehat{e}_{\varphi}$ is decreasing in productivity φ.

Proof. With a certification tax f_{φ}, the government can directly choose the certification threshold for firms of productivity φ: $\widehat{e}_{\varphi}$. Taking into account that firms maximize profits, and consumers maximize utility, we can write the social planner problem as

$$\begin{aligned}
& \max_{\widehat{e}_{\varphi},p,\tau,t} W \\
= \;& u\left(D\left(p\right)\right) - F\int_{\varphi} \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\psi_{\varphi}\left(\varphi\right)d\varphi \\
& -v\int_{\varphi}\left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_{\varphi}} \left(e - a\left(\tau\right)\right) s\left(\varphi\left(p - \tau\left(e - a\left(\tau\right)\right) - b\left(a\left(\tau\right)\right)\right)\right)\psi_e\left(e|\varphi\right)de \\ +\left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right)E\left[e|e > \widehat{e}_{\varphi},\varphi\right]s\left(\varphi\left(p - t\right)\right) \end{array}\right)\psi_{\varphi}\left(\varphi\right)d\varphi \\
& -\int_{\varphi}\left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_{\varphi}}\left(\begin{array}{c} \frac{1}{\varphi}c\left(s\left(\varphi\left(p - \tau\left(e - a\left(\tau\right)\right) - b\left(a\left(\tau\right)\right)\right)\right)\right) \\ +b\left(a\left(\tau\right)\right)s\left(\varphi\left(p - \tau\left(e - a\left(\tau\right)\right) - b\left(a\left(\tau\right)\right)\right)\right) \end{array}\right)\psi_e\left(e|\varphi\right)de \\ +\left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right)\frac{1}{\varphi}c\left(s\left(\varphi\left(p - t\right)\right)\right) \end{array}\right)\psi_{\varphi}\left(\varphi\right)d\varphi
\end{aligned}$$

subject to the market clearing condition which defines the price level

$$D\left(p\right) = \int_{\varphi}\left(\int_{\underline{e}}^{\widehat{e}_{\varphi}} s\left(\varphi\left(p - \tau\left(e - a\left(\tau\right)\right) - b\left(a\left(\tau\right)\right)\right)\right)\psi_e\left(e|\varphi\right)de + \left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right)s\left(\varphi\left(p - t\right)\right)\right)\psi_{\varphi}\left(\varphi\right)d\varphi. \tag{63}$$

We denote by λ, the Lagrange multiplier on (63). The first-order condition with respect to p leads to:

$$\begin{aligned}
& \left(p - \lambda\right)D'\left(p\right) \\
= \;& \int_{\varphi}\left(\begin{array}{c} \int_{\underline{e}}^{\widehat{e}_{\varphi}} \varphi\left(p + \left(v - \tau\right)\left(e - a\left(\tau\right)\right) - \lambda\right)s'\left(\varphi\left(p - \tau\left(e - a\left(\tau\right)\right) - b\left(a\left(\tau\right)\right)\right)\right)\psi_e\left(e|\varphi\right)de \\ +\left(1 - \Psi_e\left(\widehat{e}_{\varphi}|\varphi\right)\right)\varphi\left(p + vE\left[e|e > \widehat{e}_{\varphi},\varphi\right] - t - \lambda\right)s'\left(\varphi\left(p - t\right)\right) \end{array}\right)\psi_{\varphi}\left(\varphi\right)d\varphi.
\end{aligned} \tag{64}$$

The first order condition with respect to τ gives:

$$\begin{aligned}
& (v-\tau)\,a'(\tau)\int_{\varphi}\left(\int_{\underline{e}}^{\widehat{e}_{\varphi}} s\left(\varphi\left(p-\tau\left(e-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)\psi_{e}\left(e|\varphi\right)de\right)\psi_{\varphi}\left(\varphi\right)d\varphi \\
& +\int_{\varphi}\left(\int_{\underline{e}}^{\widehat{e}_{\varphi}}\left[\left(p-\lambda+\left(v-\tau\right)\left(e-a\left(\tau\right)\right)\right)\right]\left(e-a\left(\tau\right)\right)\varphi s'\left(\varphi\left(p-\tau\left(e-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)\psi_{e}\left(e|\varphi\right)de\right) \\
& \times\psi_{\varphi}\left(\varphi\right)d\varphi \\
= \; & 0.
\end{aligned} \tag{65}$$

The first order condition with respect to t gives:

$$\int_{\varphi}\left(\left(1-\Psi_{e}\left(\widehat{e}_{\varphi}|\varphi\right)\right)\left[p-t-\lambda+vE\left[e|e>\widehat{e}_{\varphi},\varphi\right]\right]\varphi s'\left(\varphi\left(p-t\right)\right)\right)\psi_{\varphi}\left(\varphi\right)d\varphi=0. \tag{66}$$

Combining these equations together we obtain $\lambda=p$, $\tau=v$, and (61).

For simplicity, we focus on an interior solution for $\widehat{e}_{\varphi}$. Taking the first order condition with respect to $\widehat{e}_{\varphi}$, we get:

$$\begin{aligned}
& v\left[-\left(\widehat{e}_{\varphi}-a\left(\tau\right)\right)s\left(\varphi\left(p-\tau\left(\widehat{e}_{\varphi}-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)+\widehat{e}_{\varphi}s\left(\varphi\left(p-t\right)\right)\right] \\
& -\left(\frac{1}{\varphi}c\left(s\left(\varphi\left(p-\tau\left(\widehat{e}_{\varphi}-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)\right)+b\left(a\left(\tau\right)\right)s\left(\varphi\left(p-\tau\left(\widehat{e}_{\varphi}-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)\right) \\
& +\frac{1}{\varphi}c\left(s\left(\varphi\left(p-t\right)\right)\right)+\lambda\left[s\left(\varphi\left(p-\tau\left(\widehat{e}_{\varphi}-a\left(\tau\right)\right)-b\left(a\left(\tau\right)\right)\right)\right)-\psi_{e}\left(\widehat{e}_{\varphi}|\varphi\right)s\left(\varphi\left(p-t\right)\right)\right] \\
= \; & F.
\end{aligned}$$

Using that $\lambda=p$, $\tau=v$ and rearranging terms, we obtain:

$$\frac{1}{\varphi}\pi\left(\varphi\left(p-v\widehat{e}_{\varphi}+A\left(v\right)\right)\right)-\frac{1}{\varphi}\pi\left(\varphi\left(p-t\right)\right)=F+f_{\varphi},$$

with f_{φ} defined in (62).

We define

$$h\left(e,\varphi\right)\equiv\frac{1}{\varphi}\pi\left(\varphi\left(p-ve+A\left(v\right)\right)\right)-\frac{1}{\varphi}\pi\left(\varphi\left(p-t\right)\right)-\left(t-ve\right)s\left(\varphi\left(p-t\right)\right),$$

so that the threshold $\widehat{e}_{\varphi}$ is defined by $h\left(\widehat{e}_{\varphi},\varphi\right)=F$. Note that

$$\frac{\partial h}{\partial e}=-v\left(s\left(\varphi\left(p-ve+A\left(v\right)\right)\right)-s\left(\varphi\left(p-t\right)\right)\right)<0 \text{ and}$$

$$\frac{\partial h}{\partial \varphi} = \frac{1}{\varphi^2} c\left(s\left(\varphi\left(p - ve + A\left(v\right)\right)\right)\right) - \frac{1}{\varphi^2} c\left(s\left(\varphi\left(p - t\right)\right)\right) - \left(p - t\right)\left(t - ve\right) s'\left(\varphi\left(p - t\right)\right).$$

In the isoelastic case $s(p) = s_0 p^\alpha$ (so that $c(q) = \frac{\alpha}{1+\alpha} q^{\frac{1+\alpha}{\alpha}} s_0^{-\frac{1}{\alpha}}$ given that $c(0) = 0$), then

$$\begin{aligned}
\frac{\partial h}{\partial \varphi} &= s_0 \varphi^{\alpha-1} \alpha \left[\frac{1}{1+\alpha}\left(p - ve + A\left(\tau\right)\right)^{1+\alpha} - \frac{1}{1+\alpha}\left(p - t\right)^{1+\alpha} - \left(t - ve\right)\left(p - t\right)^\alpha\right] \\
&> s_0 \varphi^{\alpha-1} \alpha \left[\left(t - ve + A\left(\tau\right)\right)\left(p - t\right)^\alpha - \left(t - ve\right)\left(p - t\right)^\alpha\right] \\
&\geq s_0 \varphi^{\alpha-1} \alpha \left[A\left(\tau\right)\left(p - t\right)^\alpha\right] \geq 0
\end{aligned}$$

where the first inequality uses that $p^{\alpha+1}$ is convex. As a result, $\widehat{e}_\varphi$ is decreasing in φ when it is interior. □

Therefore, our results on the optimal constrained policy generalize to the case with heterogeneous productivity when supply is isoelastic. Naturally, the government sets a higher certification threshold for less productive firms as the benefits from paying the social cost of certification are lower for smaller producers.[5]

The assumption that the government can impose a productivity-specific certification tax f_φ but not a productivity-specific output tax is natural in the context where the government can only observe aggregate emissions and the emissions of certified firms. In that case, the government can observe the quantity-weighted average emissions of uncertified firms $\widetilde{E}\left[e | e > \widehat{e}_\varphi\right]$ by substracting revealed emissions from the aggregate and set $t = v\widetilde{E}\left[e | e > \widehat{e}_\varphi\right]$, but the government could not observe the average emission rates of uncertified firms conditional on their productivity $E\left[e | e > \widehat{e}_\varphi, \varphi\right]$. If the government can observe sales of uncertified firms, it can observe productivity φ and set the optimal f_φ as in (62). This ensures that the optimal threshold $\widehat{e}_\varphi$ gets implemented.

The unraveling algorithm described in Section 2.4 can directly be generalized to that case (with an isoelastic supply curve): The government can set an output tax based on the quantity-weighted average emission rate of uncertified firms in the previous step $t_n = v\widetilde{E}\left[e | e > \widehat{e}_{\varphi,n-1}\right]$ and productivity (or sales)-specific certification taxes
$f_{\varphi,n} = v\left(\widetilde{E}\left[e | e > \widehat{e}_{\varphi,n-1}\right] - v\widehat{e}_{\varphi,n-1}\right) s\left(\varphi\left(p - v\widetilde{E}\left[e | e > \widehat{e}_{\varphi,n-1}\right]\right)\right)$.

[5]Even if supply is not isoelastic, one can show that $\widehat{e}_\varphi$ is decreasing in φ when $c(s(p))$ is convex in p (provided that the government implements the optimal t given by (61)).

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 entering firms produce, and we denote by N the set of 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 $\pi\left(p-\tau e+A\left(\tau\right)\right)-F \geq \pi(p-t)$.[6] In addition, we assume that the certification costs F are at most second order in τ (otherwise certification leads to welfare losses). As a result, the constraint $e \leq \tilde{e}$ binds and $\hat{e}=\tilde{e}$ provided that $\tilde{e}$ is not too close to $E\left(e|e \geq \tilde{e}\right)$.[7]

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

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

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

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

We then obtain the following corollary:

Corollary 3. *The change in emissions brought about by certification can be written as:*

$$G^V - G^U = -N_0 s'\left(p_0\right)\tau Var\left(\varepsilon\right) - N_0\frac{\tau s\left(p_0\right)\Psi\left(\widehat{e}\right)}{b''(0)} + o(\tau), \tag{69}$$

and the change in welfare as:

$$W^V - W^U = \tau\left(v-\frac{\tau}{2}\right)N_0\left(s'\left(p_0\right)Var\left(\varepsilon\right)+\frac{s\left(p_0\right)\Psi\left(\widehat{e}\right)}{b''(0)}\right) - N_0 F\Psi\left(\widehat{e}\right) + o\left(\tau^2\right), \tag{70}$$

N_0 denotes the mass of firms in laissez-faire. These expressions are identical to those without free-entry given in Corollary 1, so that our initial results are robust to considering

[6] 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.

[7] Should that occur, we get nearly full certification and $\Psi\left(\hat{e}\right)=\Psi\left(\tilde{e}\right)+O\left(\tau\right)=1+O(\tau)$. The approximations in Corollary 1 remain valid with full revelation. This is the case in particular when there is no constraint $\tilde{e}$ (i.e. $\tilde{e}=\bar{e}$).

free-entry. One important difference, however, is that 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 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:

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

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 certification does not change the number of firms at first order.

A first order Taylor expansion of (72), using (17), (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 $\left(p^V - p_0\right)_2$ denotes the second order term in the price difference $p^V - p_0$. We then take a second order expansion of (67) and get

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

from which we obtain:

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

and

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

Using (73), (75), (17) and that $D(p_0) = Ns(p_0)$, we then derive

$$\int_{p^U}^{p^V} (-D(p))\, dp = \frac{1}{2} N_0 s'(p_0) \tau^2 Var(\varepsilon) + N_0 \Psi(\widehat{e}) \frac{\tau^2}{2b''(0)} - N_0 F \Psi(\widehat{e}) + o\left(\tau^2\right).$$

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

A.6 A Model of Lobbying Against Output and Emissions Taxes

In this section we present a lobbying model to explore why our gradual "unraveling" approach may be more easily implementable than a mandatory emissions fee. Suppose firms lobby for or against policy changes. Lobbying expenditure against a policy change is given by a function $l_a(\Delta\pi)$ where $\Delta\pi \leq 0$ is the change in profits relative to the status quo. Similarly, lobbying expenditure in favor of a policy change is given by a function $l_f(\Delta\pi)$ with $\Delta\pi \geq 0$. We assume that lobbying expenditures are convex and small for small changes in profits with $l_a(0) = l_f(0) = 0$, $l_a'(0) = l_f'(0) = 0$ and $l_a''(0)$, $l_f''(0) \geq 0$. Legislation is more likely to be blocked when aggregate expenditure against the reform is higher. The economy is initially in laissez-faire. We compare aggregate lobbying effort for/against output versus Pigouvian taxation, as well as under increasing levels of voluntary certification (for which an output tax is the first step). When unraveling, we focus on a situation where the threshold increases with $\widehat{e}_{n+1} \geq E(e|e > \widehat{e}_n)$. Under the certification mechanism, certified firms pay the emission tax v per emission and uncertified firms pay the output tax $vE(e|e > \widehat{e})$.[8] All firms pay an emission tax v under the emission tax. We consider firms' incentives to lobby in the stage game, and discuss forward-looking behavior that anticipates further unraveling at the end.

Consider first the passage from laissez-faire to either an output tax or an emission tax. All firms lose, but the dirtiest firms lose less with an output tax and the cleanest firms lose less with an emission tax. With convex lobbying expenditures, aggregate lobbying is lower against an output tax than against an emission tax. Consider next the introduction of voluntary certification up to some level $\widehat{e}_1$. Pooling among high-emissions firms with an output tax is more palatable than a move to a full emissions tax: again because opposition to an emissions tax among the highest uncertified emitters more than offsets the lower lobbying against an emission tax by the cleaner uncertified firms. In addition, with the status quo now being an output tax, the voluntary certification program has the additional appeal of lowering the tax burden for relatively low-emissions firms, who now lobby in its favor. In fact, every time the certification threshold is raised, the set of firms opposed to the previous reform is now divided between some firms that favor further certification (the relatively clean uncertified firms) and a dwindling set of firms opposed to an increase in certification. This "divide and rule" tactic is in the spirit of the "gradualist" approach of Dewatripont and Roland (1992, 1995). When the final equilibrium is complete unraveling, the mechanism encounters less opposition ex ante by delaying emission taxation for the highest emitters, while reducing the tax burden up-front relative to an output tax for low-emitters.

Formally, we show that lobbying expenditures are always lower under gradual unraveling than when mandating the emission tax. The difference is directly proportional to the variance of unrevealed emission rates:[9]

[8] To focus on the comparision of mandatory emissions taxes with complete unraveling, we assume zero certification costs and fees. We do not impose additional constraints on $\widehat{e}_n$, which can be any value chosen by the government—potentially having in mind ensuring that the policy proposal is accepted. Alternatively, we could consider the implementation of the algorithm of section 2.4. In that case uncertified firms pay a lower (not yet updated) output tax, which further reduces lobbying expenditures against gradual certification compared to the emission tax.

[9] When $l'_a(0) \neq 0$, aggregate lobbying expenditures may be smaller under the output tax than under an emission tax. On one hand, compared to an output tax, an emission tax reallocates profits from high- to low-emitters by a first-order amount, this reallocation interacts with the convexity of the lobbying function, increasing lobbying expenditures by a second-order term. On the other hand, an emission tax brings efficiency

Proposition 7. *At each step of the unraveling process, aggregate opposition lobbying expenditures are lower than when the social planner directly imposes the optimal carbon tax (at second order). The difference in aggregate lobbying expenditures is proportional to* $(1 - \Psi(\widehat{e})) Var(e|e > \widehat{e})$ *where* $\widehat{e}$ *is the threshold of certified firms (*$= \underline{e}$ *in the case of the output tax).*

Proof. Implementing an output tax. Under an output tax at rate $vE(e)$, the change in profits relative to the status-quo (laissez-faire) is given by

$$\Delta\pi = \pi(p - vE(e)) - \pi(p_0) = s(p_0)(p - p_0 - vE(e)) + o(v),$$

where p is the prevailing price under an output tax, and p_0 is the price under laissez-faire. Noting that the difference in prices between output and emissions taxes are of higher order, we suppress policy-specific price superscripts for notational simplicity. We denote aggregate lobbying expenditures when moving from policy X to Y, $L^{X,Y}$ for policies $\{LF, U, V, F\}$ for laissez-faire, output tax, voluntary certification, and mandatory emission tax, respectively. Going from laissez-faire to an output tax, all firms lose by the same amount and spend

$$L^{LF,U} = l_a(\Delta\pi) = \frac{1}{2} l_a''(0) s(p_0)^2 (p - p_0 - vE(e))^2 + o(v^2),$$

on lobbying.

With a mandatory Pigouvian tax, the change in profits relative to the status-quo obeys:

$$\Delta\pi = \pi(p - ve) - \pi(p_0) = s(p_0)(p - p_0 - ve) + o(v).$$

We can then write aggregate lobbying expenditures as

$$L^{LF,F} = \frac{1}{2} l_a''(0) s(p_0)^2 E(p - p_0 - ve)^2 + o(v^2)$$

The difference in lobbying expenditures between an emissions and output tax is given by

$$L^{LF,F} - L^{LF,U} = \frac{1}{2} l_a''(0) s(p_0)^2 v^2 Var(e) + o(v^2),$$

gains relative to an output tax which raise the producer surplus by a second order term, this efficiency gains affect the amount of lobbying proportionately to $l'(0)$. With $l_a = l_f = l$ and $l'(0) \neq 0$, we can derive that aggregate lobbying expenditures are higher under the emission tax if $l''(0)/l'(0) > s'(p)/s(p)$. Given the strong opposition to emission taxes, we focus on the case where lobbying expenditures are higher for an emission tax and, to simplify the problem, assume that $l_a'(0) = l_f'(0) = 0$.

which is positive. A direct move to the optimal emission tax is more likely to be blocked than a move to an output tax.

Step-by-step certification. We now compare a move from a certification equilibrium (with threshold $\widehat{e}_n$) to either a higher certification threshold $\widehat{e}_{n+1} \geq E\left(e|e > \widehat{e}_n\right)$ or the optimal emission tax. This includes a move from the output tax which is equivalent to a certification threshold $\widehat{e} = \underline{e}$.

Since prices do not change with the certification level at first order, firms with $e \leq \widehat{e}_n$ are indifferent between the status quo and mandatory emissions taxes (at first order). Without certification costs, all firms with $e \leq \widehat{e}_{n+1}$ choose to certify. The gain in profits for newly certifying firms is given by

$$\Delta\pi = \pi\left(p - ve\right) - \pi\left(p - vE\left(e|e > \widehat{e}_n\right)\right) = s\left(p_0\right)v\left(E\left(e|e > \widehat{e}_n\right) - e\right) + o\left(v\right).$$

Uncertified firms will lose from the reform with a profit loss given by

$$\Delta\pi = \pi\left(p - vE\left(e|e > \widehat{e}_{n+1}\right)\right) - \pi\left(p - vE\left(e|e > \widehat{e}_n\right)\right) = s\left(p_0\right)v\left(E\left(e|e > \widehat{e}_n\right) - E\left(e|e > \widehat{e}_{n+1}\right)\right) + o\left(v\right).$$

We can then write aggregate lobbying expenditures going from $\widehat{e}_n$ to $\widehat{e}_{n+1}$ as:

$$\begin{aligned}
& L^{V,V}\left(\widehat{e}_n, \widehat{e}_{n+1}\right) \\
= \; & \left(1 - \Psi\left(\widehat{e}_{n+1}\right)\right) l_a\left(s\left(p_0\right)v\left(E\left(e|e > \widehat{e}_n\right) - E\left(e|e > \widehat{e}_{n+1}\right)\right) + o\left(v\right)\right) \\
& + \int_{E(e|e>\widehat{e}_n)}^{\widehat{e}_{n+1}} l_a\left(s\left(p_0\right)v\left(E\left(e|e > \widehat{e}_n\right) - e\right) + o\left(v\right)\right)\psi\left(e\right)de \\
& - \int_{\widehat{e}_n}^{E(e|e>\widehat{e}_n)} l_f\left(s\left(p_0\right)v\left(E\left(e|e > \widehat{e}_n\right) - e\right) + o\left(v\right)\right)\psi\left(e\right)de \\
= \; & \frac{\left(s\left(p_0\right)v\right)^2}{2} \\
& \times \left[\begin{array}{c}
\left(1 - \Psi\left(\widehat{e}_{n+1}\right)\right)l_a''\left(0\right)\left(E\left(e|e > \widehat{e}_n\right) - E\left(e|e > \widehat{e}_{n+1}\right)\right)^2 \\
+\left(\Psi\left(\widehat{e}_{n+1}\right) - \Psi\left(E\left(e|e > \widehat{e}_n\right)\right)\right)l_a''\left(0\right)E\left(\left(E\left(e|e > \widehat{e}_n\right) - e\right)^2|E\left(e|e > \widehat{e}_n\right) < e < \widehat{e}_{n+1}\right) \\
-\left(\Psi\left(E\left(e|e > \widehat{e}_n\right)\right) - \Psi\left(\widehat{e}_n\right)\right)l_f''\left(0\right)E\left(\left(E\left(e|e > \widehat{e}_n\right) - e\right)^2|e < E\left(e|e > \widehat{e}_n\right)\right)
\end{array}\right].
\end{aligned}$$

Moving directly to the emission tax is equivalent to updating to $\widehat{e}_{n+1} = \bar{e}$ in the previous expression (when the status quo certification level is$\hat{e}_n$). The same set of firms will lobby for or against the reform, but the the most polluting firms will lobby more actively against a switch to the emission tax. We can write aggregate lobbying expenditures as:

$$L^{V,F}\left(\widehat{e}_n\right) = \frac{\left(s\left(p_0\right)v\right)^2}{2}\left[\begin{array}{c} \left(1-\Psi\left(E\left(e|e>\widehat{e}_n\right)\right)\right)l_a''\left(0\right)E\left(\left(E\left(e|e>\widehat{e}_n\right)-e\right)^2|e>E\left(e|e>\widehat{e}_n\right)\right) \\ -\left(\Psi\left(E\left(e|e>\widehat{e}_n\right)\right)-\Psi\left(\widehat{e}_n\right)\right)l_f''\left(0\right)E\left(\left(E\left(e|e>\widehat{e}_n\right)-e\right)^2|e<E\left(e|e>\widehat{e}_n\right)\right)\end{array}\right].$$

We can then write the difference in lobbying expenditures as:

$$L^{V,F}\left(\widehat{e}_n\right)-L^{V,V}\left(\widehat{e}_n,\widehat{e}_{n+1}\right)=\frac{\left(s\left(p_0\right)v\right)^2}{2}l_a''\left(0\right)\left(1-\Psi\left(\widehat{e}_{n+1}\right)\right)Var\left(e|e>\widehat{e}_{n+1}\right).$$

which is again positive. Thus voluntary certification encounters less lobbying resistance than mandatory Pigouvian taxation, both at the outset (i.e. for an output tax which corresponds to voluntary certification with no one opting-in), and when increasing the level of certification from any arbitrary level. □

We have thus far considered the feasiblity of individual changes from an evolving status quo. Forward-looking firms will anticipate these subsequent updates and lobby according to a reform's impact on the discounted flow of future profits as unraveling progresses. In such a setting, firms would correctly predict that voluntary certification can lead to the same final outcome as a Pigovian tax. However the profits of the most polluting firms are partly shielded along the transition via the output tax. The dirtiest firms are the last to certify, and their future losses from Pigouvian taxation are discounted by a combination of interest rate and the probability of firm death. Thus the greatest discounting is applied to the largest losses, reducing the opposition from those who would otherwise lobby the most against an emissions tax. A policy of unraveling therefore encounters less ex ante opposition than a mandatory Pigouvian tax from fully-informed, forward-looking firms who anticipate eventually being subject to an emissions tax.

B Application to the Permian Basin

We now present in detail the drilling model of Section 3 which studies the introduction of our certification mechanism in the context of methane leaks in the exploitation of shale gas and oil. The main difference with the baseline analysis is that reallocation no longer occurs on the intensive margin but on the extensive margin as producers need to decide whether or not to drill. Section B.1 derives output, emissions, and welfare in equilibrium for any value of τ. Section B.2 derives the price elasticity of output and emissions in laissez-faire. Section B.3 derives Taylor approximation for output, emissions, and welfare when τ is small. Section B.4 gives details on our data. Section B.5 explains how we calibrate our model and compute the formulas from our approximations with the data at hand. Section B.6 derives expressions for the model with uniform costs of drilling and Section B.7 considers the change in welfare along the algorithm.

B.1 Equilibrium

As described in Section 3.1, a firm is characterized by a vector $(u, q, p, \mathfrak{e}, \varkappa)$, which is distributed according to the cdf $\Psi(u, q, p, \mathfrak{e}, \varkappa)$ with a corresponding pdf ψ. The domain is $[\underline{u}, \bar{u}] \times [\underline{q}, \bar{q}] \times [\underline{p}, \bar{p}] \times [\underline{\mathfrak{e}}, \bar{\mathfrak{e}}] \times [\underline{\varkappa}, \bar{\varkappa}]$ with weakly positive lower bounds and finite upper bounds, $1 \in (\underline{u}, \bar{u})$, and Ψ has full support and no mass point. We assume that $u, \varkappa, p$ are independent of each other and of $(q, \mathfrak{e})$. Therefore we can write

$$\Psi\left(u, q, p, \mathfrak{e}, \varkappa\right) = \Psi_u\left(u\right)\Psi_p\left(p\right)\Psi_\varkappa\left(\varkappa\right)\Psi_{q,\mathfrak{e}}\left(q, \mathfrak{e}\right),$$

where Ψ_u denotes the unconditional distribution of u, and $\Psi_\varkappa, \Psi_p$, and $\Psi_{q,\mathfrak{e}}$ are similarly defined. We introduce certification as follows: certified firms pay a tax τ on their emissions and uncertified firms pay an output tax t, which is equal to $\tau\times$the (quantity-weighted) average emissions rate of uncertified firms. Instead of imposing a certification tax of f, we permit firms to certify to the point of $\tilde{e}$.[1]

We now solve for the equilibrium for any level of certification limit $\tilde{e}$ and derive expressions

[1] 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. In the extensive margin case, a certification tax would directly affect both the margins of certification and of entry so that these two setups will not be equivalent. To simplify the analysis, we assume that the government directly imposes an upper bound on certification $\tilde{e}$ (this is the same issue as in the free-entry extension, see Section A.5.3). As described below, for relevant parameters, this amounts to the social planner directly choosing a certification threshold $\hat{e}$. We introduce $\tilde{e}$ to be able to show welfare gains for intermediate levels of certifications as in the curved line in Figure 2. Alternatively, we obtain intermediate level of certifications without such a bound in the algorithm, as shown in the step line in Figure 2 (see Section B.7 below).

that allow us to compute welfare.

Firm level production and emissions. Consider a well that has already been drilled. A firm with $e \leq \tilde{e}$ has the option to certify. If it does so, it pays τ per unit of emission and (conditional on producing) has an incentive to abate emissions; otherwise it faces an output tax t. The abatement rate maximizes the net gains from abatement per unit of output: $A(\tau, e) \equiv \min_a \tau ae - b(a)$, that is, a firm abates $a(e) = b'^{-1}(\tau e)$, which is positive since $b'(0) = 0$ (exactly as in Appendix A.5.1). Given the certification cost F, such a firm chooses to certify whenever

$$(p - \tau e(1-a) - b(a))\, q - F \geq (p-t)q \Longleftrightarrow t - \tau e + A(\tau, e) \geq \frac{F}{q}. \tag{1}$$

This defines a threshold function $\hat{e}(q)$: such that $\hat{e}(q) = \underline{e}$ if this inequality is violated for $\underline{e}$, $\hat{e}(q)$ is the solution to $t - \tau e(1-a) - b(a) = \frac{F}{q}$ if that solution is in $[\underline{e}, \tilde{e}]$, and $\hat{e}(q) = \tilde{e}$ otherwise (i.e. (1) holds with a strict inequality at $\tilde{e}$). Then, firms with $e \leq \hat{e}(q)$ certify and those with $e > \hat{e}(q)$ do not.[2] Uncertified firms (i.e. those with $e > \hat{e}(q)$) produce q and emit eq if and only if $p \geq t$ (otherwise the firm shuts down). Certified firms (i.e. those with $e \leq \hat{e}(q)$) produce q and emit $e(1-a)q$, if and only if $p - \tau e + A(\tau, e) \geq \frac{F}{q}$ (otherwise they shut down).

Drilling decision. For a given emission tax τ, we denote by $h(q, p, \mathfrak{e}, \tau)$ a firm's expected net revenues per unit from drilling prior to learning $\varkappa$, and therefore its full emission rate e. We get:

$$h(q, p, \mathfrak{e}, \tau) \equiv \int_{\varkappa} \left[\mathbf{1}_{\mathfrak{e}\varkappa \leq \hat{e}} \left(p - \tau \mathfrak{e}\varkappa + A(\tau, \mathfrak{e}\varkappa) - \frac{F}{q} \right)^{+} + \mathbf{1}_{\mathfrak{e}\varkappa > \hat{e}} (p-t)^{+} \right] \psi_{\varkappa}(\varkappa)\, d\varkappa, \tag{2}$$

where we use the notation $x^{+} \equiv \max\{x, 0\}$ and we recall that A is for optimally chosen abatement. At the time of drilling the firm does not yet know its emission shock $\varkappa$, it then needs to form expectations. If its eventual rate is below $\widehat{e}(q)$, the firm will certify, earning $p - \tau \mathfrak{e}\varkappa + A(\tau, \mathfrak{e}\varkappa) - \frac{F}{q}$ per unit (taking into account the fixed cost of certification), while if it is above $\widehat{e}(q)$, the firm will not certify and earn $p - t$ per unit. The firm drills if and only if the expected revenue per unit is larger than the drilling cost per unit, that is if and only if $h(q, p, \mathfrak{e}, \tau) \geq pu$.

[2] Technically, if (1) is violated for $\underline{e}$, firms with emission rates $\hat{e}(q) = \underline{e}$ do not certify, but whether we count them as certifying or not in the following is immaterial since they are a measure 0 of firms.

Aggregate production, emissions and output tax. To derive aggregate supply, we then simply need to sum across all firms' types. Firms only enter if $h(q,p,\mathfrak{e},\tau) \geq pu$ and they produce q unless they exit because ex-post production turns out to be non-profitable. Therefore, we can write:

$$S(\tau,\tilde{e}) = \int_{u,q,p,\mathfrak{e}} q \left[\mathbf{1}_{h(q,p,\mathfrak{e},\tau)\geq pu} \int_{\varkappa} \begin{pmatrix} \mathbf{1}_{\mathfrak{e}\varkappa\leq\hat{e}} \mathbf{1}_{p-\tau\mathfrak{e}\varkappa+A(\tau,\mathfrak{e}\varkappa)-\frac{F}{q}\geq 0} \\ +\mathbf{1}_{\mathfrak{e}\varkappa>\hat{e}} \mathbf{1}_{p-t\geq 0} \end{pmatrix} \psi_{\varkappa}(\varkappa)\, d\varkappa \right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}), \tag{3}$$

where we make explicit that S depends on τ and the certification threshold $\tilde{e}$. The laissez-faire equilibrium corresponds to the case $\tau = 0$ (and $\tilde{e} = \underline{e}$), and we can write

$$\begin{aligned} S^{LF} &= S(0,\underline{e}) = \int_{u,q,p,\mathfrak{e}} q \mathbf{1}_{p\geq pu} d\Psi_{u,q,p,\mathfrak{e}}(u,q,\mathfrak{e},p) \\ &= \Psi_u(1) \int_{q,\mathfrak{e}} q d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) = \Psi_u(1) E_{\psi_{\mathfrak{e},q}}[q], \end{aligned} \tag{4}$$

where we used that u, p, and $(q,\mathfrak{e})$ are independent of each other and where $E_{\psi_{\mathfrak{e},q}}[.]$ is the expectation over $(q,\mathfrak{e})$ distributed with pdf $\psi_{\mathfrak{e},q}$. Total supply without taxes therefore equals the fraction of firms which draw a low enough cost to enter, $\Psi_u(1)$, times expected production per firm.

Similarly, we can derive aggregate emissions as

$$G(\tau,\tilde{e}) = \tag{5}$$

$$\int_{u,q,p,\mathfrak{e}} \mathfrak{e}q \left[\mathbf{1}_{h(q,p,\mathfrak{e},\tau)\geq pu} \int_{\varkappa} \begin{pmatrix} \mathbf{1}_{\mathfrak{e}\varkappa\leq\hat{e}} \mathbf{1}_{p-\tau\mathfrak{e}\varkappa+A(\tau,\mathfrak{e}\varkappa)-\frac{F}{q}\geq 0} (1-a) \\ +\mathbf{1}_{\mathfrak{e}\varkappa>\hat{e}} \mathbf{1}_{p-t\geq 0} \end{pmatrix} \varkappa\psi_{\varkappa}(\varkappa)\, d\varkappa \right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

In laissez-faire, we get

$$\begin{aligned} G^{LF} &= G(0,\underline{e}) = \int_{u,q,p,\mathfrak{e}} \mathfrak{e}q \mathbf{1}_{p\geq pu} d\Psi_{u,q,p,\mathfrak{e}}(u,q,\mathfrak{e},p) \\ &= \Psi_u(1) \left(\int_{q,\mathfrak{e}} \mathfrak{e}q d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \right) = \Psi_u(1) E_{\psi_{\mathfrak{e},q}}[\mathfrak{e}q], \end{aligned} \tag{6}$$

which is the fraction of firms who enter, $\Psi_u(1)$ times expected total emissions of firms. Here we use that the emission shock $\varkappa$ has a mean of 1 and is iid. For that reason, we can take expectations over either the perceived or the true emission rate: $E_{\psi_{\mathfrak{e},q}}[\mathfrak{e}q] = E_{\psi_{e,q}}[eq]$, where $E_{\psi_{e,q}}$ denotes the expectation over e and q distributed according to the joint pdf $\psi_{e,q}$. We define $\widetilde{\psi}(e) = \frac{\int_q q\psi_{q,e}(q,e)dq}{\int_{q,e} qd\Psi_{q,e}(q,e)}$ as the quantity-weighed distribution of emission rates and we naturally get:

$$G^{LF} = \frac{E_{\psi_{e,q}}[eq]}{E_{\psi_{e,q}}[q]} S^{LF} = E_{\widetilde{\psi}_e}[e] S^{LF}. \tag{7}$$

The output tax paid by uncertified firms is equal to the emission tax τ times the quantity-weighted average emission rate of uncertified firms, that is:

$$t(\tau, \tilde{e}) = \tau \frac{\int_{u,q,p,\mathfrak{e}} \mathfrak{e} q \left[\mathbf{1}_{h(p,q,\mathfrak{e},\tau)\geq pu} \int_\varkappa \mathbf{1}_{\mathfrak{e}\varkappa>\hat{e}} \mathbf{1}_{p-t\geq 0} \varkappa \psi_\varkappa(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})}{\int_{u,q,p,\mathfrak{e}} q \left[\mathbf{1}_{h(p,q,\tau,\mathfrak{e})\geq pu} \int_\varkappa \mathbf{1}_{\mathfrak{e}\varkappa>\hat{e}} \mathbf{1}_{p-t\geq 0} \psi_\varkappa(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})}. \tag{8}$$

The output tax with no certification can be derived by setting $\hat{e} = \underline{e}$ and it is given by:

$$t^U = \tau E_{\widetilde{\psi}_e}[e]. \tag{9}$$

Welfare. Similarly to the intensive case, we can write welfare as the sum of exogenous labor income I, consumer surplus CS, producer surplus (net of certification costs) PS, emissions externality $-vG$, and tax revenues T:

$$W = I + CS + PS - vG + T. \tag{10}$$

With exogenous consumer price, the consumer surplus CS is constant (and independent of the policy considered). Since h denotes the expected profit per unit of output gross of drilling costs, we can write the producer surplus as

$$PS(\tau, \tilde{e}) = \int_{u,q,p,\mathfrak{e}} q\left(h(q,p,\mathfrak{e},\tau) - pu\right)^+ d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}). \tag{11}$$

As certified firms pay $(1-a)\tau\mathfrak{e}\varkappa$ per unit and certified firms pay t per unit in taxes, we get that tax revenues are given by:

$$T(\tau, \tilde{e}) = \int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h(p,q,\mathfrak{e},\tau)\geq pu} \int_\varkappa \begin{pmatrix} \mathbf{1}_{\mathfrak{e}\varkappa\leq\hat{e}} \mathbf{1}_{p-\tau\mathfrak{e}\varkappa+A(\tau,\mathfrak{e}\varkappa)-\frac{F}{q}\geq 0} (1-a)\tau e \\ +\mathbf{1}_{\mathfrak{e}\varkappa>\hat{e}} \mathbf{1}_{p-t>0} t \end{pmatrix} \psi_\varkappa(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

Given the definition of t, it is direct to check that we still have:

$$T(\tau, \tilde{e}) = \tau G(\tau, \tilde{e}).$$

We can the rewrite welfare as

$$W = I + CS + PS - (v - \tau) G. \tag{12}$$

B.2 Price supply and emission elasticities

In Section B.3 below, we use the derivative of the supply and emission functions with respect to prices, and we derive these expressions here. We introduce a shock ρ to the price of both oil and gas. Using (4) and (6), we can write that laissez-faire output and emissions are given by

$$S^{LF}(\rho) = \int_{u,q,p,\mathfrak{e}} q\mathbf{1}_{p+\rho \geq pu} d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \text{ and } G^{LF}(\rho) = \int_{u,q,p,\mathfrak{e}} \mathfrak{e} q\mathbf{1}_{p+\rho \geq pu} d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}). \tag{13}$$

These expressions rely on the indicator function $\mathbf{1}_{p+\rho \geq pu}$, which is not differentiable at $p+\rho = pu$.

A short review of Dirac Functions To carry out Taylor approximations, we therefore employ the Dirac Delta Function and briefly review 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 $\Xi(x, y)$ with pdf of $\xi(x, y)$ and $\xi_x(x)$ and $\xi_y(y|x)$, we consider the function:

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

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

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

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

$$\int_x \int_y x\delta_{ax=y} f(ax,x,a) d\Xi(x,y) = \int_x x f(ax,x,a)\xi_y(ax|x)\xi_x(x)dx,$$

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.

The slopes of the supply and emission functions The derivative of aggregate supply in laissez-faire with respect to a common (additive) price shock (evaluated at 0), is given by

$$\dot{S}^{LF} = E_{\psi_p}\left[\frac{1}{p}\right] E_{\psi_q}[q]\,\psi_u(1)\,, \tag{14}$$

where we used that u, p, and $(q, \mathfrak{e})$ are independent of each other. The expression captures the change in entry and therefore relies on the pdf $\psi_u(1)$. Similarly, we can express the derivative of aggregate emissions (evaluated at $\rho = 0$) as:

$$\dot{G}^{LF} = \int_{q,p,\mathfrak{e}} \frac{\mathfrak{e}q}{p}\psi_u(1)\, d\Psi_{q,\mathfrak{e},p}(q,\mathfrak{e},p) = E_{\psi_p}\left[\frac{1}{p}\right] E_{\psi_{\mathfrak{e},q}}[\mathfrak{e}q]\,\psi_u(1)\,.$$

Using (7), we then get that the ratio of the emission slope to the supply slope is equal to the ratio of emissions to output:

$$\frac{\dot{G}^{LF}}{\dot{S}^{LF}} = \frac{G^{LF}}{S^{LF}} = E_{\widetilde{\psi}_e}[e]\,.$$

This ensures that the optimal output tax is set to the social cost of methane times the quantity-weighted average emission rates.

In the data, we measure the elasticity of supply with respect to prices. We denote by r a common proportional shock to prices, so that:

$$S^{LF}(r) = \int_{u,q,p,\mathfrak{e}} q\mathbf{1}_{p(1+r)\geq pu} d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})\,. \tag{15}$$

We then get that the supply elasticity is given by

$$\epsilon^S = \frac{d\ln S^{LF}(r)}{dr}|_{r=0} = \frac{\psi_u(1)}{\Psi_u(1)}.$$

Therefore, we can infer the supply slope from the elasticity as

$$\dot{S}^{LF} = E_{\psi_p}\left[\frac{1}{p}\right] S^{LF}\epsilon^{S}. \tag{16}$$

B.3 Taylor approximations

We now assume that τ is small and derive Taylor approximations for output, emissions, and the different terms in the welfare expression. We also assume that F is second order in τ—that is we will write $F(\tau)$ and make Taylor approximation around $\tau = 0$ with $F'(0) = 0$.[3]

First, we note that

$$\frac{\partial a}{\partial \tau}|_{\tau=0} = \frac{e}{b''(0)}. \tag{17}$$

Therefore, as in the intensive margin case (Appendix A.5.1), abatement is first order in τ and the firm-level gains from abatements are second order:

$$a(e) = \frac{\tau e}{b''(0)} + o(\tau) \text{ and } A(\tau, e) = \frac{(\tau e)^2}{2b''(0)} + o(\tau^2).$$

Second, we note that for τ small, firms never exit after drilling, that is, we always have $p - t > 0$ and $p - \tau\mathfrak{e}\varkappa + A(\tau, e\varkappa) - \frac{F}{q} > 0$. In this subsection, we can then always assume that these two inequalities are satisfied. In particular, we can rewrite the output tax paid by uncertified firms as $t(\tau, \tilde{e}) = \tau e^{unc}(\tau, \tilde{e})$, where e^{unc} is the quantity-weighted average emissions rate of uncertified firms given by

$$e^{unc}(\tau, \tilde{e}) \equiv \frac{\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h(p,q,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}e\psi_{\varkappa}(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})}{\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h(p,q,\tau,\mathfrak{e})\geq pu}\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}\psi_{\varkappa}(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})}. \tag{18}$$

Using the notation e^{unc} in (1), we can now write that a firm with $e \leq \tilde{e}$ will certify if and only if the following inequality holds:

$$\tau e^{unc}(\tau, \tilde{e}) - \tau e + A(\tau, e) \geq \frac{F}{q}. \tag{19}$$

Third, we now establish that in our approximations below, we can treat $\hat{e}$ as an exogenous variable, independent of q, which is equal to $\tilde{e}$ if the latter is not close to $\overline{e}$ and to $\overline{e}$ otherwise.

[3] Of course, F is not literally a function of τ, but this notation enables us to take Taylor approximations in a rigorous way. This is simply to be interpreted as F not being too large relative to τ^2, which ensures that certification may be socially desirable.

Consider first the case where $\tilde{e}$ is not close to $\overline{e}$. Note that the right-hand side of (19), F/q is second order in τ. When $\tilde{e}$ is not close to $\overline{e}$, then e^{unc} is not close to $\tilde{e}$ either. As a result, for any $e \leq \tilde{e}$, the left-hand side in (19), namely $t - \tau e + A(\tau, e)$, is first order in τ. This ensures that the inequality (19) holds for all $e \leq \tilde{e}$: the constraint that firms can only certify up to $\tilde{e}$ binds and $\hat{e}(q) = \tilde{e}$. Consider next the case where $\tilde{e}$ is close to $\overline{e}$, then (19) holds for all e unless they are close to $\overline{e}$. Therefore, $\hat{e}(q)$ is close to $\overline{e}$ and nearly all firms certify.

Output tax. Using (18), we can differentiate $t(\tau)$ for τ close to 0:[4]

$$t'(\tau) = e^{unc}(\tau) + \tau \frac{de^{unc}}{d\tau}. \tag{20}$$

Using (2), we get $h(q, p, \mathfrak{e}, 0) = p$, so that

$$t'(0) = e^{unc}(0) = \frac{\Psi_u(1) \int_{,q,e} \mathbf{1}_{e>\widehat{e}} eq d\Psi_{q,e}(q,e)}{\Psi_u(1) \int_{,q,e} \mathbf{1}_{e>\widehat{e}} eq d\Psi_{q,e}(q,e)} = E_{\widetilde{\psi}_e}[e|e>\hat{e}].$$

Therefore, at first order we get

$$t(\tau) = \tau E_{\widetilde{\psi}_e}[e|e>\hat{e}] + o(\tau),$$

that is the output tax paid by uncertified firms is equal to $\tau\times$ the quantity-weighted average emissions rate of "potential" firms above the threshold. As in the baseline intensive-margin case, we define ε, the revealed pre-abatement emission rate from certification as:

$$\varepsilon \equiv e \text{ if } e \leq \widehat{e} \text{ and } \varepsilon \equiv E_{\widetilde{\psi}_e}[e|e>\hat{e}] \text{ if } e > \widehat{e}.$$

We note that

$$\begin{aligned} E_{\widetilde{\psi}_e}[\varepsilon] &= \int_{q,\mathfrak{e}} \frac{q}{E_{\psi_q}[q]} \int_{\varkappa} \left[\mathbf{1}_{\mathfrak{e}\varkappa \leq \widehat{e}} \mathfrak{e}\varkappa + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \frac{\int_{q,e} eq \mathbf{1}_{e>\widehat{e}} d\Psi_{q,e}(q,e)}{\int_{q,e} q \mathbf{1}_{e>\widehat{e}} d\Psi_{q,e}(q,e)} \right] \psi_{\varkappa}(\varkappa) d\varkappa d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \\ &= \int_{q,e} \left[\mathbf{1}_{e\leq\widehat{e}} e \frac{q}{E_{\psi_q}[q]} + \mathbf{1}_{e>\widehat{e}} \frac{q \int_{q,e} eq \mathbf{1}_{e>\widehat{e}} d\Psi_{q,e}(q,e)}{E_{\psi_q}[q] \int_{q,e} q \mathbf{1}_{e>\widehat{e}} d\Psi_{q,e}(q,e)} \right] d\Psi_{q,e}(q,e) \\ &= E_{\widetilde{\psi}_e}[e] = E_{\widetilde{\psi}_{\mathfrak{e}}}[\mathfrak{e}]. \end{aligned} \tag{21}$$

For some of the expressions below, we need a second order development for $t(\tau)$. We can

[4] We drop the dependence on $\tilde{e}$ since in what follows we always keep a fixed $\tilde{e}$.

write

$$t''(\tau) = 2\frac{de^{unc}}{d\tau} + \tau\frac{d^2 e^{unc}}{(d\tau)^2},$$

where

$$\frac{de^{unc}}{d\tau} = \frac{\int_{u,q,p,\mathfrak{e}} \mathfrak{e} q \left[\delta_{\frac{h(p,q,\mathfrak{e},\tau)}{p}=u} \frac{1}{p}\frac{\partial h}{\partial \tau} \int_{\varkappa} \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \varkappa \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})}{\int_{u,q,p,\mathfrak{e}} q \left[\mathbf{1}_{\frac{h(p,q,\mathfrak{e},\tau)}{p}\geq u} \int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})} \tag{22}$$

$$- \frac{\begin{pmatrix} \int_{u,q,p,\mathfrak{e}} q \left[\delta_{\frac{h(p,q,\mathfrak{e},\tau)}{p}=u} \frac{1}{p}\frac{\partial h}{\partial \tau} \int_{\varkappa} \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \varkappa \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \\ \times \int_{u,q,p,\mathfrak{e}} \mathfrak{e} q \left[\mathbf{1}_{\frac{h(p,q,\mathfrak{e},\tau)}{p}\geq u} \int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \end{pmatrix}}{\left(\int_{u,q,p,\mathfrak{e}} q \left[\mathbf{1}_{\frac{h(p,q,\mathfrak{e},\tau)}{p}\geq u} \int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})\right)^2}.$$

We note that

$$\frac{\partial h}{\partial \tau} = \int_{\varkappa} \left[\mathbf{1}_{\mathfrak{e}\varkappa \leq \widehat{e}} \left(-\mathfrak{e}\varkappa (1 - a(\tau, e)) - \frac{F'(\tau)}{q}\right) + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} (-t'(\tau))\right] \psi_{\varkappa}(\varkappa)\, d\varkappa. \tag{23}$$

Therefore,

$$\frac{\partial h}{\partial \tau}|_{\tau=0} = -\int_{\varkappa} \left[\mathbf{1}_{\mathfrak{e}\varkappa < \widehat{e}} \mathfrak{e}\varkappa + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} E_{\widetilde{\psi}_e}[e | e > \hat{e}]\right] \psi_{\varkappa}(\varkappa)\, d\varkappa = -m(\mathfrak{e}), \tag{24}$$

where we define $m(\mathfrak{e}) \equiv E_{\psi_{\varkappa}}[\varepsilon(\mathfrak{e})]$. This captures the expected pre-abatement emissions rate at which a firm with ex-ante expected rate $\mathfrak{e}$ is taxed. Note that $E_{\widetilde{\psi}_{\mathfrak{e}}}[m(\mathfrak{e})] = E_{\widetilde{\psi}_{\mathfrak{e}}}[E_{\psi_{\varkappa}}[\varepsilon(\mathfrak{e})]] = E_{\widetilde{\psi}_e}[\varepsilon] = E_{\widetilde{\psi}_e}[e]$. We can then evaluate (22) at $\tau = 0$ to get:

$$\frac{de^{unc}}{d\tau}|_{\tau=0} = -E_{\psi_p}\left[\frac{1}{p}\right] \frac{\psi_u(1)}{\Psi_u(1)} \left[\begin{array}{c} \frac{\int_{q,\mathfrak{e}} \mathfrak{e} q m(\mathfrak{e}) \int_{\varkappa} \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \varkappa \psi_{\varkappa}(\varkappa) d\varkappa d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e})}{\int_{q,\mathfrak{e}} q \int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa) d\varkappa d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e})} \\ - \frac{\begin{pmatrix} \int_{q,\mathfrak{e}} q m(\mathfrak{e}) \int_{\varkappa} \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \varkappa \psi_{\varkappa}(\varkappa)\, d\varkappa d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \\ \times \int_{q,\mathfrak{e}} \mathfrak{e} q \left[\int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,\mathfrak{e}}(q,\mathfrak{e}) \end{pmatrix}}{\left(\int_{q,\mathfrak{e}} q \left[\int_{\varkappa} \mathbf{1}_{e > \widehat{e}} \psi_{\varkappa}(\varkappa) d\varkappa\right] d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e})\right)^2} \end{array}\right]$$

$$= -\frac{E_{\psi_p}\left[\frac{1}{p}\right] \psi_u(1) \left(E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}[m(\mathfrak{e}) e | e > \widehat{e}] - E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}[m(\mathfrak{e}) | e > \widehat{e}] E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}[e | e > \widehat{e}]\right)}{\Psi_u(1)}.$$

Therefore, we obtain

$$t''(0) = -2\frac{E_{\psi_p}\left[\frac{1}{p}\right] \psi_u(1)}{\Psi_u(1)} Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}(m(\mathfrak{e}), e | e > \widehat{e}), \tag{25}$$

and we can write

$$t(\tau) = \tau E_{\widetilde{\psi}_e}[e|e > \hat{e}] - \tau^2 \frac{E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)}{\Psi_u(1)} Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m(\mathfrak{e}), e|e > \widehat{e}\right) + o\left(\tau^2\right).$$

Output. Next, we derive a Taylor approximation for the level of output. For τ small enough that firms never exit after drilling, we can rewrite (3) as:

$$S(\tau, \tilde{e})|_{\text{no post drilling exit}} = \int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h(q,p,\mathfrak{e},\tau)\geq pu}\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}). \tag{26}$$

Differentiating, we obtain that for the relevant range:

$$\frac{\partial S}{\partial \tau} = \int_{u,q,p,\mathfrak{e}} q\delta_{\frac{h(p,q,\mathfrak{e},\tau)}{p}=u} \frac{1}{p}\frac{\partial h}{\partial \tau} d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

Therefore, using (24) and (21), we get

$$\begin{aligned} \frac{\partial S}{\partial \tau}|_{\tau=0} &= -E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)\int_{q,\mathfrak{e}} q E_{\psi_\varkappa}(\mathfrak{e})\, d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \\ &= -E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)\, E_{\psi_q}[q]\, E_{\tilde{\psi}_e}(e). \end{aligned}$$

This expression is independent of $\widehat{e}$, so that, as before, the introduction of certification on top of an output tax does not change supply at first order. Compared to laissez-faire, the output tax shifts production in proportion to the slope of the supply curve and the size of the tax, namely $\tau E_{\widetilde{\psi}_e}(e)$ where $E_{\widetilde{\psi}_e}(e)$ is the $q-$weighted expectation of emissions. That is:

$$S^V = S^U + o(\tau) = S^{LF} - \tau \dot{S}^{LF} E_{\tilde{\psi}_e}(e) + o(\tau). \tag{27}$$

Emissions. We follow similar steps to derive the change in emissions. We differentiate (5) taking into account that for the relevant range of τ, there is no exit post drilling:

$$\begin{aligned} \frac{\partial G}{\partial \tau} &= \int_{u,q,p,\mathfrak{e}} \mathfrak{e}q\left[\frac{1}{p}\frac{\partial h}{\partial \tau}\delta_{\frac{h(p,q,\mathfrak{e},\tau)}{p}=u}\int_{\varkappa}\left(\mathbf{1}_{\mathfrak{e}\varkappa<\widehat{e}}(1-a) + \mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}}\right)\varkappa\psi_\varkappa(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \\ &\quad + \int_{u,q,p,\mathfrak{e}} \mathfrak{e}q\left[\mathbf{1}_{\frac{h(q,p,\mathfrak{e},\tau)}{p}\geq u}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa<\widehat{e}}\left(-\frac{\partial a}{\partial \tau}\right)\varkappa\psi_\varkappa(\varkappa)\, d\varkappa\right] d\Psi_{u,q,\mathfrak{e},p}(u,q,\mathfrak{e},p). \end{aligned}$$

Using (17) and (24), we can then write:

$$\begin{aligned}\frac{\partial G}{\partial \tau}|_{\tau=0} &= -E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)\int_{q,\mathfrak{e}} m(\mathfrak{e})\,\mathfrak{e} q d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) - \frac{\Psi_u(1)}{b''(0)}\int_{q,e} qe^2\mathbf{1}_{e<\hat{e}} d\Psi_{q,e}(q,e),\\ &= -E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)E_{\psi_q}[q]E_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e} m(\mathfrak{e})) - \frac{\Psi_u(1)E_{\psi_q}[q]}{b''(0)}E_{\widetilde{\psi}_e}\left[e^2|e<\hat{e}\right]\widetilde{\Psi}_e(\hat{e}).\end{aligned}$$

We can then write the change in emissions under certification relative to laissez-faire as:

$$G^V - G^{LF} = -\tau\dot{S}^{LF}E|_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e} m(\mathfrak{e})) - \tau\frac{S^{LF}}{b''(0)}E_{\widetilde{\psi}_e}\left[e^2|e<\hat{e}\right]\widetilde{\Psi}_e(\hat{e}) + o(\tau),$$

which captures the sum of the supply response from higher taxes and the effect from abatement. For output taxation, ε is a constant equal to $E|_{\widetilde{\psi}_\mathfrak{e}}(e)$, so that the change in emissions is given by:

$$G^U - G^{LF} = -\tau\dot{S}^{LF}\left(E|_{\widetilde{\psi}_\mathfrak{e}}(e)\right)^2 + o(\tau). \tag{28}$$

Therefore, using (21), we can write the change in emissions induced by certification above an output tax as:

$$G^V - G^U = -\tau\dot{S}^{LF}Cov_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e}, m(\mathfrak{e})) - \tau\frac{S^{LF}}{b''(0)}E_{\widetilde{\psi}_e}\left[e^2|e<\hat{e}\right]\widetilde{\Psi}_e(\hat{e}) + o(\tau). \tag{29}$$

The first term corresponds to a reallocation effect and the second term to the abatement effect. We can also rewrite this expression more intuitively as

$$\begin{aligned}G^V - G^U &= -\tau\dot{S}^{LF}\left[Var_{\widetilde{\psi}_\mathfrak{e}}(m(\mathfrak{e})) + Cov_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e} - m(\mathfrak{e}), m(\mathfrak{e}))\right] \\ &\quad -\tau\frac{S^{LF}}{b''(0)}E_{\widetilde{\psi}_e}\left[e^2|e<\hat{e}\right]\widetilde{\Psi}_e(\hat{e}) + o(\tau),\end{aligned} \tag{30}$$

which is equation (17) from Proposition 4.

This expression makes clear that the reallocation effect can be decomposed into two terms. First, there is the usual term, namely the variance of the pre-abatement revealed emission rate, except that here, it is the variance of the expected revealed *perceived* emission rate as firms decide whether or not to drill before they know their actual emission rate. Second, there is a correction term, which captures the fact that with partial unraveling, firms with perceived emissions rate $\mathfrak{e}$ will not be taxed on average at that rate but at rate $m(\mathfrak{e})$. The rates $\mathfrak{e}$ and $m(\mathfrak{e})$ generally differ because uncertified firms are taxed at the same common average rate regardless of their $\mathfrak{e}$, which benefits firms with a high $\mathfrak{e}$. Intuitively, firms adjust their entry rate in proportion to $m(\mathfrak{e}) - E_{\widetilde{\psi}_e}[e]$, so that high $\mathfrak{e}$ firms enter relatively less, reducing emissions. However, the term $Var_{\widetilde{\psi}_\mathfrak{e}}(m(\mathfrak{e}))$ understates the emission reduction because high $\mathfrak{e}$ firms have an expected tax rate above $m(\mathfrak{e})$ (and low $\mathfrak{e}$ firms below $m(\mathfrak{e})$). The covariance term adjusts for the gap.

Under full information (i.e. full unraveling), we get that $\varepsilon = e$, so that (29) becomes:

$$G^{FI} - G^{U} = -\tau \dot{S}^{LF} Var_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e}) - \tau \frac{S^{LF}}{b''(0)} E_{\widetilde{\psi}_e}\left[e^2\right] + o(\tau).$$

The reallocation effect depends on the variance of the ex-ante emission rates since drilling occurs before firms know their exact emission rate, while the abatement effects depends on ex-post emissions rate since abatement occurs post-drilling.

Producer surplus. Again, we follow the same logic to derive the change in producer surplus. As welfare changes will be second order, we need to do a second order Taylor expansion. Differentiating (21) leads to:

$$\frac{\partial PS^V}{\partial \tau} = \int_{u,q,p,\mathfrak{e}} q \mathbf{1}_{\frac{h(q,p,\mathfrak{e},\tau)}{p} \geq u} \frac{\partial h}{\partial \tau} d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}). \tag{31}$$

In particular, using (24), (21), (4), and (6), we get

$$\begin{aligned} \frac{\partial PS^V}{\partial \tau}|_{\tau=0} &= -\Psi_u(1) \int_{u,q,p,\mathfrak{e}} q m(\mathfrak{e}) \, d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \\ &= -\Psi_u(1) E_{\psi_q}[q] E_{\widetilde{\psi}_e}(e) = -G^{LF}. \end{aligned}$$

Further differentiating (31), we obtain:

$$\begin{aligned} \frac{\partial^2 PS^V}{(\partial \tau)^2} &= \int_{u,q,p,\mathfrak{e}} q \delta_{\frac{h(p,q,\tau,\mathfrak{e})}{p} = u} \frac{1}{p} \left(\frac{\partial h}{\partial \tau}\right)^2 d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \\ &\quad + \int_{u,q,p,\mathfrak{e}} q \mathbf{1}_{\frac{h(p,q,\tau,\mathfrak{e})}{p} \geq u} \frac{\partial^2 h}{\partial \tau^2} d\Psi_{u,q,\mathfrak{e},p}(u,q,\mathfrak{e},p). \end{aligned} \tag{32}$$

Differentiating (23) for τ sufficiently small that there is no exit post drilling gives:

$$\frac{\partial^2 h}{\partial \tau^2} = \int_{\varkappa} \left[\mathbf{1}_{\mathfrak{e}\varkappa \leq \widehat{e}} \left(\mathfrak{e}\varkappa \frac{\partial a}{\partial \tau} - \frac{F''(\tau)}{q} \right) + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}} \left(-t''(\tau)\right) \right] \psi_{\varkappa}(\varkappa) \, d\varkappa.$$

Using this expression, together with (17), (25), (24) in (32), we obtain:

$$\begin{aligned}\frac{\partial^2 PS^V}{(\partial\tau)^2}|_{\tau=0} &= E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)\int_{q,\mathfrak{e}} q\,(m(\mathfrak{e}))^2\,d\Psi_{q,\mathfrak{e}}(q,\mathfrak{e}) \\ &+\Psi_u(1)\int_{q,e}\mathbf{1}_{e\leq\widehat{e}}\left(\frac{qe^2}{b''(0)}-F''(0)\right)d\Psi_{q,e}(q,e) \\ &+2E_{\psi_p}\left[\frac{1}{p}\right]\psi_u(1)\,Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}(m(\mathfrak{e}),e|e>\widehat{e})\,E_{\psi_q}[q]\left(1-\widetilde{\Psi}(\widehat{e})\right)\end{aligned}$$

Therefore, using (14), we obtain that the producer surplus under voluntary certification is given by:

$$\begin{aligned}PS^V &= PS^{LF}-\tau G^{LF}+\dot{S}^{LF}\frac{\tau^2}{2}\left(E|_{\widetilde{\psi}_{\mathfrak{e}}}\left[m(\mathfrak{e})^2\right]+2Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}(m(\mathfrak{e}),e|e>\widehat{e})\left(1-\widetilde{\Psi}(\widehat{e})\right)\right) \\ &+\frac{\tau^2 S^{LF}E_{\widetilde{\psi}_e}(e^2|e<\hat{e})\,\widetilde{\Psi}_e(\hat{e})}{2b''(0)}-\Psi_u(1)\,\Psi(\widehat{e})\,F+o\left(\tau^2\right)\end{aligned}$$

For an output tax, we get that

$$PS^U-PS^{LF}=-\tau G^{LF}+\frac{\tau^2}{2}\dot{S}^{LF}E_{\widetilde{\psi}_e}(e)^2+o\left(\tau^2\right). \tag{33}$$

Therefore, the change in producer surplus (net of certification tax) from the introduction of certification on top of the output tax is given by:

$$\begin{aligned}PS^V\ -PS^U &= \dot{S}^{LF}\frac{\tau^2}{2}\left(Var_{\widetilde{\psi}_{\mathfrak{e}}}[m(\mathfrak{e})]+2Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}(m(\mathfrak{e}),e|e>\widehat{e})\left(1-\widetilde{\Psi}(\widehat{e})\right)\right) \\ &+\frac{\tau^2 S^{LF}E_{\widetilde{\psi}_e}(e^2|e<\hat{e})\,\widetilde{\Psi}_e(\hat{e})}{2b''(0)}-\Psi_u(1)\,\Psi(\widehat{e})\,F+o\left(\tau^2\right)\end{aligned}$$

The inequality $m(\mathfrak{e})\neq\mathfrak{e}$ for $\hat{e}<e$ holds solely because all uncertified firms pay the same output tax regardless of their perceived emission rate $\mathfrak{e}$. As a result

$$Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}(m(\mathfrak{e}),e|e>\widehat{e})\left(1-\widetilde{\Psi}(\widehat{e})\right)=Cov_{\widetilde{\psi}_{\mathfrak{e}}}(\mathfrak{e}-m(\mathfrak{e}),m(\mathfrak{e}))\,. \tag{34}$$

To see this formally, we use (21), to write:

$$\begin{aligned}
& Cov_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m\left(\mathfrak{e}\right),e|e>\widehat{e}\right)\left(1-\widetilde{\Psi}\left(\widehat{e}\right)\right)-Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}-m\left(\mathfrak{e}\right),m\left(\mathfrak{e}\right)\right)\\
= & \left(E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m\left(\mathfrak{e}\right)e|e>\widehat{e}\right)-E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m\left(\mathfrak{e}\right)|e>\widehat{e}\right)E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(e|e>\widehat{e}\right)\right)\left(1-\widetilde{\Psi}_{e}\left(\hat{e}\right)\right)-E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}m\left(\mathfrak{e}\right)\right)\\
& +E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}\right)E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(m\left(\mathfrak{e}\right)\right)+E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(m\left(\mathfrak{e}\right)^{2}\right)-\left(E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(m\left(\mathfrak{e}\right)\right)\right)^{2}\\
= & E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(m\left(\mathfrak{e}\right)^{2}\right)-E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}m\left(\mathfrak{e}\right)\right)+E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m\left(\mathfrak{e}\right)\mathbf{1}_{e>\widehat{e}}e\right)-E_{\widetilde{\psi}_{\mathfrak{e},\varkappa}}\left(m\left(\mathfrak{e}\right)\mathbf{1}_{e>\widehat{e}}\right)e^{unc}\left(0\right)\\
= & E|_{\widetilde{\psi}_{\mathfrak{e}}}\left[\begin{array}{c}
\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)+\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}e^{unc}\left(0\right)d\Psi_{\varkappa}\left(\varkappa\right)\right)^{2}\\
-\mathfrak{e}\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)+\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}e^{unc}\left(0\right)d\Psi_{\varkappa}\left(\varkappa\right)\right)\\
\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)+\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}e^{unc}\left(0\right)d\Psi_{\varkappa}\left(\varkappa\right)\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\\
-e^{unc}\left(0\right)\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)+\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}e^{unc}\left(0\right)d\Psi_{\varkappa}\left(\varkappa\right)\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)
\end{array}\right]\\
= & E|_{\widetilde{\psi}_{\mathfrak{e}}}\left[\begin{array}{c}
\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\right)^{2}+\left(e^{unc}\left(0\right)\right)^{2}\left(\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)\right)^{2}\\
+2e^{unc}\left(0\right)\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\right)\left(\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)\right)-\mathfrak{e}\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\\
-\mathfrak{e}e^{unc}\left(0\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)+\left(\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\right)\left(\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\right)\\
+e^{unc}\left(0\right)\left(\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\\
-e^{unc}\left(0\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)\\
-\left(e^{unc}\left(0\right)\right)^{2}\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}d\Psi_{\varkappa}\left(\varkappa\right)
\end{array}\right]\\
= & 0.
\end{aligned}$$

In the last line, we made repeated use of the fact that inside the brackets, for a given $\mathfrak{e}$, we have that $\int_{\varkappa}\mathbf{1}_{e\leq\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)+\int_{\varkappa}\mathbf{1}_{e>\widehat{e}}ed\Psi_{\varkappa}\left(\varkappa\right)=\mathfrak{e}$.

We can then rewrite the change in producer surplus as:

$$\begin{aligned}
PS^{V}-PS^{U} = & \dot{S}^{LF}\frac{\tau^{2}}{2}\left(Var_{\widetilde{\psi}_{\mathfrak{e}}}\left[m\left(\mathfrak{e}\right)\right]+2Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}-m\left(\mathfrak{e}\right),m\left(\mathfrak{e}\right)\right)\right)\\
& +\frac{\tau^{2}S^{LF}E_{\widetilde{\psi}_{e}}\left(e^{2}|e<\hat{e}\right)\widetilde{\Psi}_{e}\left(\hat{e}\right)}{2b''\left(0\right)}-\Psi_{u}\left(1\right)\Psi\left(\widehat{e}\right)F+o\left(\tau^{2}\right).
\end{aligned} \tag{35}$$

Welfare. Using (10), we obtain that the change in welfare from introducing output taxation at the Pigovian level ($\tau=v$) relative to laissez-faire is given by:

$$\begin{aligned}
W^{U}\left(v\right)-W^{LF} = & PS^{U}-\left(PS^{LF}-vG^{LF}\right),\\
= & \frac{v^{2}}{2}\dot{S}^{LF}E_{\widetilde{\psi}_{e}}\left(e\right)^{2}+o\left(\tau^{2}\right).
\end{aligned} \tag{36}$$

where the second line used (33). Then introducing certification on top of the output tax leads to a further welfare change equal to the change in the producer surplus (for $\tau = v$). Using (35), we get:

$$\begin{aligned} & W^V(v) - W^U(v) \\ = \quad & \dot{S}^{LF}\frac{v^2}{2}\left(Var_{\widetilde{\psi}_{\mathfrak{e}}}\left[m\left(\mathfrak{e}\right)\right] + 2Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e} - m\left(\mathfrak{e}\right), m\left(\mathfrak{e}\right)\right)\right) \\ & + \frac{v^2 S^{LF} E_{\widetilde{\psi}_e}\left(e^2 | e < \hat{e}\right)\widetilde{\Psi}_e\left(\hat{e}\right)}{2b''\left(0\right)} - N^{LF}\Psi\left(\widehat{e}\right)F + o\left(\tau^2\right), \end{aligned} \tag{37}$$

which is equation (16) in Proposition (4) and where we have defined $N^{LF} = \Psi_u(1)$ as the mass of firms under Laissez-Faire. The terms on the second line correspond to the welfare gains from abatement and to the costs of certification. The term on the first line corresponds to the reallocation effect. Without post-drilling uncertainty, the covariance terms drops and we recover the expression from the baseline model.[5] With uncertainty on the emissions rate after drilling, we get an additional covariance term. This covariance term is the same expression that appeared for the change in emissions (30). The intuition for why it should be included is the same as in equation (30): the variance term alone underestimates the benefit from expanding production of low $\mathfrak{e}$ firms and contracting production of high $\mathfrak{e}$ firms.

B.4 Data on the Permian basin

Detailed production and flaring data in the Permian basin are pulled from online portals of the states of Texas and New Mexico. In 2019 daily production was about 3.8 million barrels per day of oil and 2.5 million barrels of oil-equivalent of natural gas. There were approximately 66,000 leases or properties actively producing, 85% of which were in Texas.[6] Texas produced 75-80% of the basin's oil and gas. To focus on the lifetime production of unconventional wells, we include data from the $\sim 20,000$ sites whose first production began between 2012 and 2019. We remove leases that produce sporadically ($\sim 2\%$ of production in 2019).

While 33% of wells produce only gas, and 11% only oil, the overwhelming majority of output comes from wells that produce a mix of the two. Figure B.1 presents a kernel density of the oil share, weighted by well production. Most output comes from wells producing about

[5]If $\Psi_{\varkappa}$ is degenerate, we get: $Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e} - E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right], E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right]\right) = E|_{\widetilde{\psi}_{\mathfrak{e}}}\left[\mathfrak{e}\varepsilon\right] - E|_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}\right)^2 - Var_{\widetilde{\psi}_{\mathfrak{e}}}\left(\varepsilon\right) = 0$.

[6]The monthly data from Texas are reported at the lease level (which may contain more than one well), while those from New Mexico are reported at the well level for production, and the property level for venting and flaring. The median NM property has a single well in 2019, and three wells on average. The well-level data is collapsed to the property in NM, and collectively these are all referred to as "sites" or "leases."

80% oil, which may make flaring an attractive option for economically disposing of methane in the absence of sufficient collection infrastructure.

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 natural logarithm of 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 of each fuel in barrels of oil-equivalent (BOE) from site i in month t, we conduct an exponential decline curve analysis (DCA) according to

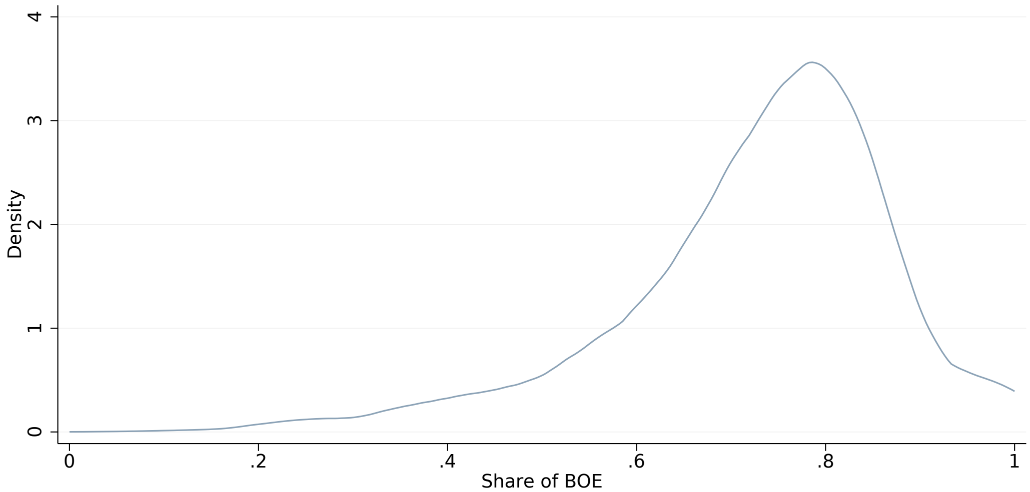
$$y_{it} = \gamma_i + \sum_v [\eta_v \mathbf{1}_{y=1} + \beta_v \mathbf{1}_{y=1} age_{it}] + \delta age_{it} + u_{it} \tag{38}$$

where γ_i are site-level fixed effects. For the first year of production we estimate a separate intercept and age decay by vintage, v (η_v, β_v), and a common rate of decay thereafter, δ. For each site we then project the expected ultimate recovery by combining the site-level fixed effects and estimated decay parameters. These estimates are conducted separately by fuel, and then added together in BOE. This is the measure of well output, q_i, that we use in our calibration exercise below.

This specification is also used to predict the lifetime well-level emissions from venting and flaring. An indicator for any emissions at well i in month t is used as the dependent variable in a linear probability model. This is then combined with the predicted values of the exponentiated DCA when the natural log of venting/flaring volume is the dependent variable, and added up over the lifetime of the well. Finally, NM separately reports volumes vented versus flared, while TX does not. We use the data in NM to estimate the relationship between the share of vented and flared in a regression analogous to (38) above, and apply those estimates to the wells in TX. For flared volumes, we adopt the emissions rate of Plant et al. (2022), who find that flares remove 91.1% of methane.

The second source of emissions estimates comes from Omara et al. (2024), which report "loss rates", i.e. the share of gas production that is emitted as methane, for a 0.1° x 0.1° grid of the Permian. These estimates are for 2021, and therefore aggregate the emissions from various vintages producing simultaneously in that year. We project grid cell loss rates on the share of grid cell production from each vintage and fuel in order to estimate location- and vintage-specific loss rates.

Figure B.1: Permian Basin Lease Oil Share in 2019



In particular, we regress loss rate r_l for location l on the vector of production shares for each fuel in l so that

$$r_l = \sum_f \sum_v s_l^{fv} \eta^{fv} + u_l$$

where f represents oil, gas, and natural gas liquids, and v represent vintages. The residual u_l is the location-specific loss rate for all wells in l. The estimated loss rate for well i of vintage v in location l is then $\sum_f s_{il}^{f,2019} \hat{\eta}^{f,2019} + \hat{u}_l$ based on the lifetime share of each fuel produced by i.

Because we observe venting/flaring emissions at the well level directly, we scale the emissions calculated from grid-level loss rates so that the total adds up to the figures in Omara et al. (2024) (i.e. to avoid double counting). Denoting the venting/flaring loss rate for site i as vf_i, gas production g_i, total production q_i, and the scaling factor k, we calculate the emissions rate as the sum from these two sources:

$$e_i = vf_i + k \frac{g_i \left[\sum_f s_{il}^{f,2019} \hat{\eta}^{f,2019} + \hat{u}_l \right]}{q_i}$$

Figure B.2: Log(Monthly Production) Profile by Vintage

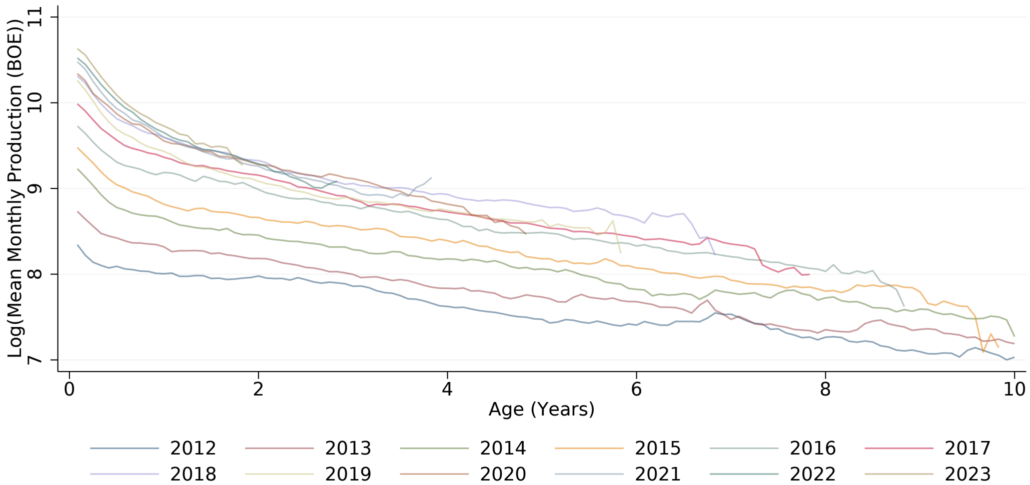
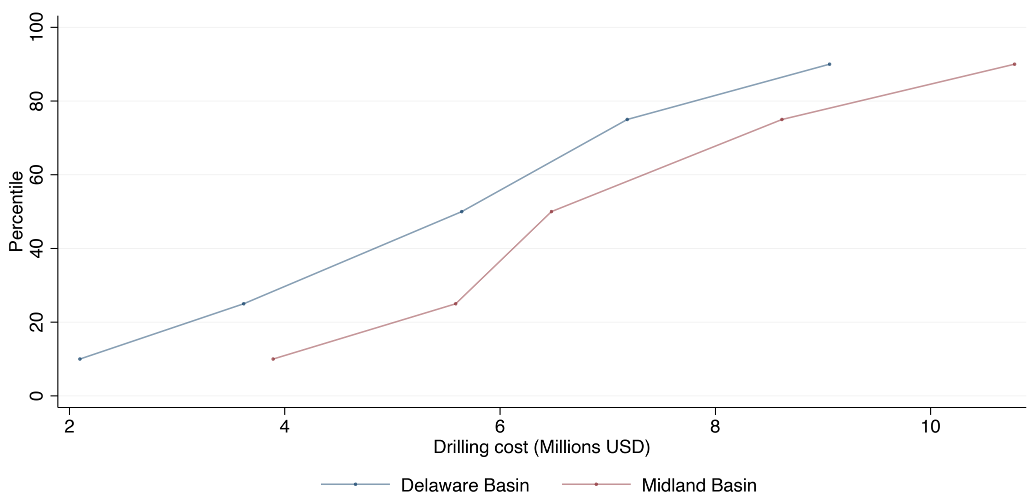


Figure B.3: Distributions of Drilling Costs in the Permian



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

B.5 Calibration and expressions needed for the empirical application

We now explain how we calibrate the model to our data and collect the expressions needed for the empirical application.

Distribution of q and p: The data directly give us the distribution of quantity and prices (through the oil share) and therefore aggregate quantity or the mean of the inverse price.

Distribution of ex ante emissions rates: We need to estimate the distribution of pre-drilling expected emissions rates, $\mathfrak{e}$, and that of the post-drilling shocks, $\varkappa$, from the distribution of emissions rates observed in the data, e_i described above.

For all wells, i, we run a quantity-weighted Poisson regression of the emissions rate on the quantity of oil and gas, location fixed effects, and vintage fixed effects with multiplicative errors. The distribution of predicted values is our distribution of $\mathfrak{e}$. Defining by $\omega_i \equiv q_i / \left(\sum_j q_j\right)$, the output share of well i, we get that by definition, the quantity-weighted mean of ex-ante emissions rates is equal to the quantity-weighted mean emission rate in the data:

$$E_{\tilde{\psi}_{\mathfrak{e}}}(\mathfrak{e}) = \sum_i \omega_i \mathfrak{e}_i = \sum_i \omega_i e_i.$$

Distribution of the post-drilling shocks: We define the raw residual for each well as $\varkappa_i^{raw} \equiv e_i/\mathfrak{e}_i$. The distribution of post-drilling shocks $\varkappa$ is an adjusted version of the quantity-weighted distribution of $\varkappa_i^{raw}$, which ensures that the quantity-weighted mean $\varkappa$ is 1 and that the quantity-weighted variance of emissions rates $e = \mathfrak{e}\varkappa$ matches the data. More specifically, after windsorising the top percentile of $\varkappa_i^{raw}$, we define

$$\varkappa_i^{mean\_adj} = \frac{\varkappa_i^{raw}}{\sum_j \omega_j \varkappa_j^{raw}}$$

which has a quantity-weighted mean of 1. We then set

$$\varkappa_i = K\left(\varkappa_i^{mean\_adj} - 1\right) + 1,$$

for a constant K. This ensures that $\varkappa_i$ is also mean 1, and we use a K such that:

$$\sum_i \omega_i e_i^2 = \left(\sum_i \omega_i \mathfrak{e}_i^2\right)\left(\sum_i \omega_i \varkappa_i^2\right).$$

This is achieved with

$$K = \left[\frac{\frac{\sum_i q_i e_i^2}{\sum_i q_i \mathfrak{e}_i^2} - 1}{\sum_i \omega_i \left(\varkappa_i^{mean\_adj} - 1 \right)^2} \right]^{\frac{1}{2}}.$$

In our data, this value is 0.57. The distribution of the post-drilling shocks is then the quantity-weighted distribution of $\varkappa_i$, which is applied to all wells' ex ante emissions rate as a multiplicative iid shock. By construction we get

$$E_{\psi_\varkappa}(\varkappa) = \sum_i \omega_i \varkappa_i = 1,$$

and $E_{\widetilde{\psi}_e}(e) = E_{\widetilde{\psi}_\mathfrak{e}}(\mathfrak{e}) E_{\psi_\varkappa}(\varkappa) = \sum_i \omega_i e_i$. In addition, we have that since $\varkappa$ is an iid shock:

$$E_{\widetilde{\psi}_e}\left(e^2\right) = E_{\widetilde{\psi}_\mathfrak{e}}\left(\mathfrak{e}^2\right) E_{\psi_\varkappa}\left(\varkappa^2\right) = \left(\sum_i \omega_i \mathfrak{e}_i^2\right)\left(\sum_i \omega_i \varkappa_i^2\right) = \sum_i \omega_i e_i^2,$$

by construction. In other words, the quantity-weighted mean and variance of emissions rates used to form expectations are the same as in the data.

Other moments. We calibrate the supply slope using equation (16), the supply elasticity of 1.63 from Newell and Prest (2019), and the sample analogue for the expected inverse price. Finally we calibrate the abatement function by noting that abatement for an individual firm is (to a first order): $a(e) = \frac{\tau e}{b''(0)}$. Marks (2022) finds that taxing at a social cost of emissions, $\tau = \nu$, firms abate 51.3% and we set $b''(0)$ such that the quantity-weighted emission reduction matches[7]:

$$\sum_i \left(\frac{\nu e_i}{b''(0)} \frac{e_i q_i}{\sum_j e_j q_j} \right) = 0.513.$$

Formulas. Table B.1 gives the formulas for the expressions used in Table 1 comparing outcomes under an output tax and in laissez-faire, using how we compute $E_{\widetilde{\psi}_e}(e)$ and $\dot{S}^{LF}$ from the data.

Table B.2 performs an analogous exercise but compares the expressions under voluntary certification to that of an output tax used both in Table 1 and in Figure 2. In the case of full unraveling (reported in Table 1), we note that we can directly compute

$$E_{\widetilde{\psi}_\mathfrak{e}}\left(\mathfrak{e}^2\right) = \sum_i \omega_i \mathfrak{e}_i^2 \text{ and } E_{\widetilde{\psi}_e}\left(e^2\right) = \sum_i \omega_i e_i^2,$$

[7]This is Marks' estimate when weighting by size, as we do here.

from which we directly obtain the variance as well.

With partial unraveling up to some threshold $\hat{e}$, we need to compute $m(\mathfrak{e}) = E_{\psi_\varkappa}[\varepsilon(\mathfrak{e})]$ for all wells. To do this, we look at all possible realizations of $e = \mathfrak{e}\varkappa$ – there are N^2 realizations corresponding to all values of $\mathfrak{e}$ and all potential residuals. Focusing on realizations with $e > \hat{e}$, we then compute the quantity-weighted emissions rate above $\hat{e}$: $e^{unc} = E_{\tilde{\psi}_e}[e|e > \hat{e}]$. For each well i, we then us the empirical counterpart to the expression for $m(\mathfrak{e})$ which is:

$$m(\mathfrak{e}_i) = \Psi_\varkappa\left(\frac{\hat{e}}{\mathfrak{e}_i}\right)\mathfrak{e}_i E_{\psi_\varkappa}\left(\varkappa|\varkappa < \frac{\hat{e}}{\mathfrak{e}_i}\right) + \left(1 - \Psi_\varkappa\left(\frac{\hat{e}}{\mathfrak{e}_i}\right)\right)e^{unc}.$$

In this expression, $\mathfrak{e}_i$ is the pre-abatement expected emission rate for a given well estimated above, $\Psi_\varkappa$ is the constructed distribution of post-drilling shocks, such that $\Psi_\varkappa\left(\frac{\hat{e}}{\mathfrak{e}_l}\right)$ is the probability that a firm will draw a $\varkappa$ such that $e = \mathfrak{e}_i\varkappa \le \hat{e}$ and then certify in which case it is taxed at its own emission rate, and $E_{\psi_\varkappa}\left(\varkappa|\varkappa < \frac{\hat{e}}{\mathfrak{e}_i}\right)$ is the conditional mean of the post-drilling shock when it is distributed according to $\Psi_\varkappa$ and satisfies $\varkappa < \frac{\hat{e}}{\mathfrak{e}_i}$. The second term captures the probability that it will not certify and will be taxed at the rate e^{unc} for uncertified wells.

Having computed $m(\mathfrak{e}_i)$,we can then compute its variance as:

$$Var_{\tilde{\psi}_\mathfrak{e}}(m(\mathfrak{e})) = \sum_i \omega_i (m(\mathfrak{e}_i))^2 - \left(\sum_i \omega_i m(\mathfrak{e}_i)\right)^2,$$

with an analogous expression for the covariance term: $Cov_{\tilde{\psi}_\mathfrak{e}}(\mathfrak{e} - m(\mathfrak{e}), m(\mathfrak{e}))$ which we use to calculate the expressions in Table B.2.

We follow a similar strategy to compute $E_{\tilde{\psi}_e}(e^2|e < \hat{e})\widetilde{\Psi}_e(\hat{e})$, namely:

$$E_{\tilde{\psi}_e}\left(e^2|e < \hat{e}\right)\widetilde{\Psi}_e(\hat{e}) = E_{\tilde{\psi}_\mathfrak{e}}\left(\mathfrak{e}^2 E_{\psi_\varkappa}\left(1_{\mathfrak{e}\varkappa<\hat{e}}\varkappa^2\right)\right) = \sum_i \omega_i \mathfrak{e}_i^2 E_{\psi_\varkappa}\left(1_{\mathfrak{e}_i\varkappa<\hat{e}}\varkappa^2\right). \tag{39}$$

Table B.1: Output Tax: Difference from Laissez Faire (Approximation)

Outcome	Formula	Reference
q.-weigt. emissions	$E_{\widetilde{\psi}_e}(e) = \frac{\sum_i e_i q_i}{\sum_i q_i}$	
Supply slope	$\dot{S}^{LF} = E_{\psi_p}\left[\frac{1}{p}\right] S^{LF}\epsilon^S$	
Production	$-v\dot{S}^{LF}E_{\widetilde{\psi}_e}(e)$	eq (27
Emissions	$-v\dot{S}^{LF}\left(E_{\widetilde{\psi}_e}(e)\right)^2$	eq (28)
Producer Surplus	$-S^{LF}vE_{\widetilde{\psi}_e}(e) + \frac{1}{2}\dot{S}^{LF}\left(vE_{\widetilde{\psi}_e}(e)\right)^2$	eq (33)
Tax Revenue	$v\left(S^{LF} - \dot{S}^{LF}E_{\widetilde{\psi}_e}(e)\right)E_{\widetilde{\psi}_e}(e)$	$v(G^{LF}$+eq (28))
External Cost	$-v^2\dot{S}^{LF}\left(E_{\widetilde{\psi}_e}(e)\right)^2$	$v\times$eq (28)
Welfare	$\frac{v^2}{2}\dot{S}^{LF}\left(E_{\widetilde{\psi}_e}(e)\right)^2$	eq (36)

Table B.2: Voluntary certification Tax: Difference from Output tax (Approximation)

Outcome	Formula	Reference
Production	0	eq (27)
Emissions	$-\dot{S}^{LF}\begin{pmatrix} Var_{\widetilde{\psi}_{\mathfrak{e}}}\left(E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right]\right) \\ +Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e} - E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right], E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right]\right)\end{pmatrix}$ $-vS^{LF}\frac{1}{b''(0)}E_{\widetilde{\psi}_e}\left(e^2 \mid e < \hat{e}\right)\widetilde{\Psi}_e(\hat{e})$	eq (30)
Gross Prod. Surplus	$\frac{v^2}{2}\begin{pmatrix} \dot{S}^{LF}\begin{pmatrix} Var_{\widetilde{\psi}_{\mathfrak{e}}}\left(E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right]\right) \\ +2Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e} - E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right], E_{\psi_{\varkappa}}\left[\varepsilon(\mathfrak{e})\right]\right)\end{pmatrix} \\ +\frac{S^{LF}}{b''(0)}E_{\widetilde{\psi}_e}\left(e^2 \mid e < \hat{e}\right)\widetilde{\Psi}_e(\hat{e})\end{pmatrix}$	eq (35)
Tax Revenue	$v\left(G^V - G^U\right)$	
External Cost	$-v\left(G^V - G^U\right)$	
Certification Cost	$-N^{LF}\Psi_e(\hat{e})F$	
Welfare	Gross Prod. Surplus$-N^{LF}\Psi_e(\hat{e})F$	

B.6 Emissions Tax Relative to Laissez Faire with Uniformly Distributed Costs

We now solve the model exactly when the idiosyncratic cost shocks, u, are distributed uniformly over the interval $[\underline{u}, \bar{u}]$, with $1 \in (\underline{u}, \bar{u})$. To make the mapping with the data more transparent, we introduce a mass of entrants M (which was previously normalized to 1). Throughout this section we also introduce how we map the various expressions to the data.

Using (15) and the uniform u, we can write aggregate supply in laissez-faire as a function of a proportional price shock r (assuming that $1+r \in (\underline{u}, \bar{u})$) as:

$$S^{LF}(r) = M\frac{1+r-\underline{u}}{\overline{u}-\underline{u}}\int_{q,p} qd\Psi_q(q).$$

We can then derive the supply elasticity (at $r=0$) as:

$$\epsilon^S = \frac{d\ln S^{LF}(r)}{dr}|_{r=0} = \frac{d\ln(1+r-\underline{u})}{dr}|_{r=0} = \frac{1}{1-\underline{u}}.$$

Since we take ϵ^S from the data, we can then calibrate $\underline{u}$ as:

$$\underline{u} = 1 - \frac{1}{\epsilon^S}.$$

We then express supply as:

$$S^{LF} = M\frac{1-\underline{u}}{\overline{u}-\underline{u}}E_{\psi_q}(q) \Rightarrow \frac{M}{\overline{u}-\underline{u}} = \frac{S^{LF}}{(1-\underline{u})E_{\psi_q}(q)} = \frac{N^{LF}}{1-\underline{u}}, \tag{40}$$

where N^{LF} is the number of wells that we observe. We cannot separately identify M and $\overline{u}$, but none of the expressions below require us to do so.

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, $\beta\frac{a}{1-a} = \tau e$, yields an optimal abatement of $a^* = \frac{\tau e}{\beta+\tau e}$. We then get that

$$\tau e(1-a^*) + b(a^*) = \beta\ln\left(1+\frac{\tau e}{\beta}\right). \tag{41}$$

We use the same estimate from Marks (2022) that the quantity-weighted average abatement rate at the social cost of emissions is 51.3%, and calibrate β such that

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

We only use the exact model in the case of full certification, therefore we simply assume that the government imposes an emission tax on all producers of $\tau = v$. Using (2), we get that in this case:

$$\begin{aligned} h\left(q,p,\mathfrak{e},\tau\right) &= \int_{\varkappa}\left(p-\tau\mathfrak{e}\varkappa\left(1-a\right)-b\left(a\right)-\frac{F}{q}\right)^{+}\psi_{\varkappa}\left(\varkappa\right)d\varkappa \\ &= E_{\psi_x}\left[\left(p-\beta\ln\left(1+\frac{\tau\mathfrak{e}\varkappa}{\beta}\right)-\frac{F}{q}\right)^{+}\right] \end{aligned}$$

For each well i, we predict $\mathfrak{e}_i$ and derive the distribution ψ_x of $\varkappa$ as before. We can then compute $h\left(q_i,p_i,\mathfrak{e}_i,\tau\right)$ as an expected value using all possible values of the shocks $\varkappa$.

Using (3), we can then write output as

$$\begin{aligned} S &= M\int_{u,q,p,\mathfrak{e}}q\left[\mathbf{1}_{h(q,p,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\mathbf{1}_{p-\tau\mathfrak{e}\varkappa(1-a)-b(a)-\frac{F}{q}\geq 0}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right) \\ &= \frac{M}{\overline{u}-\underline{u}}\int_{q,p,\mathfrak{e}}q\left[\left(\frac{h\left(q,p,\mathfrak{e},\tau\right)}{p}-\underline{u}\right)^{+}\int_{\varkappa}\mathbf{1}_{p-\tau\mathfrak{e}\varkappa(1-a)-b(a)-\frac{F}{q}\geq 0}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]d\Psi_{q,p,\mathfrak{e}}\left(q,p,\mathfrak{e}\right) \end{aligned}$$

Using (40), we can then compute output in the data as:

$$\begin{aligned} S &= \frac{M}{\overline{u}-\underline{u}}\frac{1}{N^{LF}}\sum_i q_i\left(\frac{h\left(p_i,\mathfrak{e}_i,\tau\right)}{p_i}-\underline{u}\right)^{+}E_{\psi_{\varkappa}}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\right) \\ &= \frac{S^{LF}}{1-\underline{u}}\sum_i\omega_i\left(\frac{h\left(p_i,\mathfrak{e}_i,\tau\right)}{p_i}-\underline{u}\right)^{+}E_{\psi_{\varkappa}}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\right) \end{aligned} \quad (42)$$

Note that in line with what we do in Section B.3, this corresponds to expected output before the realization of the $\varkappa$ shocks. Using (40), we then get that the change in supply as a function of the tax is given by:

$$S-S^{LF}=S^{LF}\left(\frac{\sum_i\omega_i\left(\frac{h(p_i,\mathfrak{e}_i,\tau)}{p_i}-\underline{u}\right)^{+}E_{\psi_{\varkappa}}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\right)}{1-\underline{u}}-1\right). \quad (43)$$

We follow the same strategy for emissions. We get that, emissions are given by:

$$G=M\int_{u,q,p,\mathfrak{e}}q\left[\left(\frac{h\left(q,p,\mathfrak{e},\tau\right)}{p}-\underline{u}\right)^{+}\int_{\varkappa}\mathbf{1}_{p-\tau\mathfrak{e}\varkappa(1-a)-b(a)-\frac{F}{q}\geq 0}e\left(1-a\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right),$$

which in the data corresponds to:

$$G = \frac{S^{LF}}{1-\underline{u}} \sum_i \omega_i \left(\frac{h\left(p_i, \mathfrak{e}_i, \tau\right)}{p_i} - \underline{u} \right)^+ E_{\psi_\varkappa} \left(\mathbf{1}_{p_i - \beta \ln\left(1+\frac{v \mathfrak{e}_i \varkappa}{\beta}\right) - \frac{F}{q_i} \geq 0} \frac{\beta \mathfrak{e}_i \varkappa}{\beta + \tau \mathfrak{e}_i \varkappa} \right). \tag{44}$$

Laissez-faire emissions are the same as in the data, therefore we get that the change in emissions is given by:[8]

$$G - G^{LF} = \frac{S^{LF}}{1-\underline{u}} \sum_i \omega_i \left(\frac{h\left(p_i, \mathfrak{e}_i, \tau\right)}{p_i} - \underline{u} \right)^+ E_{\psi_\varkappa} \left(\mathbf{1}_{p_i - \beta \ln\left(1+\frac{v \mathfrak{e}_i \varkappa}{\beta}\right) - \frac{F}{q_i} \geq 0} \frac{\beta \mathfrak{e}_i \varkappa}{\beta + \tau \mathfrak{e}_i \varkappa} \right) - G^{LF}. \tag{45}$$

Using (11) and a similar approach, the producer surplus is given by:

$$PS = \frac{M}{\overline{u} - \underline{u}} \frac{1}{2} \int_{q,p,\mathfrak{e}} \frac{q}{p} \left(\left(h\left(q, p, \mathfrak{e}, \tau\right) - p \underline{u} \right)^+ \right)^2 \left(\int_\varkappa \mathbf{1}_{p - \tau \mathfrak{e} \varkappa (1-a) - b(a) - \frac{F}{q} \geq 0} \psi_\varkappa \left(\varkappa\right) d\varkappa \right) d\Psi_{q,p,\mathfrak{e}} \left(q, p, \mathfrak{e}\right) \tag{46}$$

We can then compute the producer surplus in the data (inclusive of certification costs) as:

$$PS = \frac{1}{2} \frac{S^{LF}}{1-\underline{u}} \sum_i \frac{\omega_i}{p_i} \left(\left(h\left(q_i p_i, \mathfrak{e}_i, \tau\right) - p_i \underline{u} \right)^+ \right)^2 E_{\psi_\varkappa} \left(\mathbf{1}_{p_i - \beta \ln\left(1+\frac{v \mathfrak{e}_i \varkappa}{\beta}\right) - \frac{F}{q_i} \geq 0} \right) \tag{47}$$

In laissez-faire, $\tau = 0$ and no firm pays certification so that:

$$PS^{LF} = \frac{1}{2} S^{LF} \sum_i p_i \omega_i \left(1 - \underline{u}\right). \tag{48}$$

Hence the producer surplus difference is:

$$\begin{aligned} & PS - PS^{LF} \\ = \; & \frac{1}{2} \frac{S^{LF}}{1-\underline{u}} \sum_i \omega_i \left[\frac{1}{p_i} \left(\left(h\left(q_i p_i, \mathfrak{e}_i, \tau\right) - p_i \underline{u} \right)^+ \right)^2 E_{\psi_\varkappa} \left(\mathbf{1}_{p_i - \beta \ln\left(1+\frac{v \mathfrak{e}_i \varkappa}{\beta}\right) - \frac{F}{q_i} \geq 0} \right) - p_i \left(1 - \underline{u}\right)^2 \right]. \end{aligned} \tag{49}$$

Tax revenues and changes in the social cost of emissions are trivially obtained from (44) as $T - T^{LF} = \tau G$ and $-v\left(G - G^{LF}\right)$. With $\tau = v$ and using (12), we can then express the change in welfare relative to laissez-faire as

$$W - W^{LF} = PS - PS^{LF} + vG^{LF}. \tag{50}$$

[8]With $\tau = 0$ and $F = 0$ (as there is no certification in laissez-faire), (44) gives $G^{LF} = S^{LF} \sum_i \omega_i \mathfrak{e}_i$, which by definition corresponds to laissez-faire emissions in the data.

We collect the expressions for the differences in the quantities of interest between the equilibrium with full certification and the laissez-faire equilibrium in the exact model in Table B.3. Note that in Table B.3, we compute the equilibrium in the exact model with $F = 0$.

Table B.3: Full certification equilbrium: Difference from Laissez Faire (Exact)

Outcome	Formula	Reference
Production	$S^{LF}\left(\frac{1}{1-\underline{u}}\sum_i \omega_i\left(\frac{h(p_i,\mathfrak{e}_i,\tau)}{p_i}-\underline{u}\right)^+ E_{\psi_\varkappa}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\right)-1\right)$	eq (43)
Emissions	$\frac{S^{LF}}{1-\underline{u}}\sum_i \omega_i\left(\frac{h(p_i,\mathfrak{e}_i,\tau)}{p_i}-\underline{u}\right)^+ E_{\psi_\varkappa}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\frac{\beta\mathfrak{e}_i\varkappa}{\beta+\tau\mathfrak{e}_i\varkappa}\right)-G^{LF}$	eq (45)
Producer Surplus	$\frac{1}{2}\frac{S^{LF}}{1-\underline{u}}\sum_i \omega_i\left[\begin{array}{c}\frac{1}{p_i}\left(\left(h\left(q_i,p_i,\mathfrak{e}_i,\tau\right)-p_i\underline{u}\right)^+\right)^2 E_{\psi_\varkappa}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\right)\\ -p_i\left(1-\underline{u}\right)^2\end{array}\right]$	eq (49)
Tax Revenue	$\frac{vS^{LF}}{1-\underline{u}}\sum_i \omega_i\left(\frac{h(p_i,\mathfrak{e}_i,\tau)}{p_i}-\underline{u}\right)^+ E_{\psi_\varkappa}\left(\mathbf{1}_{p_i-\beta\ln\left(1+\frac{v\mathfrak{e}_i\varkappa}{\beta}\right)-\frac{F}{q_i}\geq 0}\frac{\beta\mathfrak{e}_i\varkappa}{\beta+\tau\mathfrak{e}_i\varkappa}\right).$	$v$eq (44))
External Cost	$-v\left(G-G^{LF}\right)$	$v\times$eq (45)
Welfare	$PS-PS^{LF}+vG^{LF}$	eq (50)

B.7 Welfare change along the algorithm

Figure 2 shows the welfare gains with progressive unraveling along the lines of the algorithm presented in Section 2.4. At different steps of the algorithm, welfare differs from welfare in a certification equilibrium with the same threshold of certification $\widetilde{e}$. This is because in a certification equilibrium non-certified firms pay the output tax given by (8) while under the algorithm, non-certified firms pay an output tax, $t_n(\tau)$ which corresponds to the certification threshold of the previous round, denoted $\widehat{e}_{n-1}$. That is, the output tax under the algorithm at step n is given by:[9]

$$t_n\left(\tau\right)=\tau\frac{\int_{u,q,p,\mathfrak{e}}\mathfrak{e}q\left[\mathbf{1}_{h^{\text{alg}}_{n-1}(p,q,\mathfrak{e},\tau)\geq pu}\int_\varkappa \mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_{n-1}}\mathbf{1}_{p-t_{n-1}\geq 0}\varkappa\psi_\varkappa\left(\varkappa\right)d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right)}{\int_{u,q,p,\mathfrak{e}}q\left[\mathbf{1}_{h^{\text{alg}}_{n-1}(p,q,\mathfrak{e},\tau)\geq pu}\int_\varkappa \mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_{n-1}}\mathbf{1}_{p-t_{n-1}\geq 0}\psi_\varkappa\left(\varkappa\right)d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right)}\text{ for }n\geq 1. \tag{51}$$

This expression differs from (8) in that taxes and therefore the function h^{alg}_{n-1} now depends on the certification in the previous round. As no firms certify initially, $\widehat{e}_0=\underline{e}$ and $t_1=\tau E\left(e\right)$.

[9]Here we anticipate that such a threshold exists for $n-1$. The next equations show that there is a threshold for step n, given that there is a threshold at 0 (the lower bound), then by induction, there is always a threshold.

Under the algortihm, the firm's expected net revenue from drilling per unit before it learns its emission rate is given by

$$h_n^{\text{alg}}(q,p,\mathfrak{e},\tau) \equiv \int_{\varkappa} \max\left(p - \tau\mathfrak{e}\varkappa(1-a) - b(a) - \frac{F}{q}, p - t_n\right)\psi_{\varkappa}(\varkappa)\,d\varkappa \tag{52}$$

where we ignored the possibility that firms exit post drilling (as the focus in this section is on small τ). Note that

$$h_n^{\text{alg}}(q,p,\mathfrak{e},0) = p. \tag{53}$$

Threshold. Under the algorithm, firms certify whenever

$$(p - \tau e(1-a) - b(a))\,q - F \geq (p - t_n)q$$

$$\Longleftrightarrow t_n - \tau e(1-a) - b(a) \geq \frac{F}{q},$$

given that the left-hand side is decreasing in e, firms certify whenever their emission rate is below a threshold $\widehat{e}_n(q,\tau)$ defined implicitly through

$$t_n - \tau\widehat{e}_n(q,\tau)(1 - a(\tau,\widehat{e}_n(q,\tau))) - b(a(\tau,\widehat{e}_n(q,\tau))) = \frac{F}{q}. \tag{54}$$

We can then write (52) as

$$h_n^{\text{alg}}(q,p,\mathfrak{e},\tau) = \int_{\varkappa}\left(\mathbf{1}_{\mathfrak{e}\varkappa\leq\widehat{e}_n(q,\tau)}\left(p - \tau\mathfrak{e}\varkappa(1-a) - b(a) - \frac{F}{q}\right) + \mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}(p - t_n)\right)\psi_{\varkappa}(\varkappa)\,d\varkappa \tag{55}$$

Using (53), we can write (51) as

$$t_n(\tau) = \tau E_{\widetilde{\psi}_e}[e|e > \widehat{e}_{n-1}(q,0)] + o(\tau). \tag{56}$$

Taking a first order Taylor approximation of (54), using (56) and recalling that F is second order in τ, we can write:

$$\tau E_{\widetilde{\psi}_e}[e|e > \widehat{e}_{n-1}(q,0)] - \tau\widehat{e}_n(q,0) + o(\tau) = 0$$

$$\Rightarrow \widehat{e}_n(q,0) = E_{\widetilde{\psi}_e}[e|e > \widehat{e}_{n-1}(q,0)] + o(1),$$

so that at 0th order $\widehat{e}_n$ is independent of q. We can then write

$$\widehat{e}_n(0) = E_{\widetilde{\psi}_e}[e|e > \widehat{e}_{n-1}(0)] \text{ with } \widehat{e}_0 = \underline{e}, \tag{57}$$

with $\widehat{e}_n(q,0) = \widehat{e}_n(0) + o(1)$, where the function $o(1)$ depends on q. We can also then rewrite (56) as

$$t_n(\tau) = \tau E_{\widetilde{\psi}_e}[e|e > \widehat{e}_{n-1}(0)] + o(\tau) = \tau\widehat{e}_n(0) + o(\tau). \tag{58}$$

Welfare. Ignoring the possibility that any firm exit post drilling (given the focus on small τ), producer surplus under the algorithm can be written as:

$$PS_n^{\text{alg}}(\tau) = \int_{u,q,p,\mathfrak{e}} q\left(h_n^{\text{alg}}(q,p,\mathfrak{e},\tau) - pu\right)^+ d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}), \tag{59}$$

as firms make expected revenues $h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)$ per unit after drilling. We can then similarly express tax revenues and emissions as

$$T_n^{\text{alg}}(\tau) = \int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\begin{pmatrix}\mathbf{1}_{\mathfrak{e}\varkappa\leq\widehat{e}_n(q,\tau)}(1-a(\tau,e))\tau e\\+\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}t_n\end{pmatrix}\psi_{\varkappa}(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e});$$

$$G_n^{\text{alg}}(\tau) = \int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\begin{pmatrix}\mathbf{1}_{\mathfrak{e}\varkappa\leq\widehat{e}_n(q,\tau)}(1-a(\tau,e))\\+\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\end{pmatrix}\mathfrak{e}\varkappa\psi_{\varkappa}(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

Welfare under the algorithm is then given by

$$W_n^{\text{alg}} = I + CS + PS_n^{\text{alg}}(\tau) - vG_n^{\text{alg}}(\tau) + T_n^{\text{alg}}(\tau),$$

where income I and consumer surplus CS are constant given fixed prices. Similarly under laissez-faire, we get

$$W^{LF} = I + CS + PS^{LF} - vG^{LF}.$$

We focus here on the case $\tau = v$,[10] so that we can write

$$\begin{aligned} W_n^{\text{alg}} - W^{LF} &= PS_n^{\text{alg}}(\tau) - vG_n^{\text{alg}}(\tau) + T_n^{\text{alg}}(\tau) - PS^{LF} + vG^{LF} \qquad (60) \\ &= \int_{u,q,p,\mathfrak{e}} q\left(h_n^{\text{alg}}(q,p,\mathfrak{e},\tau) - pu\right)^+ d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) - PS^{LF} + vE_{\widetilde{\psi}_e}(e)\, S^{LF} \\ &\quad + \int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu} \int_{\varkappa} \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}_n(q,\tau)} \left(t_n - \tau\mathfrak{e}\varkappa\right)\psi_{\varkappa}(\varkappa)\, d\varkappa\right] d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}) \end{aligned}$$

Taylor expansion. As before, we use Taylor expansions to find an approximation to $W_n^{\text{alg}} - W^{LF}|_{\tau=v}$. Differentiating (55), we get:

$$\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial \tau} = \int_{\varkappa} \left(\mathbf{1}_{\mathfrak{e}\varkappa \leq \widehat{e}_n(q,\tau)} \left(-\mathfrak{e}\varkappa\left(1 - a(\tau,e)\right) - \frac{F'(\tau)}{q}\right) + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}_n(q,\tau)} \left(-\frac{dt_n}{d\tau}\right)\right)\psi_{\varkappa}(\varkappa)\, d\varkappa, \qquad (61)$$

where we used the envelope theorem to cancel terms in $\frac{\partial a}{\partial \tau}$ and equation (54) to cancel terms in $\frac{\partial \widehat{e}_n(q,\tau)}{\partial \tau}$. We then get

$$\begin{aligned} &\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial \tau}|_{\tau=0} \qquad (62) \\ = &-\int_{\varkappa}\left(\mathbf{1}_{\mathfrak{e}\varkappa \leq \widehat{e}_n(0)}\mathfrak{e}\varkappa + \mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}_n(0)} E_{\widetilde{\psi}_e}\left[e|e > \widehat{e}_{n-1}(0)\right]\right)\psi_{\varkappa}(\varkappa)\, d\varkappa \\ = &-\left[E|_{\psi_{\varkappa}}(\varepsilon(\mathfrak{e})) + \left(E_{\widetilde{\psi}_e}\left[e|e > \widehat{e}_{n-1}(0)\right] - E_{\widetilde{\psi}_e}\left[e|e > \widehat{e}_n(0)\right]\right)\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa > \widehat{e}_n(0)}\psi_{\varkappa}(\varkappa)\, d\varkappa\right], \end{aligned}$$

where, as before, we define

$$\varepsilon = e \text{ if } e \leq \widehat{e}_n(0) \text{ and } \varepsilon = E_{\widetilde{\psi}_e}\left[e|e > \widehat{e}_n(0)\right] \text{ if } e > \widehat{e}_n(0).$$

[10] In particular, as before, τ and v are both small.

We can then differentiate (60) to get:

$$
\begin{aligned}
&\frac{\partial\left(W_n^{\text{alg}}-W^{LF}\right)}{\partial\tau} \qquad (63)\\
= &\int_{u,q,p,\mathfrak{e}} q\mathbf{1}_{\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\geq u}\frac{\partial h_n^{\text{alg}}\left(q,p,\mathfrak{e},\tau\right)}{\partial\tau}d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right)+E_{\widetilde{\psi}_e}\left(e\right)S^{LF}\\
&+\int_{u,q,p,\mathfrak{e}} q\psi_u\left(\frac{h_n^{\text{alg}}\left(q,p,\mathfrak{e},\tau\right)}{p}\right)\frac{1}{p}\frac{\partial h_n^{\text{alg}}\left(q,p,\mathfrak{e},\tau\right)}{\partial\tau}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\left(t_n-\tau\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\Psi_{q,p,\mathfrak{e}}\left(q,p,\mathfrak{e}\right)\\
&+\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\left(-\frac{\partial\widehat{e}_n\left(q,\tau\right)}{\partial\tau}\right)\left(t_n-\tau\widehat{e}_n\left(q,\tau\right)\right)\psi_{\varkappa}\left(\frac{\widehat{e}_n\left(q,\tau\right)}{\mathfrak{e}}\right)\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right)\\
&+\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\left(\frac{dt_n}{d\tau}-\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}\left(u,q,p,\mathfrak{e}\right)
\end{aligned}
$$

Evaluating at 0, using (53), (58), (62) and that $E_{\widetilde{\psi}_e}\left(\varepsilon\right)=E_{\widetilde{\psi}_e}\left(e\right)$, we obtain:

$$
\begin{aligned}
&\frac{\partial\left(W_n^{\text{alg}}-W^{LF}\right)}{\partial\tau}|_{\tau=0}\\
= &-\Psi_u\left(1\right)\int_{q,\mathfrak{e}} q\left[\begin{array}{c}E|_{\psi_{\varkappa}}\left(\varepsilon\left(\mathfrak{e}\right)\right)\\+\left(E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_{n-1}\left(0\right)\right]-E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]\right)\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\end{array}\right]d\Psi_{q,\mathfrak{e}}\left(q,\mathfrak{e}\right)\\
&+E_{\widetilde{\psi}_e}\left(e\right)S^{LF}+\Psi_u\left(1\right)\int_{q,\mathfrak{e}} q\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\left(E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_{n-1}\left(0\right)\right]-\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\Psi_{q,\mathfrak{e}}\left(q,\mathfrak{e}\right)\\
= &\ 0.
\end{aligned}
$$

Therefore $W_n^{\text{alg}}=W^{LF}+o\left(\tau\right)$: intuitively, for a small social cost of carbon (v), the effect on production allocation of any policy commensurate with the cost is first order, leading to a second order effect on welfare. To get a second order development of the welfare effect of the algorithm, we differentiate (63) a second time and obtain:

$$\frac{\partial^2\left(W_n^{\text{alg}}-W^{LF}\right)}{\partial\tau^2}$$

$$= \int_{q,p,\mathfrak{e}} q\frac{1}{p}\psi_u\left(\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\right)\left(\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}\right)^2 d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$+\int_{u,q,p,\mathfrak{e}} q\mathbf{1}_{\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\geq u}\frac{\partial^2 h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau^2}d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})$$
$$+\int_{q,p,\mathfrak{e}} q\frac{\partial\left[\psi_u\left(\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\right)\frac{1}{p}\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}\right]}{\partial\tau}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}(t_n-\tau\mathfrak{e}\varkappa)\psi_\varkappa(\varkappa)\,d\varkappa d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$-2\int_{q,p,\mathfrak{e}}\frac{q}{p}\psi_u\left(\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\right)\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}\frac{\partial\widehat{e}_n(q,\tau)}{\partial\tau}(t_n-\tau\widehat{e}_n(q,\tau))\,\psi_\varkappa\left(\frac{\widehat{e}_n(q,\tau)}{\mathfrak{e}}\right)d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$+2\int_{q,p,\mathfrak{e}}\frac{q}{p}\psi_u\left(\frac{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{p}\right)\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\left(\frac{dt_n}{d\tau}-\mathfrak{e}\varkappa\right)\psi_\varkappa(\varkappa)\,d\varkappa d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$-\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\frac{\partial\left(\frac{\partial\widehat{e}_n(q,\tau)}{\partial\tau}\psi_\varkappa\left(\frac{\widehat{e}_n(q,\tau)}{\mathfrak{e}}\right)\right)}{\partial\tau}(t_n-\tau\widehat{e}_n(q,\tau))\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})$$
$$-2\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\frac{\partial\widehat{e}_n(q,\tau)}{\partial\tau}\left(\frac{dt_n}{d\tau}-\widehat{e}_n(q,\tau)\right)\psi_\varkappa\left(\frac{\widehat{e}_n(q,\tau)}{\mathfrak{e}}\right)\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})$$
$$+\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)\geq pu}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\frac{d^2t_n}{d\tau^2}\psi_\varkappa(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

Evaluating this expression at $\tau=0$ and using (53) and (58), we get:

$$\frac{\partial^2\left(W_n^{\text{alg}}-W^{LF}\right)}{\partial\tau^2}|_{\tau=0} \tag{64}$$

$$= \int_{q,p,\mathfrak{e}}\frac{q}{p}\psi_u(1)\left(\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}|_{\tau=0}\right)^2 d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$+\int_{u,q,p,\mathfrak{e}} q\mathbf{1}_{1\geq u}\frac{\partial^2 h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau^2}|_{\tau=0}d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e})$$
$$+2\int_{q,p,\mathfrak{e}}\frac{q}{p}\psi_u(1)\frac{\partial h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau}|_{\tau=0}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}(\widehat{e}_n(0)-\mathfrak{e}\varkappa)\,\psi_\varkappa(\varkappa)\,d\varkappa d\Psi_{q,p,\mathfrak{e}}(q,p,\mathfrak{e})$$
$$+\int_{u,q,p,\mathfrak{e}} q\left[\mathbf{1}_{1\geq u}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\frac{d^2t_n}{d\tau^2}|_{\tau=0}\psi_\varkappa(\varkappa)\,d\varkappa\right]d\Psi_{u,q,p,\mathfrak{e}}(u,q,p,\mathfrak{e}).$$

To derive $\frac{\partial^2 h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau^2}|_{\tau=0}$, we differentiate (61):

$$\frac{\partial^2 h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau^2} \tag{65}$$

$$= \int_{\varkappa} \frac{\partial \widehat{e}_n(q,\tau)}{\partial\tau}\delta_{\mathfrak{e}\varkappa=\widehat{e}_n(q,\tau)}\left[\left(-\mathfrak{e}\varkappa(1-a(\tau,e)) - \frac{F'(\tau)}{q}\right) + \frac{dt_n}{d\tau}\right]\psi_\varkappa(\varkappa)\,d\varkappa \tag{66}$$

$$+\int_{\varkappa}\left(\mathbf{1}_{\mathfrak{e}\varkappa\le\widehat{e}_n(q,\tau)}\left(\mathfrak{e}\varkappa\frac{\partial a(\tau,e)}{\partial\tau} - \frac{F''(\tau)}{q}\right) + \mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,\tau)}\left(-\frac{d^2t_n}{d\tau^2}\right)\right)\psi_\varkappa(\varkappa)\,d\varkappa. \tag{67}$$

Evaluating this expression at $\tau=0$, using again (58), we get:

$$\frac{\partial^2 h_n^{\text{alg}}(q,p,\mathfrak{e},\tau)}{\partial\tau^2}|_{\tau=0} = \int_{\varkappa}\left(\mathbf{1}_{e\le\widehat{e}_n(0)}\left(\frac{e^2}{b''(0)} - \frac{F''(0)}{q}\right) + \mathbf{1}_{e>\widehat{e}_n(0)}\left(-\frac{d^2t_n}{d\tau^2}|_{\tau=0}\right)\right)\psi_\varkappa(\varkappa)\,d\varkappa. \tag{68}$$

Plugging this expression in (64), using (4) and (14), we get:

$$\frac{\partial^2\left(W_n^{\text{alg}} - W^{LF}\right)}{\partial\tau^2}|_{\tau=0} \tag{69}$$

$$= \dot{S^{LF}}\int_{\mathfrak{e}}\left(\frac{\partial h_n^{\text{alg}}}{\partial\tau}|_{\tau=0}\right)^2 d\widetilde{\Psi}_{\mathfrak{e}}(\mathfrak{e}) + 2\dot{S^{LF}}\int_{\mathfrak{e}}\frac{\partial h_n^{\text{alg}}}{\partial\tau}|_{\tau=0}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\left(\widehat{e}_n(0) - \mathfrak{e}\varkappa\right)\psi_\varkappa(\varkappa)\,d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}(\mathfrak{e})$$
$$+S^{LF}\widetilde{\Psi}\left(\widehat{e}_n(0)\right)\frac{E_{\widetilde{\psi}_e}\left[e^2|e<\widehat{e}_n(0)\right]}{b''(0)} - \Psi_u(1)\Psi\left(\widehat{e}_n(0)\right)F''(0).$$

We can then find a second order approximation of welfare under the algorithm (for $\tau = v$) as:

$$W_n^{\text{alg}} = W^{LF} + \frac{\tau^2}{2}\frac{\partial^2\left(W_n^{\text{alg}} - W^{LF}\right)}{\partial\tau^2}|_{\tau=0} + o\left(\tau^2\right). \tag{70}$$

Using (36) and (37), we get that at second order, welfare under the certification program (with $v=\tau$ and a revelation threshold set at $\widehat{e}_n(0)$) is given by:

$$W^V\left(\widehat{e}_n(0)\right) = W^{LF} + \frac{\tau^2}{2}\left[\begin{array}{c}\dot{S}^{LF}\left(E_{\widetilde{\psi}_e}\left[E|_{\psi_\varkappa}\left(\varepsilon(\mathfrak{e})\right)^2\right] + 2Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e} - E_{\psi_\varkappa}\left[\varepsilon(\mathfrak{e})\right], E_{\psi_\varkappa}\left[\varepsilon(\mathfrak{e})\right]\right)\right) \\ +\frac{S^{LF}E_{\widetilde{\psi}_e}\left(e^2|e<\widehat{e}_n(0)\right)\widetilde{\Psi}_e\left(\widehat{e}_n(0)\right)}{b''(0)} - \Psi_u(1)\Psi\left(\widehat{e}_n(0)\right)F''(0)\end{array}\right] + o\left(\tau^2\right). \tag{71}$$

Taking the difference between (70) and (71), and using (69), we obtain:

$$\begin{aligned}
& W_n^{\text{alg}} - W^V\left(\widehat{e}_n\left(0\right)\right) \\
= & \frac{\tau^2 \dot{S^{LF}}}{2}\left[\begin{array}{c}\int_{\mathfrak{e}}\left(\frac{\partial h_n^{\text{alg}}}{\partial\tau}|_{\tau=0}\right)^2 d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) + 2\int_{\mathfrak{e}}\frac{\partial h_n^{\text{alg}}}{\partial\tau}|_{\tau=0}\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\left(\widehat{e}_n\left(0\right)-\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
-\left(E_{\widetilde{\psi}_e}\left[E|_{\psi_{\varkappa}}\left(\varepsilon\left(\mathfrak{e}\right)\right)^2\right] + 2Cov_{\widetilde{\psi}_{\mathfrak{e}}}\left(\mathfrak{e}-E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right], E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right]\right)\right)\end{array}\right] + o\left(\tau^2\right)
\end{aligned}$$

Plugging in (62) and using (34) and (57) we obtain:

$$\begin{aligned}
& W_n^{\text{alg}} - W^V\left(\widehat{e}_n\left(0\right)\right) \\
= & \frac{\tau^2 \dot{S^{LF}}}{2}\times \\
& \left[\begin{array}{c}
2\left(\widehat{e}_n\left(0\right)-E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]\right)\int_{\mathfrak{e}}E|_{\psi_{\varkappa}}\left(\varepsilon\left(\mathfrak{e}\right)\right)\left[\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
+\left(\widehat{e}_n\left(0\right)-E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]\right)^2\int_{\mathfrak{e}}\left[\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right]^2 d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
-2\int_{\mathfrak{e}}E|_{\psi_{\varkappa}}\left(\varepsilon\left(\mathfrak{e}\right)\right)\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\left(\widehat{e}_n\left(0\right)-\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
-2\left(\widehat{e}_n\left(0\right)-E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]\right)\int_{\mathfrak{e}}\left(\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right)\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}\left(\widehat{e}_n\left(0\right)-\mathfrak{e}\varkappa\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
-2\int_{\mathfrak{e}}E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right]\left(\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(0)}e\psi_{\varkappa}\left(\varkappa\right)d\varkappa\right)d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right) \\
+2E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]\int_{\mathfrak{e}}E_{\psi_{\varkappa}}\left[\varepsilon\left(\mathfrak{e}\right)\right]\int_{\varkappa}\mathbf{1}_{\mathfrak{e}\varkappa>\widehat{e}_n(q,0)}\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right)+o\left(1\right)
\end{array}\right].
\end{aligned}$$

Simplifying further, we get:

$$\begin{aligned}
& W_n^{\text{alg}} - W^V\left(\widehat{e}_n\left(0\right)\right) \\
= & -\tau^2 \dot{S^{LF}}\left(E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]-\widehat{e}_n\left(0\right)\right) \\
& \times\left[\int_{\mathfrak{e}}\left(1-\Psi_{\varkappa}\left(\frac{\widehat{e}_n\left(0\right)}{\mathfrak{e}}\right)\right)\int_{\varkappa}\mathbf{1}_{e>\widehat{e}_n(0)}\left(e-\frac{\widehat{e}_n\left(0\right)+E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]}{2}\right)\psi_{\varkappa}\left(\varkappa\right)d\varkappa d\widetilde{\Psi}_{\mathfrak{e}}\left(\mathfrak{e}\right)\right] \\
& +o\left(\tau^2\right).
\end{aligned}$$

This expression reflects the welfare loss from implementing the certification algorithm where uncertified firms are undertaxed compared to the certified equilibrium where the output tax of uncertified firms immediately adjusts to the certification threshold: Under the algorithm, uncertified firms are taxed as if their emission rate were $\widehat{e}_n\left(0\right)$ instead of $E_{\widetilde{\psi}_e}\left[e|e>\widehat{e}_n\left(0\right)\right]$.

In the data, we can then compute the law of motion for $\widehat{e}_n$ as

$$\widehat{e}_{n+1}(0) = \frac{\sum_{i|e>\widehat{e}_n(0)} e_i q_i}{\sum_{i|e>\widehat{e}_n(0)} q_i},$$

The welfare change relative to certification can then be expressed as:

$$\begin{aligned}
& W_n^{\text{alg}} - W^V\left(\widehat{e}_n(0)\right) \\
= \; & -v^2 \dot{S}^{LF}\left(\widehat{e}_{n+1}(0) - \widehat{e}_n(0)\right) \\
& \times \left[\sum_i \omega_i \left(1 - \Psi_{\varkappa}\left(\frac{\widehat{e}_n(0)}{\mathfrak{e}_i}\right)\right) E_{\psi_{\varkappa}}\left(\mathbf{1}_{\mathfrak{e}_i \varkappa > \widehat{e}_n(0)}\left(\mathfrak{e}_i \varkappa - \frac{\widehat{e}_n(0) + \widehat{e}_{n+1}(0)}{2}\right)\right)\right],
\end{aligned}$$

where the distribution of $\mathfrak{e}_i$ and the distribution $\Psi_{\varkappa}$ are still the same as those derived in Section B.5.

C Theory Appendix: International

We consider the international setting. Individuals in Home and Foreign have potentially different preferences that can be represented by the utility of a representative agent in each country over aggregate consumption as:

$$U_H = C_{0,H} + u_H(C_H) - \alpha v G \tag{1}$$

$$U_F = C_{0,F} + u_F(C_F) - (1-\alpha)\, vG, \tag{2}$$

where $\alpha \in (0,1)$ denotes the population weight of H, so that v is the global social cost of carbon. 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 assume that labor endowments are sufficiently large so that the outside sector is active in both countries. Trading the outside good is free and we normalize its price to 1. As a result, wages are equal to 1 in both countries (though this is without loss of generality). It costs κ units of the outside good to ship the polluting good.

In this Appendix, we allow for Home production of the polluting good. There is a mass 1 of Home producers with a supply function s_H and a c.d.f. of emission rates Ψ_H. We assume that all emissions in Home are observable and taxed at τ_H. As a result, 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. World market clearing for the polluting good in that case can be written as

$$\begin{aligned} D_H(p_H) + D_F(p_F) \quad &= \quad E_H[s_H(p_H - \tau_H e + A_H)] \\ &\quad + \Psi_F(\hat{e}) E_F\left[s_F(p_H - \tau_F e - \kappa + A_F) | e < \hat{e}\right] + (1 - \Psi_F(\hat{e})) s_F(p_F). \end{aligned} \tag{3}$$

Since our focus will be on the case where τ_F, τ_H are small, we ignore the possibility of exit in this section (this is without loss of generality when there is no Home production since there is always demand for uncertified Foreign firms from their domestic market).

In this Appendix, we first derive the price changes in the pooling equilibrium from an increase in certification, then prove Proposition 5, and finally discuss the optimal policy.

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

Lemma 3. *In the pooling equilibrium, the effect of rising $\hat{e}$ on p_H and p_F is given by:*

$$\frac{dp_H}{d\hat{e}} = \frac{\left(D'_F - (1-\Psi_F(\hat{e}))s'_F(p_F)\right)\tau_F \frac{\partial E_F(e|e>\hat{e})}{\partial \hat{e}} + \left[s_F(p_F+A_F) - s_F(p_F)\right]\psi_F(\hat{e})}{\left\{D'_H(p_H) + D'_F(p_F) - \int_{\underline{e}}^{\hat{e}}\left[s'_F(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right] - (1-\Psi_F(\hat{e}))s'_F(p_F)\right\}}, \tag{4}$$

$$\frac{dp_F}{d\hat{e}} = \frac{-D'_H(p_H) + \int_{\underline{e}}^{\hat{e}}\left[s'_F(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right]\tau_F \frac{\partial E_F(e|e>\hat{e})}{\partial \hat{e}} + \left[s_F(p_F+A_F) - s_F(p_F)\right]\psi_F(\hat{e})}{\left\{D'_H(p_H) + D'_F(p_F) - \int_{\underline{e}}^{\hat{e}}\left[s'_F(p_H - \tau_F e - \kappa + A_F)\psi_F(e)de\right] - (1-\Psi_F(\hat{e}))s'_F(p_F)\right\}} < 0, \tag{5}$$

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 $\rho = 0$, so that (18) leads to:

$$p_F = (p_H - \tau_F E_F(e|e > \hat{e}) - \kappa). \tag{6}$$

Plugging this expression into equation (3) and differentiating with respect to $\hat{e}$ gives (4) from which one can get (5). The sign of (5) 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 5

We seek to establish Proposition 5 which gives the change in world welfare from introducing our certification mechanism when the default is an output-based tariff. With Home production, equation (22) generalizes to:

$$\begin{aligned} & W^V - W^U \\ = \; & \underbrace{s'_F\tau_F\left(v - \frac{\tau_F}{2}\right)Var_F(\varepsilon)}_{\text{Reallocation Effect}} + \underbrace{\tau_F\left(v - \frac{\tau_F}{2}\right)\frac{s_F\Psi_F(\hat{e})}{b''(0)}}_{\text{Abatement Effect}} \underbrace{- \left[(v-\tau_H)E_H(e)s'_H + (v-\tau_F)E_F(e)s'_F\right]\Delta p_H}_{\text{Price Effect on Untaxed Emissions}} \\ & \underbrace{-\frac{D'_F}{2}\left(\tau_F\left(E_F(e) + E_F(e|e>\hat{e})\right) - \rho\right)\Delta p_F}_{\text{Consumption Leakage Effect}} \underbrace{-\rho s'_F\left(1 - \Psi^F(\hat{e})\right)\left(\frac{\Delta p_H + \rho}{2} + (v-\tau_F)E(e|e>\hat{e})\right)}_{\text{Backfilling Effect}} \\ & \underbrace{-F\Psi_F(\hat{e})}_{\text{Cost of Certification}} + o(\tau^2), \end{aligned} \tag{7}$$

with the price changes given by

$$\Delta p_H = \underbrace{\frac{-\epsilon_F^D \theta_F^D}{\epsilon^S - \epsilon^D} \tau_F \left(E_F(e|e > \hat{e}) - E_F(e)\right)}_{>0} \underbrace{- \rho \frac{\left(1 - \Psi_F(\hat{e})\right) \epsilon_F^S \theta_F^S - \epsilon_F^D \theta_F^D}{\epsilon^S - \epsilon^D}}_{\leq 0} + o\left(\tau\right), \text{ and} \tag{8}$$

$$\Delta p_F = \underbrace{\frac{\epsilon_H^D \theta_H^D - \epsilon^S}{\epsilon^S - \epsilon^D} \tau_F \left(E_F(e|e > \hat{e}) - E_F(e)\right)}_{<0} + \underbrace{\rho \frac{\epsilon^S - \epsilon_F^S \theta_F^S (1 - \Psi_F) - \epsilon_H^D \theta_D^H}{\epsilon^S - \epsilon^D}}_{\geq 0} + o(\tau). \tag{9}$$

The change in emissions in Foreign remains given by (27) at first order, while the change in emissions at Home is given by:

$$G_H^V - G_H^U = E_H\left(e\right) s_H^{'} \Delta p_H + o\left(\tau\right). \tag{10}$$

We proceed in three steps: We 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 (1) and (2), we can write W^V(up to a constant equal to world exogenous labor income) as:

$$W^V = \underbrace{CS_H + CS_F}_{\text{consumer surpluses}} + \underbrace{PS_H + PS_F}_{\text{producer surpluses}} \underbrace{- \left[\left(v - \tau_H\right) G_H + \left(v - \tau_F\right) G_F + \tau_F G_{F,F}\right]}_{\text{non-internalized emissions}} - F \Psi_F\left(\hat{e}\right). \tag{11}$$

Consumer and producer surpluses are given by:

$$CS_H = u_H\left(C_H\right) - p_H C_H \text{ and } CS_F = u_F\left(C_F\right) - p_F C_F, \tag{12}$$

$$PS_H = \int_{\underline{e}}^{\infty} \left(p_H - \tau_H e + A_H\right) s_H\left(p_H - \tau_H e + A_H\right) \psi_H\left(e\right) de \tag{13}$$
$$- \int_{\underline{e}}^{\infty} c_H\left(s_H\left(p_H - \tau_H e + A_H\right)\right) \psi_H\left(e\right) de,$$

$$\begin{aligned} PS_F &= \int_{\underline{e}}^{\hat{e}} \left(p_H - \tau_F e + A_F - \kappa\right) s_F \left(p_H - \tau_F e + A_F - \kappa\right) \psi_F\left(e\right) de \\ &\quad - \int_{\underline{e}}^{\hat{e}} c_F\left(s_F\left(p_H - \tau_F e + A_F - \kappa\right)\right) \psi_F\left(e\right) de + \left(p_F s_F\left(p_F\right) - c_F\left(s_F\left(p_F\right)\right)\right)\left(1 - \Psi_F\left(\hat{e}\right)\right), \end{aligned} \tag{14}$$

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:

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

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

The Foreign emissions for domestic consumption, $G_{F,F}$, are given by:

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

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 $\tau_F E_F\left(e|e > \widehat{e}\right)$, 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 4.3.[1]

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.

[1] The different equilibrium possibilities are as follows. Foreign certified firms always export to Home. Some uncertified Foreign firms export to Home and Home firms do not export to Foreign when $\rho = 0 \Leftrightarrow p_H - p_F = \tau_F E_F(e|e > \hat{e}) + \kappa$. Neither uncertified Foreign firms nor Home firms export when $\tau_F E_F(e|e > \hat{e}) - \kappa < p_H - p_F < \tau_F E_F(e|e > \hat{e}) + \kappa$. Uncertified Foreign firms do not export but at least some Home firms export to Foreign when $p_H - p_F \leq \tau_F E_F(e|e > \hat{e}) - \kappa$, which is the case that we mostly ignore here.

A first-order Taylor approximation of the global market clearing equation (3) gives

$$\begin{aligned} & D'_H(p_0)\left(p_H^V - p_0\right) + D'_F(p_0-\kappa)\left(p_F^V - p_0 + \kappa\right) \qquad (15)\\ = \; & s'_H(p_0)\left(p_H^V - \tau_H E_H(e) - p_0\right) + s'_F(p_0-\kappa) \\ & + \left(\Psi_F(\hat{e})\left(p_H^V - \tau_F E_F(e|e<\hat{e})\right) + (1-\Psi_F(\hat{e}))\left(p_F^V+\kappa\right) - p_0\right) + o(\tau), \end{aligned}$$

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

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

$$\begin{aligned} & D'_H(p_0)\left(p_H^U - p_0\right) + D'_F(p_0-\kappa)\left(p_F^U - p_0 + \kappa\right) \qquad (16)\\ = \; & s'_H(p_0)\left(p_H^U - \tau_H E_H(e) - p_0\right) + s'_F(p_0-\kappa)\left(p_F^U - p_0 + \kappa\right) + o(\tau), \end{aligned}$$

and

$$p_F^U = p_H^U - \kappa - \tau_F E_F(e)\,. \qquad (17)$$

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

$$p_H^U - p_0 = \frac{s'_H(p_0)\,\tau_H E_H(e) + s'_F(p_0-\kappa)\,\tau_F E_F(e) - D'_F(p_0-\kappa)\,\tau_F E_F(e)}{s'_H(p_0) + s'_F(p_0-\kappa) - D'_H(p_0) - D'_F(p_0-\kappa)} + o(\tau)\,. \qquad (18)$$

Separating equilibrium. In the setting with certification, we first focus on a separating equilibrium. At order 0, (21) implies that:

$$s_F(p_0-\kappa)(1-\Psi_F(\hat{e})) = D_F(p_0-\kappa) + o(1), \qquad (19)$$

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

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

If $\hat{e}$ is such that equation (19) 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 that $\hat{e}$ 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 (22) in the main text. This is correct, though a more stringent adherence to convention would have us evaluate at expressions at$\hat{e}_0$ for the separating equilibrium.

From the certification condition (20), one gets that at first order:

$$p_H - \tau_F \hat{e} - p_F = \kappa + \frac{F+f}{s_F\left(p_0-\kappa\right)} + o\left(\tau\right), \tag{20}$$

where 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 (26).

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 $\tau_F E_F\left(e|e>\widehat{e}_F\right)$ (i.e. $p_H > p_F + \tau_F E_F\left(e|e>\widehat{e}_F\right) - \kappa$).

We use equation (20), label p_H and p_F, p_H^V and p_F^V respectively, and substitute for p_F^V in equation (15) to get:

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

and combine this with equation (18) to get

$$p_H^V - p_H^U = \frac{\left(\begin{array}{c} \left(1-\Psi_F\left(\widehat{e}\right)\right)s_F^{'}\left(p_0-\kappa\right)\left(\left(\tau_F\widehat{e}+\frac{F+f}{s_F(p_0-\kappa)}\right)-\tau_F E_F\left(e|e>\widehat{e}\right)\right) \\ -D_F^{\prime}\left(p_0-\kappa\right)\left(\tau_F\widehat{e}+\frac{F+f}{s(p_0-\kappa)}-\tau_F E_F\left(e\right)\right)\end{array}\right)}{s_H^{\prime}\left(p_0\right)+s_H^{\prime}\left(p_0-\kappa\right)-D_H^{\prime}\left(p_0\right)-D_F^{\prime}\left(p_0-\kappa\right)} + o\left(\tau\right), \tag{21}$$

and similarly:

$$p_F^V - p_F^U = \frac{\left(\begin{array}{c} s_H^{'}\left(p_0\right)\left(\tau_F E_F\left(e\right)-\left(\tau_F\hat{e}+\frac{f}{s_F(p_0)}\right)\right) \\ +s_F^{'}\left(p_0-\kappa\right)\left(\tau_F E_F\left(e\right)-\left(1-\Psi_F\left(\hat{e}\right)\right)\tau_F E_F\left(e|e>\hat{e}\right)-\Psi_F\left(\hat{e}\right)\left(\tau_F\hat{e}+\frac{f}{s_F(p_0-\kappa)}\right)\right) \\ -\left(\tau_F E_F\left(e\right)-\left(\tau_F\hat{e}+\frac{f}{s^F(p_0-\kappa)}\right)\right)D_H^{\prime}\left(p_0\right)\end{array}\right)}{s_H^{'}\left(p_0\right)+s_F^{'}\left(p_0-\kappa\right)-D_H^{\prime}\left(p_0\right)-D_F^{\prime}\left(p_0-\kappa\right)} + o(\tau). \tag{22}$$

We define ϵ_F^D as the elasticity of demand wrt. prices in Foreign as $\epsilon_F^D = D_F^{\prime}(p_0-\kappa)(p_0-\kappa)/D_F(p_0-\kappa) = D_F^{\prime}(p_0-\kappa)p_0/D_F(p_0-\kappa)+o(\tau)$ 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:

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

Using (26), we obtain (8). Using (17) and (18), we get (9). Equations (23) and (24) then follow from removing Home production.

Pooling equilibrium. In the pooling equilibrium (20) 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 (6) holds, in (15) 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 (18) 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 (8) still holds in the pooling equilibrium. Using that $\rho = 0$, we then further get that (9) also holds in the pooling equilibrium.

C.2.3 Taylor approximations for welfare and emission changes

First, we sum the changes in consumer and producer surplus (from (12), (13) and (14)) to get:

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

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

We handle each of these terms in order:

$$
\begin{aligned}
& \text{price effects} \qquad (23) \\
= \; & -D_F\left(p_0 - \kappa\right)\left(p_F^V - p_H^V + \tau_F E_F(e) + \kappa\right) \\
& + \left(p_H^V - p_H^U\right)\left[\begin{array}{c} s_H'\left(p_0\right)\left(\frac{p_H^V + p_H^U}{2} - \tau_H E_H(e) - p_0\right) \\ + s^{F\prime}\left(p_0 - \kappa\right)\left(\frac{p_H^V + p_H^U}{2} - \tau_F E_F(e) - p_0\right) - D_H'\left(p_0\right)\left(\frac{p_H^V + p_H^U}{2} - p_0\right) \end{array}\right] \\
& - D_F'\left(p_0 - \kappa\right)\left(p_F^V - p_F^U\right)\left(\frac{p_F^V + p_F^U}{2} - p_0 + \kappa\right) + o\left(\tau^2\right).
\end{aligned}
$$

Summing up (15) and (16) gives:

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

Plugging this expression in (23) delivers:

$$\begin{aligned}
& \textbf{price effects} \qquad (24)\\
= & -D_F(p_0-\kappa)\left(p_F^V-p_H^V+\tau_F E_F(e)+\kappa\right) \\
& +\left(p_H^V-p_H^U\right)s_F^{'}(p_0-\kappa)\frac{(1-\Psi_F(\hat{e}))\left(p_H^V-p_F^V-\kappa-\tau_F E_F\left(e|e>\widehat{e^F}\right)\right)}{2} \\
& +D_F^{'}(p_0-\kappa)\left(p_H^V-p_F^V-\tau_F E_F(e)-\kappa\right)\left(\frac{p_F^V+p_F^U}{2}-p_0+\kappa\right)+o\left(\tau^2\right)
\end{aligned}$$

We continue with the reallocation term:

$$\textbf{reallocation term} = \frac{s_F'(p_0-\kappa)}{2}(\tau_F)^2 Var_F(\varepsilon)+o\left(\tau^2\right), \qquad (25)$$

the abatement term:

$$\textbf{abatement term} = \Psi_F(\hat{e})\, s_F(p_0-\kappa)\, A_F+o\left(\tau^2\right), \qquad (26)$$

and the adjustment term:

$$\begin{aligned}
& \textbf{adjustment term} \qquad (27)\\
= & (1-\Psi_F(\hat{e}))\left(p_F^V+\kappa-p_H^V+\tau_F E(e|e>\hat{e})\right)\left[\begin{array}{c} s_F(p_0-\kappa) \\ +s_F^{'}(p_0-\kappa)\left[\frac{p_F^V+\kappa+p_H^V-\tau_F E_F(e|e>\hat{e})}{2}-p_0\right]\end{array}\right]+o\left(\tau^2\right).
\end{aligned}$$

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 (10) 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 (17) and (18) then gives the first order change in Foreign emissions as (27), which again is sufficient to compute the second order change in welfare.

We note that we can write:

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

Plugging this term, (24), (25), (26) and (27) in (11), and using the definition of ρ, we obtain the welfare change from certification as:

$$\begin{aligned}
& W^V - W^U \qquad (28) \\
= \; & \left(1-\Psi_F\left(\hat{e}\right)\right)s_F^{'}\left(p_0-\kappa\right)\rho\left(\frac{p_F^V+\kappa+p_H^U-\tau_F E_F(e|e>\hat{e})}{2} - p_0\right) \\
& +D_F^{'}\left(p_0-\kappa\right)\left(-\rho\left(\frac{p_F^V+p_F^U}{2} - p_0 + \kappa\right) + \tau_F\frac{E_F\left(e|e>\hat{e}\right)+E_F\left(e\right)}{2}\left(p_F^U - p_F^V\right)\right) \\
& +\rho\,\left[\left(1-\Psi_F\left(\hat{e}\right)\right)s_F\left(p_0-\kappa\right) - D_F\left(p_0-\kappa\right)\right] \\
& +\frac{s_F^{\prime}\left(p_0-\kappa\right)}{2}\left(\tau_F\right)^2 Var_F\left(\varepsilon\right) + \Psi_F\left(\hat{e}\right)s_F\left(p_0-\kappa\right)A_F + o\left(\tau^2\right) \\
& -\left(v-\tau_H\right)\left(G_H^V - G_H^U\right) - \left(v-\tau_F\right)\left(G_F^V - G_F^U\right) - F\Psi_F\left(\hat{e}\right)
\end{aligned}$$

Taking a first-order expansion of the market-clearing equation in Foreign in the separating equilibrium (21), 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 (28) and using (18), we get:

$$\begin{aligned} & W^V - W^U \\ = & -\left(1-\Psi_F\left(\hat{e}\right)\right)s_F^{\prime}\left(p_0-\kappa\right)\rho\left(\ \tfrac{\rho+\Delta p_H}{2}\ \right) \\ & -\frac{D_F^{\prime}\left(p_0-\kappa\right)}{2}\Delta p_F\left(\tau_F\left(E_F\left(e|e>\hat{e}\right)+E_F\left(e\right)\right)-\rho\right) \\ & +\frac{s_F^{\prime}\left(p_0-\kappa\right)}{2}\left(\tau_F\right)^2 Var_F\left(\varepsilon\right)+\Psi_F\left(\hat{e}\right)s_F\left(p_0-\kappa\right)A_F+o\left(\tau^2\right) \\ & -\left(v-\tau_H\right)\left(G_H^V-G_H^U\right)-\left(v-\tau_F\right)\left(G_F^V-G_F^U\right)-F\Psi_F\left(\hat{e}\right). \end{aligned} \tag{29}$$

Plugging in the expressions for the change in emissions (10) and (27) 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 (7).

C.2.4 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:

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

This is zero in the pooling equilibrium 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 (18) and (17), we get

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

Next using the expression for Δp_F (equation 22), we get:

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

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.

C.3 Optimal policy

In the following, we look at the optimal unilateral policy for a Home policy maker who maximizes world welfare. We consider a Home policy maker that can implement any allocation at home, but is restricted in Foreign: She can only use trade taxes, an emission tax on certified firms, and tax/subsidy on certification f. That is, the Home policy maker may offer a price $p_E^R - \tau_F(e - a)$ to a Foreign firm which reveals its e, exports and undertakes abatement, and a price p_E^U to exporters who do not reveal. 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:

$$D\left(p^F\right) = s^F\left(p^F\right)\left(1 - \Psi^F\left(\hat{e}^F\right)\right) - M, \tag{30}$$

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

$$
\begin{aligned}
& W \\
= \; & u_H \left(\int_0^{\infty} q_H(e)\,\psi_H(e)\,de + \int_0^{\hat{e}} s_F\left(p_E^R - \tau_F(e-a_F) - b_F(a_F) - \kappa\right)\psi_F(e)\,de + M \right) \\
& - \int_0^{\infty} \left(c_H(q_H(e)) + b_H(a_H(e))\,q_H(e)\right)\psi_H(e)\,de + u_F(D_F(p_F)) \\
& - \int_0^{\hat{e}} \left(c_F\left(s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right) + \left(b_F(a_F)+\kappa\right)s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right)\psi_F(e)\,de \\
& - c_F(s_F(p_F))\left(1-\Psi_F(\hat{e})\right) - \kappa|M| - \int_0^{\infty}(e - a_H(e))\,q_H(e)\,\psi_H(e)\,de \\
& - v\left(\int_0^{\hat{e}} (e-a_F)\,s_F\left(p_E^R - \tau_F(e-a_F) - b_F(a_F) - \kappa\right)\psi_F(e)\,de + E_F(e|e>\hat{e})\,s_F(p_F)\left(1-\Psi_F(\hat{e})\right) \right) \\
& - F\Psi_F(\hat{e}),
\end{aligned}
\tag{31}
$$

with the constraint that the Foreign market clears (30). We solve the corresponding Lagrange problem and let λ be the Lagrange multiplier on (30).[2] 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 $p_H = u'_H(C_H)$.

$$
\text{wrt. } p_E^R : \int_0^{\hat{e}} \left(p_H - p_E^R + (\tau_F - v)(e - a_F)\right) s'_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\psi_F(e)\,de = 0, \tag{32}
$$

where as before $A_F \equiv \tau_F a_F - b_F(a_F)$,

$$
\text{wrt. } p_F\text{: } \begin{array}{c} D'_F(p_F)\,p_F - s'_F(p_F)\left(p_F + vE_F(e|e>\hat{e})\right)\left(1-\Psi_F(\hat{e})\right) \\ +\lambda\left(s'_F(p_F)\left(1-\Psi_F(\hat{e})\right) - D'_F(p_F)\right) \end{array} = 0, \tag{33}
$$

FOC wrt. to M:

$$
\lambda = p_H - \kappa \text{ if } M > 0; \text{ and } \lambda = p_H + \kappa \text{ if } M < 0, \tag{34}
$$

[2] 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 $p_E^U < p_F + \kappa$.

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:

$$\int_0^{\hat{e}} \left[\begin{array}{c} \left(p_E^R - p_H + (\nu - \tau_F)(e - a_F)\right) s_F'\left(p_E^R - \tau_F e + A_F - \kappa\right)(-(e-a_F)) \\ +\frac{da_F}{d\tau^F}\left(b_F'(a_F) - \nu\right)\left(s_F\left(p_E^R - \tau_F e + A_F - \kappa\right)\right) \end{array} \right] \psi_F(e)\, de = 0 \quad (35)$$

FOC wrt. $\hat{e}$

$$\begin{aligned} & p_H s_F\left(p_E^R - \tau^F \hat{e} + A_F - \kappa\right) - c_F\left(s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right)\right) \\ -&\left(b_F(a_F) + \kappa\right) s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right) + c_F(s_F(p_F)) + c_F(s_F(p_F)) \\ &-v\left((e - a_F) s_F\left(p_E^R - \tau_F \hat{e} + A_F - \kappa\right) - \hat{e} s_F(p_F)\right) - F - \lambda s_F(p_F) \quad = 0. \end{aligned} \quad (36)$$

Together equations (32) and (35) 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 (36) as

$$\pi_F\left(p_H - \tau_F(\hat{e} - a_F) - b_F(a_F) - \kappa\right) - F = (\lambda - v\hat{e}) s_F(p_F) - c_F(s_F(p_F)). \quad (37)$$

We then solve for the system of (33), (34) and (37) 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 (33), gives $p_F = p_H - t^* - \kappa$, with

$$t^* = v\frac{s_F'(p_F) E_F(e|e>\hat{e})(1-\Psi_F(\hat{e}))}{s_F'(p_F)(1-\Psi_F(\hat{e})) - D_F'(p_F)}, \quad (38)$$

so that uncertified Foreign firms face an output-based tariff given by t^*. We can then rewrite (37) as

$$\pi_F\left(p_H - \tau_F(\hat{e} - a_F) - b_F(a_F) - \kappa\right) - F - f^* = \pi_F(p_F), \quad (39)$$

$$\text{with } f^* = (t^* - v\hat{e}) s_F(p_F). \quad (40)$$

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 (33), gives $p_F = p_H - t^* + \kappa$, with t^* still defined by (38), that is Home firms receive an export subsidy given by t^* when they export to Foreign. This again delivers (39), so that Foreign firms still pay a certification tax given by (40).

Finally, in the separating case, $M = 0$. Then p_H, p_F and $\hat{e}$ are defined by Foreign market clearing (30) which becomes $D\left(p^F\right) = s^F\left(p^F\right)\left(1 - \Psi^F\left(\widehat{e}^F\right)\right)$, a Home market clearing equation, and the first order condition on $\hat{e}$, (37) which can be written as (39) with f^* given by (40). This allocation can be implemented with the trade tax t^* defined in (39) 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 $p_H - p_F + \kappa$ determines the type of allocation that is optimal. Bringing the three cases together, we obtain:

Proposition 8. *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* $f^* = (t^* - v\hat{e})\, s_F(p_F)$*.*

We compare the resulting welfare from solving problem (31) 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:

$$\begin{aligned} W^* - W^{LF} &= s_F'(p_0 - \kappa)\frac{v^2}{2}Var(\varepsilon) + \Psi_F(\hat{e})A_F s_F(p_0 - \kappa) - F\Psi_F(\hat{e}) \\ &\quad + D_F'(p_0 - \kappa)\frac{t^*}{2}E_F(e|e > \hat{e}) - s_F'(p_0 - \kappa)\frac{v}{2}E_F(e)\left(p_H^* - vE_F(e) - p_H^{LF}\right), \end{aligned}$$

where t^* obeys (38). This welfare expression holds in all three cases, though strictly speaking t^* only applies as a tax in the pooling equilibrium.

D International Empirical Application: Brazilian Steel Trade

To illustrate how to calibrate our model and the welfare implications of a voluntary certification program in the international setting, we now conduct an analysis based on our approximation formulas 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. In this Appendix, we present our calibration, followed by our main results, and some additional results.

D.1 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.[1] We use the welfare formula given in Proposition 5. This exercise requires a handful of key statistics and economic parameters of supply and demand. We use publically available data to calibrate the model.

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).[2] 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)).

[1] 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 largest exporter to the US after Canada.

[2] 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).

Table D.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)

Emissions rates. The production shares for each process in each country is directly available in the report of the World Steel Association (World Steel, 2020): the EAF shares are 22.2% in Brazil and 69.7% in the US. A study by Hasanbeigi and Springer (2019) reports mean emission rates by country for each process: In Brazil, the mean emission rates are 2.07 tCO_2/t for the more energy-intensive BF-BOF, and 0.46 for EAF . Corresponding numbers for the US are 1.82 and 0.62, respectively, implying that EAF producers are relatively clean in Brazil but the average plant for the whole industry is dirtier. We use the same data sources to estimate mean emission rates in the OECD.[3]

[3]More specifically, Hasanbeigi and Springer (2019) report the emission rates of the BF-BOF and EAF processes 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 D.1 reports the values.

There is also substantial variation in emissions rates within each process. We could not find direct data on this, so instead we calibrate the emission distributions as follows – recall, however, that should certification be implemented progressively, the social planner would not need to collect ex-ante data on the distribution of emission rates beyond the mean. We assume that for each process, the emission rate distribution is a double-bounded log-normal distribution, so that we need to calibrate 4 parameters: the two bounds, the mean μ and the standard-deviation σ of the unbounded distribution. We already have the average emission rate as 1 moment for each process. For a second moment, we assume that there is the same amount of heterogeneity for each process, and that once one controls for the process-type, most heterogeneity reflects differences in energy intensity and that those move with TFP differences. To get an estimate for the standard deviation of log productivity in the steel sector in Brazil, we use a value for the basic metal products sector from Schor (2004): 0.409.[4] Using US data, Collard-Wexler and De Loecker (2015) report a small productivity premium for the EAF process (of 0.074), we use that value for Brazil and the share of EAF production in Brazil (0.222) to get a common estimate for the standard deviation of log productivity within each process given by $(0.409^2 - 0.222 \cdot (1 - 0.222) \cdot 0.074^2)^{1/2}$.[5] Finally, we (arbitrarily) set the lower bound to be equal to 2/3 of the emission rate in the cleanest country as reported by Hasanbeigi and Springer (2019) (France for EAF and Canada for BOF) and the upper-bound 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 process. Figure D.1 shows the c.d.f. $\Psi_F(e)$ for the resulting overall distribution of emission rates in Brazil.

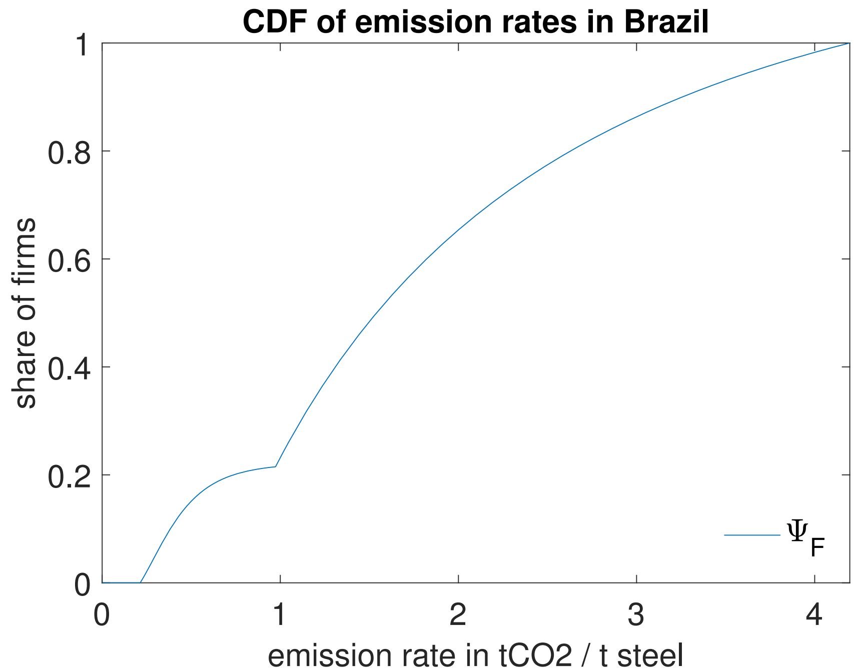
Social cost of carbon and abatement. We set the social cost of carbon at \$51 (Interagency Working Group on Social Cost of Greenhouse Gases, 2021). To parameterize abatement costs, we match abatement in our model, $a = \tau / b''(0)$, with abatement according to a marginal abatement cost curve for the 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.[6] This is conservative relative

[4]This implicitly assumes that in the basic metal sector, within-subsector heterogeneity dominates heterogeneity across sub-sectors. This assumption likely pushes us to overestimate the true amount of heterogeneity in emission rates.

[5]Evidence from Lyubich et al. (2018) shows that emissions rates tend to be more heterogeneous than TFP, so that our assumption that within-process heterogeneity is similar to TFP heterogeneity may underestimates the true amount of heterogeneity – nevertheless since the mean productivity premium of EAF is much lower than the mean gap in emission rates, the standard deviation of log emission rates of the joint distribution of steel firms is already larger than that of log TFP in our calibration.

[6]To adjust for inflation and technological progress, we assume that the ratio between the abatement cost curve and the price of steel in Brazil is constant (i.e. technology in abatement and in steel are "proportional"

Figure D.1: CDF of emission rates in Brazil



to McKinsey & Co. (2009), which predicts a marginal abatement cost of \$50 for 0.526 tCO_2/ t steel in 2030.

Certification cost. To estimate the certification cost F, we rely on Gallaher and Depro (2002) from the EPA who find that the annualized cost of monitoring hazardous air pollutants (manganese, lead, benzene, etc. but not CO_2) in the iron and steel sector was \$1.04M per plant in 2001 (their Table 3.5). Using the production of these plants, we get a certification cost of \$0.49 per ton in 2019. [7] We assume 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 (\$16 M).

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

to each other). We use data from Instituto Aço Brasil (2016 and 2021) to compute the average price of for steel produced in Brazil in 2010 and 2019. For simplicity, we ignore the distinction between EAF and BOF abatement technologies.

[7] These pollutants are generated by the BOF process and the monitoring costs are computed for 18 plants which produced with the BOF process. 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. We then estimate that these plants produced in 2001 53.4 Mt of steels with the BOF process. This gives us a cost of certification in 2001 \$ per ton of steel of $1.04 \times 18/53.4 = 0.35$, which we convert into 2019 values using the US GDP deflator.

Table D.2: Emission and Welfare Effects from Environmental Trade Policies

	First Best	CBA	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.

3.5, which is the supply elasticity used by EPA (2002). Table D.1 gives all parameters and their sources.

D.2 Main results

Table D.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 $vE_F(e)$= \$87 per ton) reduces steel production in Brazil and increases welfare by \$714M 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 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 versus \$137). As a result, without any certification tax, voluntary certification very slightly reduces welfare relative to an output-based tariff by \$6M—though this ceases to be the case if abatement costs were just 25% lower (see Section D.3).

Table D.3 decomposes the welfare change from the output-based tariff to the voluntary certification program into the different channels discussed above (Table D.4 below shows how

Table D.3: Decomposition of Welfare Effects from Voluntary Certification Relative to CBA

	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 CBA	-6	152

Note: This table decomposes the welfare changes from voluntary certification relative to an output-based tariff (CBA) 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.

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 $\$-228$M, while the "reallocation effect" leads to gains of \$150M.

A tax or other program to limit certification to the optimal level, however, brings \$152M 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. This leads to a "backfilling effect" of a much smaller magnitude at $\$-10$M, while the reallocation effect remains \$150M since the certification threshold does not move much with the introduction of the optimal certification tax.[8] With our baseline calibration, abatement gains are comparatively small at \$37M.[9] These gains would be substantially larger using the estimates of McKinsey & Co. (2009), so that even

[8] 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 D.2 below.

[9] The second best policy described in section 4.3 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.

an unrestricted voluntary certification program would deliver gains over a standard carbon border adjustment (CBA).

Finally, note that while the output-based tariff reduces emissions by 5.6Mt of CO_2, the optimal certification program reduces emissions by 11.1Mt 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 14.5Mt of CO_2.

To illustrate how the benefits of an opt-in emissions taxation program change with the extent of certification, Figure D.2 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) as long as the certification fee remains above 4% of revenues, and they only disappear for trivial certification taxes.[10]

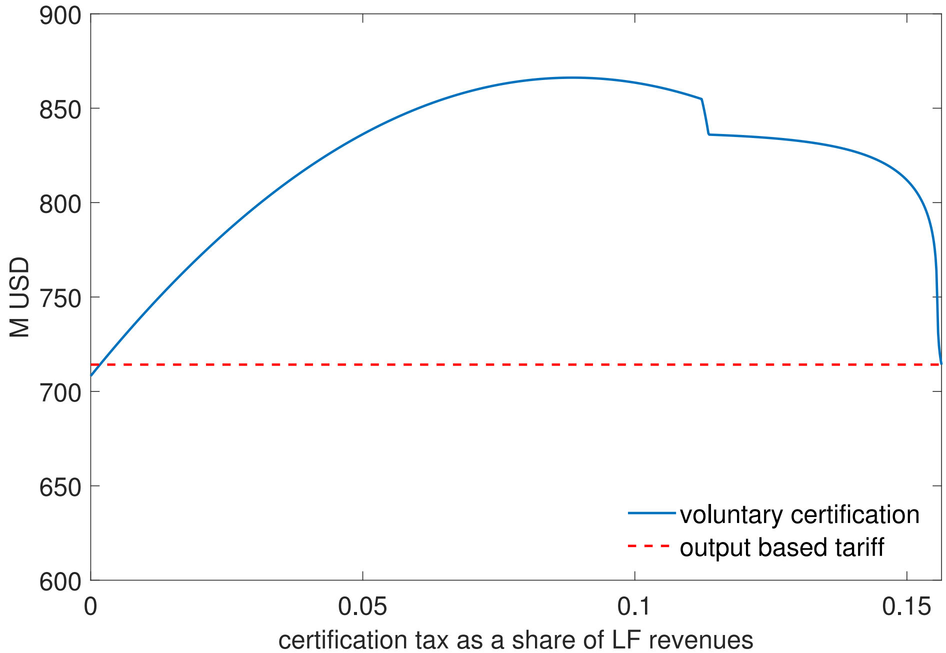
We stress that this exercise is a simple proof-of-concept and that our numbers are indicative of orders of magnitude but not exact values. It shows that there are potentially large emission reductions and significant welfare gains from such a certification program, though it may not be desirable to let certification occur without restriction if it leads to a large reduction in the price gap between the two countries. Fortunately, prevailing price gaps relative to border adjustment fees are observable. This makes it possible for policymakers to adjust certification criteria in order to avoid significant backfilling losses.

D.3 Additional results

We now assess the sensitivity of our results. Table D.4 reproduces the first row of Table D.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

[10]The small jump in Figure D.2 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 D.2: 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 CBA. The extent of certification $\Psi_F(\hat{e})$ rises as certification taxes fall

laissez-faire. The Table adds one column for the welfare gains under the optimal output-based tariff which differs from a carbon-border adjustment for the reason discussed in Section 4.3 – 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 D.4 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 without 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 carbon border adjustment. 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 $v = \$102$ 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 $\tau_F \neq v$ (implying that the certification tax is not set optimally). Correcting the social cost of carbon and implementing the true optimal CBA (with $\tau_F = v$) 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 D.4: Parameters and Sources for the Steel Numerical Example

	Carbon border 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.