Asymmetric information and costly participation in voluntary conservation markets*

Willard Robinson†
North Carolina State University

October 27, 2026
[Updated frequently; please click HERE for the latest version.]

Abstract

Emissions trading systems now cover roughly 30% of global greenhouse gas emissions, and offsets promise substantial gains from trade between capped sectors and cheaper abatement elsewhere. I extend van Benthem and Kerr's (2013) model of nature-based offset supply: landowners hold private information about the commercial value of their land and respond to a uniform baseline-price mechanism, but must also pay a privately known fixed cost to participate. Acting alone, these two friction reduce the social value an offset market creates, but that value remains non-negative. When combined, offset markets can be welfare-reducing, the point that closing it welfare-dominates an active market. The two frictions play different roles: landowners holding private information about their land's commercial value determines who participates in the market while the costly participation determines how much welfare they create. Willingness to participate is highest for those landowners who would not develop their land anyway, meaning the hardest landowners to exclude are precisely those who generate the least social value. Separating welfare-positive from welfare-negative landowners is possible by exploiting differences in their willingness to participate at different levels of policy stringency. Improving baseline accuracy, the focus of recent policy debate, is therefore necessary but not sufficient to make these markets work.

Keywords: Carbon offsets, transaction costs, additionality

JEL Classification: Q58, Q57, D82, D23, Q24

*I thank Spencer Banzhaf and Ben Leard for helpful feedback, as well as participants at Camp Resources XXXII and seminar participants at North Carolina State University for thoughtful discussion.

†Department of Economics, North Carolina State University; wmrobin3@ncsu.edu

1 Introduction

Carbon markets now cover roughly 30% of global greenhouse gas (GHG) emissions, generating revenue upwards of $107 billion in 2025 alone. This rise is driven largely by emissions trading systems, which tripled their coverage in the past decade (World Bank, 2026). Most of this activity is concentrated in the industrial sectors of high- and upper middle-income countries where average abatement costs are around $44 per ton. Substantial mitigation opportunities exist outside of this setting, for example in forestry, agriculture, other nature-based activities, and in lower-income countries where abatement costs are far lower, on the order of $10 per ton in Brazil (Araujo et al., 2026). Because GHGs mix globally, a ton of abatement in California is as beneficial as a ton of abatement in the Amazon. The offsets mechanism facilitates trade between these regions with vastly different abatement costs, generating gains from trade.

Whether these gains are real depends critically on their "additionality": that the credited reductions represent emissions reductions that truly would not have occurred otherwise. In 2023, a team of researchers intensely analyzed 18 projects expected to "sequester" 90.5 million tons of greenhouse gases (GHGs). According to their estimates, only 6.4 million (7%) of these would be associated with truly incentive-produced emissions reductions (West et al., 2023). At the time of their publication, 62 million of these credits had been issued, with about 24% of those having been used to offset real global emissions. In effect, the projects had been used to offset more than twice as many carbon emissions as they actually contributed. Each of these non-additional credits used to offset true GHG emissions reflects a real increase in global emissions that appears as a net-zero change on paper. This result is not necessarily caused by enforcement failure or fraud, but is a natural reaction by landowners responding to incentives.

The scale of the programs relying on this counterfactual is substantial. Offsets are central to the Clean Development Mechanism, which generated more than $90 billion of renewable energy investment in developing countries (Calel et al., 2025), and to California's cap-and-trade program, where 223 million of these credits have been used (Fowlie et al., 2024). They extend beyond GHGs to biodiversity and wetland conservation, creating $2.4 billion in value added from 1995 to 2020 (Aronoff & Rafey, 2026). Each of these programs awards credits against a predicted baseline.

While the counterfactual problem has received sustained attention, the frictions governing who shows up to the market have not. Suppliers of nature-based offsets must put forth funds for project design documentation, third-party validation, ongoing monitoring, and periodic

verification before receiving a single credit. Pearson et al. (2014) document these costs for tropical forest sequestration. The costs were made up largely of insurance, monitoring, and regulatory costs and can reach as much as 2.7 times project income. On the demand side, Jaraitė-Kazukauskė and Kazukauskas (2015) identify participation costs associated with trading activity in the EU ETS, and Naegele (2018) finds significant additional costs borne by firms adopting offsets in particular. In other work, I have demonstrated that such demand-side transaction costs in California's cap-and-trade program can be up to 35% of a firm's total compliance value (Robinson, 2026). Three features of these costs drive the results that follow. They are largely fixed at the project level rather than proportional to credits issued, so they determine whether a landowner participates rather than how much he conserves once enrolled.[1] They are heterogeneous and unobserved, so the regulator cannot mitigate them through a uniform adjustment to the price. And, crucially, they are a pure loss to the participating landowner: unlike the payment for non-additional conservation, which is a transfer from regulator to landowner, resources spent on registration and verification are absorbed by the certification process rather than passed to any party in this model. This third feature is what permits a voluntary participant to make a strictly negative contribution to social welfare. The interaction between transaction costs and asymmetric information can make an offset market welfare-reducing. As such, the choice to restrict offsets or limit their use in some capacity need not reflect concerns about additionality alone: when participation is costly and the regulator cannot perfectly distinguish projects, the welfare contribution of the projects attracted to the market can itself be negative. This paper provides a framework for understanding this possibility and for evaluating when restricting participation or closing the market is preferable to maintaining an open market.

Non-additional credits are costly through three distinct channels. First, when the purchasing entity faces a binding budget constraint, every dollar spent on a non-additional offset comes at the opportunity cost of a dollar spent on an additional offset. As such, welfare maximizing means choosing projects with the greatest return per dollar spent (i.e., maximizing dollar value of public funds, Hahn et al., 2026). This is the channel van Benthem and Kerr's (2013) (hereafter "vBK") cost-effectiveness objective captures. Second, when a regulated firm surrenders offsets against its cap, a non-additional credit licenses emissions with no corresponding reduction, and global emissions rise; non-additional offsets used to reduce attributed emissions of regulated firms increase global emissions. Third, offset supply requires a transaction cost (i.e., participation is costly). Even when neither of the first two channels operates, a non-additional supplier incurs registration and verification costs while

[1] I show this to be the case for offset demand in California in Robinson (2026), and that firms are generally persistent in their offset adoption: firms that use offsets once are much more likely to continue using them.

Figure 1: The dual role of asymmetric information and transaction costs

Asymmetric information? \ Transaction costs?	No	Yes
No	Non-negative welfare guaranteed	Non-negative welfare guaranteed
Yes	Non-negative welfare guaranteed	Welfare non-negativity not guaranteed

Notes: This figure illustrates the outcome of offset markets under 4 different scenarios. The asymmetric information and transaction costs are presented as binary for illustrative purposes. In a more realistic representation, these would be continuous distortions. Negative welfare can occur when the sum of participation costs overpowers any social benefit that the offset market provides.

producing no environmental benefit, so his participation is a pure deadweight loss.

To focus on the role of participation costs, I take the perspective of a government procuring land conservation from an unlimited budget (as in van Benthem & Kerr, 2013) rather than as credits surrendered against a cap. I show that, even with the first two channels switched off, the transaction costs can make an offset market welfare-reducing, meaning closing the market is the optimal policy. The assumptions are therefore conservative: relaxing either of the first two would strengthen the case for closure.

In this paper I develop a theoretical framework for offset markets with asymmetric information and costly participation. The paper's central conceptual contribution can be summarized in the two-by-two table in Figure 1. With neither asymmetric information nor transaction costs, the offset market is efficient and guaranteed not to reduce aggregate welfare. With asymmetric information alone, it remains optimal to open the market, as all non-additional offsets are simple monetary transfers, generating non-negative welfare. With transaction costs alone, the regulator has perfect information and can prohibit welfare-reducing offset suppliers from participating. Only when both are present simultaneously does the market become potentially welfare-negative.

The central point of this paper is that asymmetric information and costly participation play different roles in determining the welfare effects of an offset market. Asymmetric information determines which landowners are attracted to participation: because participa-

tion thresholds are lower for landowners with more commercially valuable land, the market disproportionately attracts landowners whose conservation is less likely to be additional. Participation costs determine the welfare contribution of those participants once they enter: unlike offset payments, which are transfers, the resources devoted to registering, monitoring, and verifying projects are socially costly. The interaction therefore matters for the sign of welfare, not merely its magnitude. When participation is costless, adverse selection can reduce the gains from an offset market but cannot make an otherwise appropriately priced market welfare-reducing. When participation is costly, however, the landowners most likely to participate can also be those whose participation generates negative social value. The resulting screening problem is the central object of the paper.

The basis for this screening is hinted at in Aspelund and Russo's (2026) analysis of the US's Conservation Reserve Program. They identify that, empirically, even the most non-additional landowners had a reservation price, below which they would not supply conservation, despite the fact that they would not have developed that land anyway. While Aspelund and Russo (2026) do not explicitly state this, the presence of that reservation price is consistent with an existing transaction cost, and it provides for the possibility of screening.

Three further results follow from the framework. First, the optimal baseline is generally interior, and its role is one of screening rather than transfer determination. Because a landowner's participation threshold is decreasing in the average return to his land, the baseline screens from the top: it excludes owners whose parcels are commercially valuable (thus socially beneficial to conserve), but provide too little conservation to be worthwhile. At the Pigouvian price, the optimal baseline is exactly the expected business-as-usual conservation of the marginal participant; away from that price it is adjusted by a wedge reflecting the gap between price and externality value. With heterogeneous participation costs the baseline ceases to be a binary type-selector and becomes an intensity controller, thinning participation within every type at once, and the optimality condition equates the average welfare contribution of the marginal crowd to zero. Second, the optimal price may lie strictly below the Pigouvian benchmark. In a two-type economy, the ordering of participation thresholds can switch as the price varies, because a landowner whose parcels are entirely non-additional receives only lump-sum payments and cannot adjust his opportunity cost, while a landowner with a mix of returns can. Pricing just below this crossing lets the regulator separate welfare-positive from welfare-negative types, and this separation can dominate a Pigouvian pooling policy despite inducing less conservation per participant. Third, when both frictions are severe, the value of the regulator's problem can be negative at every price and baseline, in

which case closing the market strictly dominates every active policy.

Many of the largest offset programs in the world have high and heterogeneous participation costs and imperfectly observable counterfactuals. As a result, much of the existing literature and policy discussion has focused on improving the identification of additionality through stricter enforcement, better monitoring, and increasingly sophisticated uses of satellite imagery and machine learning (Badgley et al., 2022; Kim & Sanford, 2026). These advances improve the regulator's ability to distinguish additional from non-additional, and go a long way toward improving market efficiency. However, not addressing the underlying participation incentives might leave other, less costly welfare improvements uncaptured.

1.1 Connection to the literature

Voluntary conservation framework This analysis builds on vBK. The main departure in this paper is that landowners must pay a fixed, privately known cost to enter the market. The distinction is not that a second friction compounds the first. In vBK, payments for non-additional conservation are transfers; non-additional transfers hurt the policy because they increase the average cost per unit of additional conservation, not because any resources are destroyed. Proposition 1(i) shows that in such an environment welfare is non-negative at any price at or below the externality value, whatever the informational structure, when there are no participation costs. Adverse selection can shrink the gains arbitrarily, but it cannot reverse their sign. In the current paper, the regulator is a planner with a similarly unlimited budget, the most favorable possible treatment. What breaks the result is that the participation cost is absorbed by the certification process rather than received by anyone. It essentially destroys resources.

Bento et al. (2015) show that restricting eligibility improves efficiency in an offset market operating alongside a regulated sector, subject to distributional constraints on who bears the cost. Two differences follow. Their regulator chooses the eligible set of offset providers whereas types in this paper are unobserved, and any restriction must be induced through the baseline-price mechanism. And their distributional constraint is a constraint on transfers; relaxing it makes their market unambiguously better, while relaxing the analogous constraint here changes nothing, since transfers are already neutral.

Relative to this literature, the present paper's contribution is to show that transaction costs can transform an adverse selection problem from one that reduces surplus into one that actively reduces welfare.

Asymmetric information in voluntary markets The adverse selection problem in offset markets is a result of asymmetric information. The information problem in environmental policy has been studied before.[2] Ferraro (2008) provides an initial mechanism design treatment of payments for ecosystem services under hidden provision cost, showing that first-best conservation is generally unattainable.[3] Mason and Plantinga (2013) characterize the additionality problem in a principal-agent model, and derive second-best offset supply contracts. Bento et al. (2015) study how distributional constraints affect the design of carbon offset markets which operate in conjunction with regulated markets, showing that targeting restrictions on who can participate can significantly improve the efficiency of the market's outcome. These papers all point to targeting participants as a way to improve welfare outcomes. This paper complements the literature by embedding transaction costs as another source of adverse selection, and providing potential adaptations policymakers can make.

Costly participation in voluntary markets Transaction costs in environmental markets have received some empirical attention, but limited theoretical treatment. Stavins (1995) provides the foundational analysis of how transaction costs reduce the efficiency gains from emissions trading systems. Baudry et al. (2021) and Singh and Weninger (2017) extend this to cap-and-trade settings explicitly, documenting how fixed and proportional transaction costs reduce participation, thereby distorting the optimal allocation of emissions across firms. Jaraitė-Kazukauskė and Kazukauskas (2015) identify participation costs associated with general trading activity in the EU ETS; Naegele (2018) extends their analysis and finds significant additional transaction costs associated with adopting offsets on top of the standard compliance permits. Pearson et al. (2014) document similar costs in the supply of offsets. What the existing literature has not established is that transaction costs can interact with asymmetric information to make a market welfare-reducing rather than merely less efficient. This paper provides that result.

Offsets are not naturally homogeneous commodities. Each offset differs in its location, baseline conditions, monitoring requirements, permanence risks, and likelihood of satisfying additionality criteria. As a result, offset markets rely heavily on certification, verification, and intermediation to transform these varied projects into tradeable credits: each of which entails transaction costs. Such frictions create an important role for regulators and market makers who match buyers and sellers. Whereas in an ETS, the regulator chooses a cap and the market determines the price, in an offset market the regulator chooses the rules

[2]This literature is, itself, related to a branch of research motivated by insurance markets. See Dionne and Doherty (1992) and Einav and Finkelstein (2011) for reviews of this sibling literature.

[3]Twelve years later, Wunder et al. (2020) wrote a review describing the success of these policies in practice.

determining who participates and how many credits they receive. These choices effectively shape the supply of offsets, giving the regulator greater influence over prices than in a conventional commodity market (like permits in cap-and-trade). Thus, the price of an offset is a result of certification rules and not just the marginal cost of abatement. This is the reason for the baseline-price mechanism.

2 Model setup

As in vBK, two types of agents participate in this model. These are (i) landowners with privately known economic returns to deforestation, and (ii) a regulator who offers the landowners a payment for each plot of land conserved beyond an assigned baseline. Landowners make plot-by-plot conservation decisions, and all participation in the market for offsets is voluntary.[4]

That economic returns are known only to the owner of a plot introduces adverse selection, a well-studied issue in the payments-for-ecosystem-services markets (Ferraro, 2008; Mason & Plantinga, 2013). The innovation in this paper—and the main departure from vBK—is pairing this with transaction costs. In addition to hidden information regarding deforestation profits, landowners incur a fixed cost to participate in the market.

2.1 Primitives

The set of **landowners** $\mathcal{I}$ is indexed by $i = 1, \ldots, I$. Each landowner i is endowed with N **parcels** $\mathcal{J}(i)$ indexed by j. Parcel j owned by i carries a privately known economic return to deforestation $r_{ij} \in \mathbb{R}$. Each parcel is either **developed** or **conserved**: if i develops parcel j, he deforests it permanently, earning r_{ij}, and generating no external benefit; if he conserves it, he earns nothing and generates an external benefit $\delta > 0$, constant across all parcels.

The regulator. I take the perspective of a government procuring conservation from an unlimited budget, purchasing offsets for their conservation value rather than surrendering them against an emissions cap. Payments from the regulator to landowners are therefore transfers that net out of aggregate welfare. This is the most favorable possible treatment of non-additionality: it switches off the channels through which non-additional credits are conventionally harmful, leaving only the participation costs.

[4] Though, for expository purposes, I sometimes describe what a landowner *would* contribute *if* he enrolled.

Participation costs. In addition to his returns, landowner i is characterized by a participation cost $\Omega_i \geq 0$ that he must pay to enter the market. It represents project design documentation, third-party validation, monitoring, verification, and any other idiosyncratic barriers such as reputational exposure. The cost is fixed at the landowner level, so it governs whether i participates and has no impact on how much he conserves once enrolled. It is unobserved by the regulator, and is absorbed by the certification process, meaning it provides a benefit to no party. As such, it acts as a pure loss to aggregate welfare.

Types. Returns are drawn from a common family parameterized by the average return of i's parcels, γ_i.

Assumption 1 (Log-concave deforestation returns)**.** *The private returns to deforestation of a plot r have a distribution $F_r(r;\gamma) = F\left(\frac{r-\gamma}{\sigma}\right)$ with σ common across landowners. The base density F is log-concave and continuous.*[5] *The distribution has an analogous probability density function $f_r(r;\gamma)$, also log-concave and continuous, which approaches zero as $r \to \pm\infty$.*

A landowner is fully described by the independent pair (γ_i, Ω_i), which is privately known only to i. The regulator does not observe either, but she knows the distributions for private deforestation returns (F_r), average deforestation returns (F_γ), and participation costs (F_Ω), so she makes policy based on expected values. Log-concavity nests the normal and uniform distributions and delivers the unimodality used throughout.

Business-as-usual. Absent any policy, i deforests parcel j if and only if $r_{ij} > 0$. His business-as-usual (BAU) conservation is thus

$$B_i = \sum_{j \in \mathcal{J}(i)} \mathbb{1}\{r_{ij} \leq 0\},$$

which the regulator cannot observe but can forecast for a given γ:

$$\mathbb{E}[B(\gamma)] = N\, F_r(0;\, \gamma). \tag{1}$$

vBK's expected BAU conservation is nested here: where they sum N plot-level conservation probabilities, this expression computes the same object from a single sufficient statistic for a landowner's returns.

[5] The returns distribution is parameterized by γ as well as σ, but, because all landowners have the same σ, I express the density as $F_r(r;\gamma)$ and not $F_r(r;\gamma,\sigma)$.

2.2 The baseline-price mechanism and participation

A policy consists of a price $p \geq 0$ and a baseline $b \in [0, N]$, written (b, p). The regulator announces the policy; each landowner then decides independently whether to participate. A participant is paid p for each parcel he conserves in excess of the baseline, so his **accredited abatement** is his conservation less b. Non-participants are unaffected by the policy and behave as in BAU.

Real offset policies set project-specific baselines based on observable data such as location, common practice, and forest type. These observable data do not truly reveal (γ_i, Ω_i), however, so such a policy is really a uniform baseline across a set of types it cannot distinguish. As such, b can be interpreted as one project or group's baseline and the primitives F_r, F_γ, and F_Ω as further heterogeneity within that group.

A participant maximizes his payoff parcel-by-parcel. For each, he can earn r_{ij} by developing or p by conserving, so his dominant strategy is to conserve every parcel with $r_{ij} \leq p$ and develop the rest.[6] Conservation under the policy therefore consists of two distinct pieces: parcels with $r_{ij} \leq 0$, which he would have conserved anyway, and the parcels with $r_{ij} \in (0, p]$, which the payment has induced. Only the second is additional.

The participation decision. A participant earns p on each conserved parcel and r_{ij} on each developed one, less the dollar value of the baseline and his participation cost. His outside option is the BAU. Differencing these two yields:

$$\underbrace{p(B_i - b)}_{\substack{\text{payment for BAU} \\ \text{conservation}}} + \underbrace{\sum_{j:\, r_{ij} \in (0,p]} (p - r_{ij})}_{\text{parcelwise profit}} - \underbrace{\Omega_i}_{\substack{\text{participation} \\ \text{cost}}} .$$

The first term is the non-additionality channel. A landowner with $B_i > b$ collects p for each parcel he would have conserved regardless. It is a transfer and, hence, welfare-neutral. Because these generate private value, the least additional landowners are the most willing to join any market. The second term is the surplus on induced conservation, and it represents the profit earned on every parcel with returns between 0 and p. The third is a pure welfare loss.

[6] He is indifferent at $r_{ij} = p$; I assume he conserves these parcels throughout.

Landowner i participates when the payoff is non-negative. Rearranging yields

$$b \leq \theta_i(p) \coloneqq B_i + \sum_{j\colon r_{ij} \in (0,p]} \left(\frac{p - r_{ij}}{p}\right) - \frac{\Omega_i}{p}.$$

The **participation threshold** $\theta_i(p)$ is the highest baseline at which i will participate, measured in units of parcels. Dividing the parcelwise profit by p converts dollars into parcel units.

Note the asymmetric role of price. It scales the parcelwise-profit term, which increases with p, while the parcel-equivalent value of the participation cost decreases with p. A landowner whose parcels are all non-additional has zero parcelwise profit, so his threshold moves with price only through the participation cost element. This asymmetry is the source of the threshold crossing exploited in Section 5. A landowner whose parcels are entirely non-additional attains the maximum threshold $N - \frac{\Omega_i}{p}$, so he participates under any policy that attracts someone with the same cost.

Expected thresholds. Since (γ_i, Ω_i) is unobservable by the regulator, she works with the expected threshold

$$\begin{aligned}\mathbb{E}\Big[\theta(\gamma, \Omega;\, p)\Big] &= N \cdot F_r(0;\, \gamma) + N \int_0^p \left(\frac{p - r}{p}\right) \cdot f_r(r;\, \gamma)\, \mathrm{d}r - \frac{\Omega}{p} \\ &= \underbrace{\mathbb{E}[B(\gamma)]}_{\substack{\text{avg. BAU} \\ \text{conservation}}} + \underbrace{\left(\frac{p - \mathbb{E}[r_\gamma \mid r \in (0,p]]}{p}\right)}_{\substack{\text{avg. parcelwise} \\ \text{profit}}} \underbrace{N \int_0^p f_r(r;\, \gamma)\, \mathrm{d}r}_{\substack{\text{incentivized} \\ \text{conservation}}} - \underbrace{\frac{\Omega}{p}}_{\substack{\text{participation} \\ \text{cost}}}, \end{aligned} \tag{2}$$

where the middle term is scaled by the expected number of parcels. Under Assumption 1,

$$\frac{\partial\, \mathbb{E}[\theta(\gamma, \Omega; p)]}{\partial \gamma} = -\frac{N}{p} \int_0^p f_r(r; \gamma)\, \mathrm{d}r < 0,$$

so expected thresholds decrease as land becomes more commercially valuable: the most valuable landholders are, ceteris paribus, the least willing to participate. Land in downtown Raleigh produces substantial social value when developed, and landowners downtown are the least likely to enter a conservation market.

The participation set. Because thresholds are monotone in γ, participation at a given cost is determined by a cutoff. For any policy and any cost Ω, define $\bar{\gamma}(\Omega; b, p)$ by the condition that $\mathbb{E}[\theta(\bar{\gamma}, \Omega; p)] = b$. A landowner with cost Ω participates if and only if $\gamma \leq$

$\bar{\gamma}(\Omega; b, p)$. Formally, the participation set is

$$\mathcal{P}(b, p) \coloneqq \Big\{(\gamma, \Omega) \colon \gamma \leq \bar{\gamma}(\Omega; b, p)\Big\}.$$

At every cost, the participation set is a lower interval in γ. Lowering the baseline (i.e., reducing barriers to entry) increases $\bar{\gamma}(\Omega; b, p)$, which enlarges $\mathcal{P}(b, p)$. As such, the baseline can be used by the regulator as a means to allow higher-γ participants, but cannot impose any lower bound on γ.

2.3 Welfare

The social value of conservation. Every conserved parcel generates the same external benefit δ, but conserving a parcel comes at the cost of forgone development returns. Because those returns reflect genuine economic value to the landowner, the social value of conserving parcel j is $\delta - r_{ij}$ if $r_{ij} \in (0, p]$, and 0 otherwise.

Incentivizing the conservation of a parcel j with $r_{ij} \leq 0$ contributes no social value because it would have been conserved under the BAU. The policy changes none of the behavior, and payment for its conservation is a transfer which nets out. Conserving a parcel with $r_{ij} > \delta$ contributes negatively, since it generates more economic value being developed than being conserved. Under a policy with $p \leq \delta$, no such parcel is ever induced into conservation—for this reason, Pigouvian pricing is a natural benchmark. The parcels that matter are those that generate socially valuable additional conservation; i.e., those with returns in $(0, p]$.

Individual welfare contributions. Let

$$\psi_i(p) \coloneqq \sum_{j \colon r_{ij} \in (0,p]} (\delta - r_{ij}) \tag{3}$$

denote the **gross social value** of the conservation that a policy with price p induces from landowner i. His net contribution to welfare is that value less his participation cost:

$$w_i(p) = \psi_i(p) - \Omega_i. \tag{4}$$

Note that these are defined for any landowner, regardless of whether the policy actually incentivizes his participation. Thus, $w_i(p)$ measures what i would contribute if he enrolled. The baseline does not appear in (4) because a participant's strategy is to conserve all plots

with returns up to p, and is unaffected by how many parcels he is paid for.

These terms carry the paper's central tension. A landowner with mostly non-additional parcels has $\psi_i(p)$ near zero, so when Ω_i is significant, his participation contributes negatively to welfare, even though he privately benefits. A landowner with very commercially valuable parcels has a small $\psi_i(p)$ for the opposite reason: most of his parcels' returns are above p. Positive welfare contributions come from landowners in the middle, i.e., those with many additional parcels that are not too commercially valuable.

In expectation over a type's returns,

$$\psi(\gamma; p) = N \int_0^p (\delta - r) f_r(r; \gamma)\, \mathrm{d}r \quad \text{and} \quad \mathbb{E}[w(\gamma, \Omega; p)] = \psi(\gamma; p) - \Omega.$$

Aggregate welfare. Only enrolled landowners contribute to aggregate welfare, which is the sum of contributions over the participation set:

$$W(b, p) = \sum_{i \in \mathcal{P}(b,p)} w_i(p) \quad \text{and} \quad W(b, p) = I \iint_{\mathcal{P}(b,p)} \Big[\psi(\gamma; p) - \Omega\Big] f_\gamma(\gamma) f_\Omega(\Omega)\, \mathrm{d}\Omega\, \mathrm{d}\gamma.$$

The baseline as a screening device. Note that price enters W twice: through each participant's welfare contribution and through the participation set itself. At the same time, the baseline enters only through the participation set. Raising b removes the landowners nearest the participation margin while having no other impact on aggregate welfare. If those landowners were contributing negative welfare, the aggregate welfare rises.

This separation of the impact of price and baseline makes the baseline a useful screening instrument for the regulator. She can adjust who is in the market without distorting what anyone in it does. Section 4 illustrates the usefulness of screening alone while Section 5 illustrates the power of leveraging both the baseline and the price.

3 Adverse selection

Section 2 defined the baseline-price mechanism: the baseline determines who participates while the price determines how much they conserve. This section discusses the limits of the mechanism. The first is that no uniform policy reaches the first-best: the outcome where only those who create positive social value participate, and socially efficient conservation. Under the first-best, the sum of the collective welfare (4) attains its maximum possible value.

The second is that the market can be harmful, but only when both asymmetric information and participation costs are present.

3.1 Impossibility of first-best

Landowner i participates whenever offset market revenue exceeds participation costs

$$p(B_i - b) + \sum_{j\colon r_{ij} \in (0,p]} (p - r_{ij}) \geq \Omega_i,$$

while his participation raises welfare only when the social value of his conservation covers that same cost,

$$\sum_{j\colon r_{ij} \in (0,\delta]} (\delta - r_{ij}) \geq \Omega_i.$$

The two conditions coincide for landowner i if and only if $p = \delta$ and $b = B_i$. The price is common across landowners, so $p = \delta$ can be satisfied for everybody. The baseline is also common, but BAU conservation is not: a scalar baseline can satisfy $b = B_i$ for at most one i. Thus, the regulator can align private and social participation incentives only for a single type. Then, every other type is at risk of being either a welfare-negative inclusion (where welfare-reducing participants enter the market) or a welfare-positive exclusion (where welfare-generating landowners choose not to participate).

This is a multidimensional screening problem. The offsets literature has modeled adverse selection in offset markets along one hidden dimension: either (i) hidden cost (Ferraro, 2008; Mason & Plantinga, 2013) or (ii) hidden BAU conservation (Bento et al., 2015; van Benthem & Kerr, 2013). Neither attains the first-best under a uniform instrument, but with a single hidden dimension, the type space can be ordered, and a scalar cutoff used to separate it. The participation cost Ω_i is a second private dimension, and Armstrong (1996) and Rochet and Choné (1998) establish that no scalar instrument (even a non-uniform and non-linear one) can separate a multidimensional type space.

The regulator's problem departs from that literature in one important respect. Armstrong (1996) and Rochet and Choné (1998) study a profit-maximizing monopolist facing rational buyers, so participation is always weakly beneficial to both parties and shutting down the market is never optimal.[7] In the current model, one of the hidden dimensions (Ω_i) is a pure

[7] See Cesa-Bianchi et al. (2025) for a welfare maximization approach under a different type of uncertainty. Determining welfare maximizing policy in their case requires a process of eliciting hidden values for the good, something that is beyond the scope of this paper.

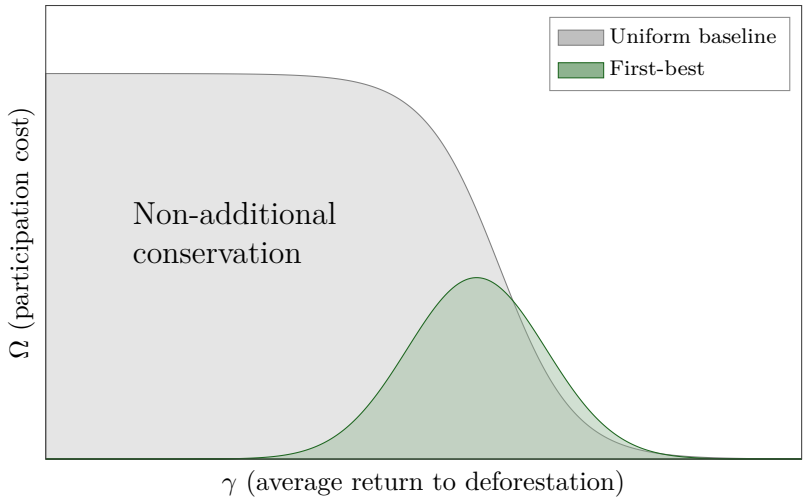


Figure 2: Participation sets compared to the first-best set

Note: This figure illustrates the gap between the uniform baseline's participation frontier and the first-best frontier in the (γ, Ω) plane.

surplus loss. This fact, outside the classical screening and offsets literature, is what makes market shutdown a potential optimum.

Figure 2 shows the gap in the (γ, Ω) plane. The first-best frontier traces the combinations of average returns and participation costs for which a landowner's gross social value exactly covers his participation cost. This is the socially efficient participation boundary: everybody in the shaded region under this boundary should participate. The uniform baseline frontier traces the types who actually do participate. The shaded region between the two frontiers represents the unavoidable distortion caused by the constraint that the baseline is a scalar. Landowners in the gray region above the green dashed line participate under the uniform baseline, but would be excluded under first-best. Any landowners in the green region, but not the gray region, would participate under first-best but are excluded by the uniform baseline.

A regulator with better prediction technology can forecast B_i more accurately and can set b closer to the BAU conservation of any type. This is a genuine improvement, but it can only shift the uniform baseline's participation frontier in one dimension. It cannot lie on top of the first-best frontier without also improving the predicted Ω_i.

3.2 Joint effects of asymmetric information and transaction costs

Proposition 1 (Both frictions are necessary for negative welfare)**.**
Let $W(b,p) = \sum_{i \in \mathcal{P}(b,p)} w_i(p)$ be aggregate net welfare of a policy (b,p). The following three

statements hold.

(i) (No transaction costs.) *If* $\Omega_i = 0$ *for every* i*, then* $W(b,p) \geq 0$ *for every policy with* $0 < p \leq \delta$*. Hence,* $\max\{W(b,p)\} \geq 0$ *and an active market weakly dominates shutdown,* irrespective of the informational environment.

(ii) (No asymmetric information.) *If the regulator observes each landowner's development profits, she can guarantee* $W \geq 0$*,* regardless of the distribution of transaction costs. *In fact, by using a type-contingent baseline* $b = B_i$ *and an eligibility rule, she can guarantee the first-best aggregate welfare.*

(iii) (Both frictions.) *There exist economies with* $\Omega_i > 0$ *for a positive mass of landowners and privately known returns for which* $\max_{(b,p)} W(b,p) < 0$*; for these, market shutdown strictly dominates every active market.*

Consequently, an offset market is strictly welfare-reducing only if asymmetric information and transaction costs are present simultaneously.

Proposition 1 isolates the distinct roles of the two frictions. Part (i) says transaction costs are what create the possibility of negative welfare: with $\Omega_i = 0$ and a price at or below the externality value, every additional plot contributes $\delta - r_{ij} \geq 0$, so no participant can be a net negative welfare contributor and the sign of aggregate welfare is always non-negative regardless of who selects in. Adverse selection can *shrink* the gains—by drawing in low-γ, largely non-additional landowners whose plots contribute little—but it cannot make them negative.

Part (ii) says asymmetric information about γ_i and Ω_i is what strips the regulator of her ability to undo the adverse selection damage. When she observes types, she can simply set $b_i = B_i$, and landowners will properly account for Ω_i when choosing whether to participate. The transaction costs thin the admitted set, but welfare non-negativity is still guaranteed.

Only when both frictions are present can the participation set be populated by landowners who are simultaneously welfare-negative and unremovable. This is because the participation set consists of a lower interval in γ and the baseline can only screen from the top of the γ distribution. Example 1 constructs such an economy.

Example 1 (Closure dominates every active market)**.** Let $\delta = 0.5$, and consider an economy consisting only of the two following landowners, one of whom has entirely inframarginal plots (ℓ) while the other (h) has a mix of inframarginal and true conservation.[8]

[8]The returns and participation cost values are chosen purely to serve as an example and are not meant

Table 1: Two landowners in Example 1

Type	Returns	Ω_i	B_i	$w_i(\delta)$
ℓ	{-0.4, -0.3, -0.2, -0.1}	0.3	4	-0.3
h	{-0.1, -0.1, 0.3, 0.4}	0.2	2	0.1

Notes: The two types of landowners in Example 1 vary in their returns to deforestation and in their participation costs. Their baselines and welfare (at any price) are calculated based on the returns and participation costs. Welfare contributions at the Pigouvian price are presented as this price incentivizes the highest amount of socially-efficient conservation, illustrating the difference in magnitude between ℓ's negative welfare contribution and h's greatest possible welfare contribution.

Observe the following facts about this economy:

(a) Every plot owned by ℓ has a non-positive return, so no plot ever lies in $(0, p]$ and $w_\ell(p) = -\Omega_\ell = -0.3$ for every price $p > 0$. Whenever ℓ participates it reduces W by 0.3.

(b) The maximum welfare contribution h can make is less than ℓ's participation cost: $w_h(p) = \sum_{r_{hj} \in (0,p]} (\delta - r_{hj}) - \Omega_h \leq (0.2 + 0.1) - 0.2 = 0.1 < 0.3 = \Omega_\ell$.

(c) h cannot be admitted without ℓ wherever h is beneficial. Using the definition of the baseline threshold for i, $\theta_i(p)$,

$$\theta_\ell(p) - \theta_h(p) = \begin{cases} 2 - 0.1/p \lesseqgtr 0, & p < 0.3, \\ 1 + 0.2/p > 0, & 0.3 \leq p < 0.4, \\ 0.6/p > 0, & p \geq 0.4. \end{cases}$$

Since $w_h(p) \geq 0$ at any $\{p \geq 0.3\}$ but $\theta_\ell(p) < \theta_h(p)$ only for $p < 0.05 < 0.3$, no baseline can enroll h while excluding ℓ at any price where h contributes non-negative welfare.

These pin down welfare at *every* policy. For any (b, p), exactly one of the following outcomes is possible:

- $\ell \in \mathcal{P}(b, p)$: by (a) and (b), $W(b, p) = -0.3 + \max\{0, w_h(p)\} \leq -0.3 + 0.1 = -0.2 < 0$; or
- $\ell \notin \mathcal{P}(b, p)$ but $h \in \mathcal{P}(b, p)$: this needs $\theta_\ell(p) < \theta_h(p)$. By (c), this requires $p < 0.05$, where $w_h(p) = -0.2$. Thus, $W(b, p) = -0.2 < 0$; or, finally,
- neither participates: $W(b, p) = 0$ (i.e. shutdown).

to represent any two landowners.

It follows, then, that every active market yields $W(b,p) \leq -0.2 < 0$. As such,

$$\max_{(b,p):\ \mathcal{P}(b,p)\neq\emptyset} W(b,p) = -0.2 < 0 = W_{\text{shutdown}}.$$

Closure strictly dominates every active policy. This establishes Proposition 1 (iii).

Finally, note that removing either friction restores a viable market. With no transaction costs ($\Omega_\ell = \Omega_h = 0$), $w_\ell \equiv 0$ and $w_h(\delta) = 0.3$. The Pigouvian policy ($p = \delta$) gives $W = 0.3 > 0$. The ℓ-type still cannot be separated at this price ($\theta_\ell(\delta) = 4 > \theta_h(\delta) = 2.6$), so adverse selection remains, but its non-additional conservation is now a costless transfer rather than a welfare reduction. With full information, the regulator can forbid ℓ from participating, enrolling h alone at, e.g., $p = \delta$, for $W = w_h(\delta) = 0.1 > 0$. Each friction alone leaves a weakly beneficial market; only their conjunction forces closure. ■

Proposition 1 is an existence result showing that welfare-negative markets are possible. However, it says nothing about when they arise. Section 4 provides such a condition and characterizes the optimal baseline for any price. Section 5 then asks whether the price can be leveraged to screen better than the baseline.

4 Results with continuous type

Section 2 established that the baseline affects welfare only through participation. This section demonstrates the power that this gives the regulator. I first characterize the optimal baseline for any given price, first when participation costs are common across landowners and then when they are heterogeneous. Two features emerge that have no counterpart in the frictionless model. The optimal baseline is interior and screens from the top, excluding the most commercially valuable land rather than the least additional; and with heterogeneous costs it is determined not by a single marginal type but by an average over the landowners sitting at the participation margin. Because the characterization holds at every price, it reduces the regulator's two-dimensional problem to a single dimension. That reduced problem is not tractable in the continuum as price enters welfare through the participant's contribution, his threshold, and the identity of the marginal type simultaneously. In Section 5 I turn to a two-type economy to study the price directly.

In this section, I treat the economy as continuous. This allows for generalized results which use calculus in calculating "population" statistics. Unless otherwise stated, all proofs are relegated to Appendix A.

Assumption 1 implies that $\gamma \to \pm\infty$ causes $\mathbb{E}[w(\gamma, \Omega; p)] \to -\Omega$, for any Ω. An implication of the log-concavity is that for each Ω, $\mathbb{E}[w(\gamma, \Omega; p)]$ is unimodal on its γ support (Bagnoli & Bergstrom, 2005); this guarantees that the expected welfare contribution of any landowner with participation cost Ω is shaped like an inverted-U with γ-asymptotes at $-\Omega$. These two implications jointly imply that for any $\Omega \geq 0$, expected welfare contribution is either always negative, or contains a single interval which is positive and unimodal. Thus, there are exactly two zero-crossings in expected welfare: γ_L and γ_R, with $\gamma_L < \gamma_R$. The following lemma formalizes this.

Lemma 1 (Expected welfare zero-crossings)**.** *Under Assumption 1, the expected welfare contribution of a (γ, Ω)-type landowner is continuous in γ for a given Ω, with $\mathbb{E}[w(\gamma, \Omega; p)] \to -\Omega$ as $\gamma \to \pm\infty$. Moreover, if $\max_\gamma \mathbb{E}[w(\gamma, \Omega; p)] > 0$, then $\mathbb{E}[w(\gamma, \Omega; p)]$ has exactly two zero-crossings, $\gamma_L(p) < \gamma_R(p)$, with $\mathbb{E}[w(\gamma, \Omega; p)] > 0$ for all $\gamma \in (\gamma_L(p), \gamma_R(p))$ and $\mathbb{E}[w(\gamma, \Omega; p)] < 0$ otherwise.*

4.1 Screening with a common participation cost

Suppose every landowner faces the same participation cost $\Omega > 0$. This is the special case of a degenerate cost distribution, generalized in the next section.

With Ω common, participation depends completely on γ. More specifically, every γ-type below $\bar{\gamma}(b, p)$ participates and every γ-type above stays out. As such, the baseline acts purely as a type-selection mechanism. Choosing b is equivalent to choosing the highest-return landowner admitted to the market. I show in the next section that this is a knife-edge property of the common participation cost rather than a general feature of the mechanism.

The regulator's problem is

$$\max_b I \int_{-\infty}^{\bar{\gamma}(b,p)} \mathbb{E}\Big[w(\gamma, \Omega; p)\Big] f_\gamma(\gamma)\, \mathrm{d}\gamma, \quad \bar{\gamma}(b, p) \mapsto \mathbb{E}\Big[\theta(\bar{\gamma}, \Omega; p)\Big] = b.$$

Because $\mathbb{E}[w(\gamma, \Omega; p)]$ does not depend on b, the baseline enters only through the limit of integration, and

$$\frac{\partial W}{\partial b} = I\, \mathbb{E}\Big[w(\bar{\gamma}, \Omega; p)\Big] f_\gamma(\bar{\gamma}) \frac{\partial \bar{\gamma}}{\partial b}.$$

Since $\frac{\partial \bar{\gamma}}{\partial b} < 0$ by Assumption 1, the entire change in welfare is driven by the removal of the marginal type. Raising the baseline improves welfare only if the expelled type contributed negatively. The optimum is where he contributes nothing, $\mathbb{E}\Big[w(\bar{\gamma}^*, \Omega; p)\Big] = 0$. Lemma 1 defines two such types, $\gamma_L < \gamma_R$, and the optimum is the upper one since welfare between

them is positive.

Proposition 2 (Optimal baseline with common cost)**.** *Let Assumption 1 hold and suppose all landowners share a participation cost* $\Omega > 0$ *with* $\max_\gamma \mathbb{E}\big[w(\gamma, \Omega; p)\big] > 0$*. The optimal baseline admits every* γ*-type up to* $\bar{\gamma}^* = \gamma_R(p)$*, and equals*

$$b^*(p) = \mathbb{E}[B(\bar{\gamma}^*)] - N\left(\frac{\delta - p}{p}\right)\int_0^p f_r(r; \bar{\gamma}^*)\, dr.$$

At the Pigouvian price, the wedge vanishes and $b^*(\delta) = \mathbb{E}[B(\bar{\gamma}^*)]$*.*

The optimal baseline is thus the marginal participant's expected BAU conservation, adjusted by a wedge when price is non-Pigouvian.

The baseline screens participants from the top of the γ distribution. Because thresholds fall as land becomes more valuable, $\bar{\gamma}^*$ is an upper bound on participants' returns. What the baseline excludes are those landowners incentivized to provide too little conservation to provide a positive welfare contribution. This is the exact opposite of the non-additional landowner.

The wedge links price and baseline. When $p < \delta$, each conserved plot generates social value that is not fully compensated by the price. To substitute for this, the baseline is made more favorable to encourage greater participation. This is the substitutability exploited in Section 5 where lowering the price becomes a device for separating types at the cost of less conservation.

Along the Pigouvian locus, $b^*(\delta) = \mathbb{E}[B(\bar{\gamma}^*)]$ depends on δ only through the marginal type, and $\bar{\gamma}^*$ rises with δ. The optimal baseline therefore falls as conservation becomes more valuable. Because the optimal baseline screens from the top, raising δ makes some of the commercially valuable land the baseline was excluding worth conserving. As such, the regulator widens the participation margin.

Example 2 makes the screening decision concrete and shows what the optimal baseline excludes.

Example 2 (Normal returns, common participation cost)**.** Suppose that returns follow a standard normal distribution. There are $N = 100$ plots owned by each of $I = 1,000$ landowners, whose average returns γ are drawn uniformly from $[-5, 5]$. Every landowner faces the same participation cost $\Omega = 2.5$, the external benefit is $\delta = 0.5$, and the price is Pigouvian (i.e., $p = \delta$).

Figure 3a shows the participation region over candidate baselines. On the horizontal axis are candidate baselines and on the vertical axis are landowner types γ. At $b = 0$, every type participates; as the baseline rises, the highest-γ types exit first, and the dashed lines mark b^* and the marginal type $\bar{\gamma}^*$ it implements.

Figure 3b shows the welfare contributions across different γ-types. The curve is the inverted-U of Lemma 1, with zero-crossings at γ_L and $\gamma_R = \bar{\gamma}^*$. The red shaded region is the negative welfare the regulator avoids by screening, and the dotted continuation shows what an open market ($b = 0$) would have admitted. Finally, Figure 3c converts this to aggregate terms, and the gain from screening is substantial. Starting from an open market, raising the baseline initially increases aggregate welfare because it removes high-γ landowners who make negative welfare contributions. Aggregate welfare reaches its maximum when the marginal landowner has $\gamma \approx \gamma_R$, after which point it decreases as b increases.

Figure 4 shows what is being excluded. Across the γ support, it decomposes each type's parcels into non-additional, additional, and developed. The types screened out at b^* are overwhelmingly developers: nearly all of their parcels earn returns above $p = \delta$, so they conserve little and would have contributed almost nothing beyond their participation cost. Those at the far left, whose parcels are almost entirely non-additional, remain in the market, and their participation costs are absorbed as pure loss.

■

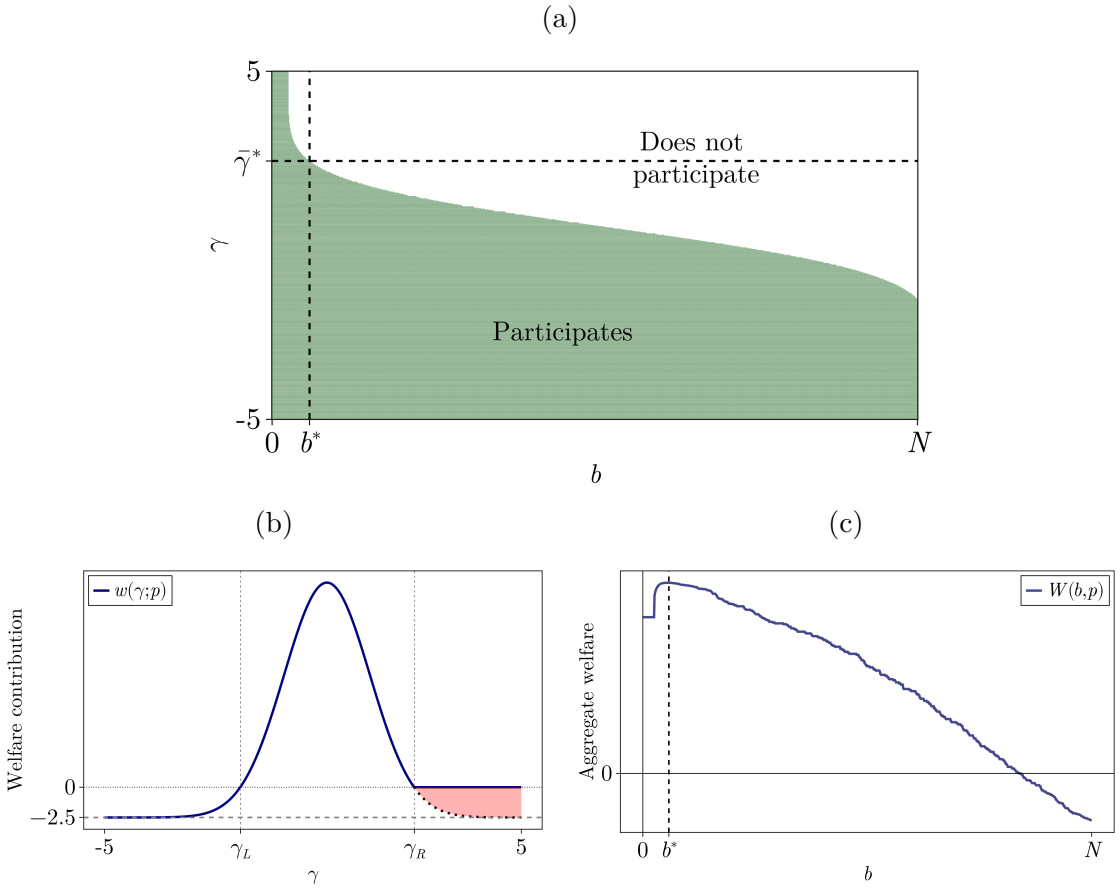
The example illustrates the central limitation of baseline screening. The optimal baseline can remove the high-γ landowners whose participation is socially harmful, but it can do nothing about the lowest γ-types whose participation is socially costly. The baseline therefore improves welfare by screening out the worst participants, but it cannot implement the first-best participation set. This limitation becomes more consequential when participation costs themselves differ across landowners: a low-γ landowner with a sufficiently high participation cost may be worth excluding, while a higher-γ landowner with a sufficiently low cost may be worth retaining. The next section allows participation costs to vary and asks how the regulator should choose a baseline when screening must operate along both dimensions.

4.2 Heterogeneity in participation costs

I now let participation costs vary, with each Ω_i drawn independently from F_Ω.

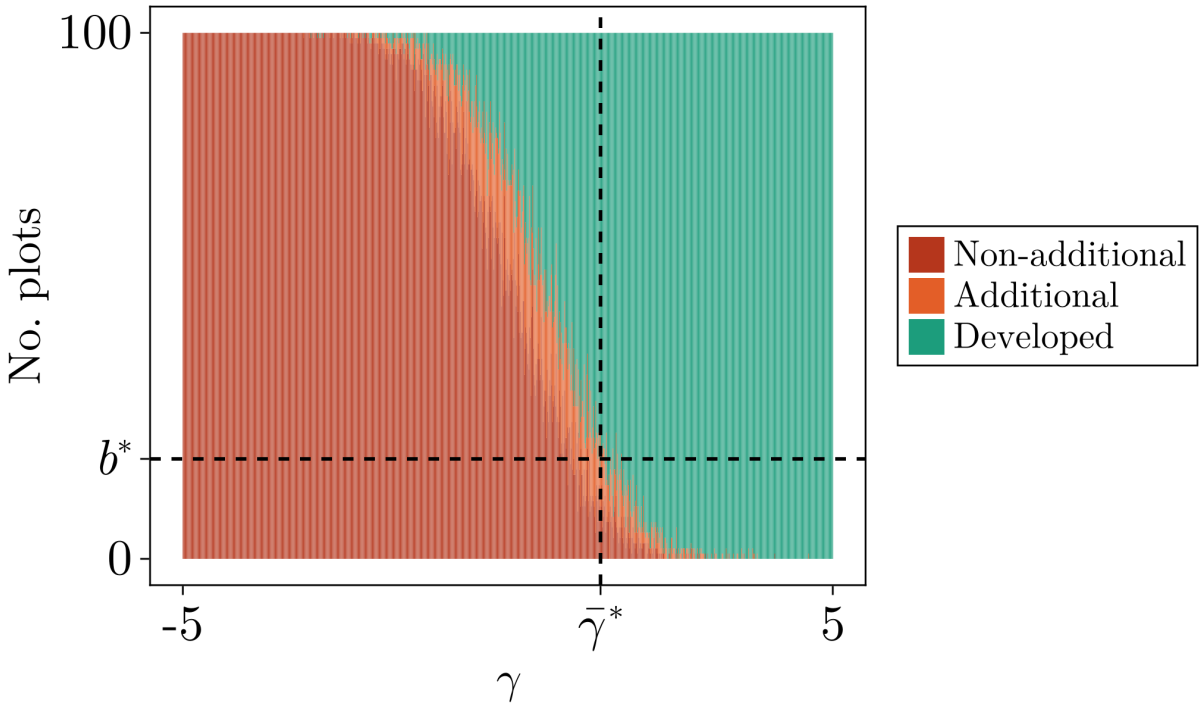
Assumption 2 (Cost distribution regularity)**.** *The participation-cost distribution F_Ω is independent and continuous and strictly positive on $(0, \Omega_{\max})$, with no atom at zero, so that*

Figure 3: Welfare effects of screening out at Pigouvian price



Notes: Figure 3a shows the participation region for this simulation. On the horizontal axis is candidate baselines and on the vertical axis is the observed types γ, with horizontal dashed line representing $\bar{\gamma}^*$ and the vertical dashed line representing b^*. Figure 3b shows the welfare contribution of the different γ-types. The two zero-crossings are labeled, and $\bar{\gamma}^* = \gamma_R$. Figure 3c shows the real value in terms of aggregate welfare from screening out excessively high γ-types.

Figure 4: Plots conserved and screened out at Pigouvian price



Notes: This figure demonstrates the kinds of plots that are being screened out by the optimal baseline. The horizontal axis covers the entire γ support, with the number of non-additional plots colored red, the number of additional plots colored orange, and the number of developed plots colored green. The horizontal and vertical dashed lines show b^* and $\bar{\gamma}^*$, respectively.

$F_\Omega(0) = 0$.

Participation now depends on both dimensions of a landowner's type, and the participation set is two-dimensional. Rearranging $b \leq \mathbb{E}[\theta(\gamma, \Omega; p)]$ for Ω gives, for each γ, a cost cutoff below which that type participates:

$$\bar{\Omega}(\gamma;\, b, p) \coloneqq \underbrace{p\Big[NF_r(p;\, \gamma) - b\Big]}_{\text{avg. conservation income}} - \underbrace{\mathbb{E}\Big[r_\gamma \mid r \in (0, p]\Big]}_{\text{avg. returns | add'l}} \underbrace{N \int_0^p f_r(r;\, \gamma)\, \mathrm{d}r}_{\substack{\text{avg. add'l} \\ \text{conservation size}}}. \tag{5}$$

The first term in (5) is expected conservation income, which raises the cost a landowner will bear; the second enters negatively because participation requires forgoing the returns on newly conserved plots. The cutoff type $\bar{\gamma}(b, p)$ is the γ whose threshold reaches b even at zero cost, so $\mathbb{E}[\theta(\bar{\gamma}, 0; p)] = b$ if and only if $\bar{\Omega}(\bar{\gamma}; b, p) = 0$.

The cutoff type is well defined: by Section 2.2, $\frac{\partial\, \mathbb{E}[\theta(\gamma, 0; p)]}{\partial \gamma} = -\frac{N}{p} \int_0^p f_r(r; \gamma)\, dr < 0$. As such, $\gamma \mapsto \mathbb{E}[\theta(\gamma, 0; p)]$ is continuous and strictly decreasing from N to 0. Hence for each $b \in (0, N)$ there is exactly one $\bar{\gamma}(b, p)$ with $\mathbb{E}[\theta(\bar{\gamma}, 0; p)] = b$, equivalently $\bar{\Omega}(\bar{\gamma}; b, p) = 0$; and $\bar{\Omega}(\gamma; b, p) > 0$ if and only if $\gamma < \bar{\gamma}(b, p)$. At $b = N$ no type satisfies the participation condition, so the market is empty and $W(N, p) = 0$.

The regulator's objective becomes

$$\max_b I \int_{-\infty}^{\bar{\gamma}} \Big[\underbrace{\psi(\gamma; p) F_\Omega(\bar{\Omega}(\gamma; b, p))}_{\substack{\text{gross value, weighted} \\ \text{by participation}}} - \underbrace{\int_0^{\bar{\Omega}(\gamma; b, p)} \Omega f_\Omega(\Omega)\, \mathrm{d}\Omega}_{\substack{\text{costs borne by} \\ \text{participants}}}\Big] f_\gamma(\gamma)\, \mathrm{d}\gamma,$$

with $\psi(\gamma; p) \coloneqq N \int_0^p (\delta - r) f_r(r; \gamma)\, \mathrm{d}r$ the gross social value of a γ-type's additional conservation.

The boundary term disappears. Differentiating the baseline moves both the bracketed term and the limit $\bar{\gamma}$. But at $\gamma = \bar{\gamma}$ only landowners with $\Omega = 0$ participate, a measure-zero set under Assumption 2. Shifting the cutoff type therefore has no first-order effect, and the entire effect of the baseline is through the cost cutoff, which decreases by p for every participating γ.

Lemma 2 (Baseline derivative of aggregate welfare)**.** *Let Assumptions 1 and 2 hold, and define $\psi(\gamma; p) \coloneqq N \int_0^p (\delta - r) f_r(r; \gamma)\, dr$ and $\bar{\Omega}(\gamma; b, p)$ as in (5), with $\bar{\gamma}(b, p)$ the cutoff type solving $E[\theta(\bar{\gamma}, 0; p)] = b$ (equivalently $\bar{\Omega}(\bar{\gamma}; b, p) = 0$). Then aggregate welfare is differentiable*

in b, with

$$\frac{\partial W}{\partial b} = -pI \int_{-\infty}^{\bar{\gamma}(b,p)} \Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big]\, f_\Omega\Big(\bar{\Omega}(\gamma;b,p)\Big)\, f_\gamma(\gamma)\, d\gamma. \tag{6}$$

Corollary 2.1 (Sign of the baseline effect)**.** *Define the marginal-participant density*

$$h(\gamma;b,p) = \frac{f_\Omega\left(\bar{\Omega}(\gamma;b,p)\right)\, f_\gamma(\gamma)}{\int_{-\infty}^{\bar{\gamma}} f_\Omega\left(\bar{\Omega}(\gamma';b,p)\right)\, f_\gamma(\gamma')\, d\gamma'},$$

which integrates to 1 over $(-\infty, \bar{\gamma})$. Then

$$\operatorname{sign}\left(\frac{\partial W}{\partial b}\right) = -\operatorname{sign}\left(\mathbb{E}_h\Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big]\right),$$

where $\psi(\gamma;p) \coloneqq N\int_0^p (\delta - r) f_r(r;\gamma)\, dr$ and $\mathbb{E}_h[X(\gamma;b,p)] \coloneqq \int_{-\infty}^{\bar{\gamma}(b,p)} X(\gamma;b,p)\, h(\gamma;b,p)\, d\gamma$.

The object h deserves a name because it is what the regulator is acting on. Raising b by one unit removes, at each γ, the landowner currently at that type's cost cutoff. There are $f_\Omega(\bar{\Omega}(\gamma;b,p))$ such landowners, weighted by how common the type is; h normalizes that product. The regulator is not choosing a single type to expel (as in the common-cost case), she is expelling a crowd of types. In particular, she is removing the highest-cost participant at every γ-type all at once. Corollary 2.1 says she should keep raising the baseline as long as the crowd being removed is welfare-negative on average. Setting the derivative to zero gives Proposition 3.

Proposition 3 (Optimal baseline with heterogeneous participation costs)**.** *Under Assumptions 1 and 2, the optimal baseline is the h-weighted expected value of a wedge-adjusted conservation measure, taken over the density of marginal participants:*

$$b^*(p) \mapsto b = \mathbb{E}_h\left[B(\gamma) - N\left(\frac{\delta - p}{p}\right)\int_0^p f_r(r;\gamma)\ dr\right]. \tag{7}$$

At the Pigouvian price $p = \delta$ the wedge vanishes and the optimal baseline is the h-weighted expected BAU conservation of the marginal landowners, $b^*(\delta) \mapsto b = \mathbb{E}_h\Big[\mathbb{E}[B(\gamma)]\Big]$.

What does this marginal-participant weighting mean? The formula coincides with the common-cost case in that the optimal baseline is an expected wedge-adjusted conservation measure. This baseline is again equal to the expected BAU conservation of the marginal participant in the case of a Pigouvian price. In vBK, and in the common-cost case above, the baseline is set to one landowner's counterfactual, and the regulator's information problem is to predict that single counterfactual accurately. In this case, the baseline is set to an

average over the set of landowners at the participation margin. That set is determined by the baseline itself; Proposition 3 is therefore a fixed point, not a formula.

Condition (7) is necessary for an interior optimum, not sufficient. It does not assert that a solution is unique, nor does it rule out a corner optimum at $b = 0$ or $b = N$, at which the first-order condition may not be satisfied. However, the proposition does show that at an optimal interior baseline, the crowd of landowners sitting at the participation margin contributes zero welfare on average, and the regulator gains nothing by thinning it further.

Pinning down h therefore requires measuring F_Ω directly. That is a different empirical program from the one the reform debate currently funds, but it is not an infeasible one: Pearson et al. (2014) estimate these costs for tropical forest projects while Naegele (2018) and Robinson (2026) estimate their demand-side counterparts in the EU and California, respectively.

This is consequential for the policy debate on baseline accuracy. Improving prediction of B_i improves the regulator's estimate of ψ and $\bar{\Omega}$, but it does not tell her h. That is because h depends on the unobserved joint distribution of returns and participation costs of landowners at the participation margin. A regulator with a perfectly accurate estimate of the returns distribution still has uncertainty over the costs required to supply offsets. When only one dimension is hidden (i.e., participation costs are homogeneous), the scalar cutoff separates the type space so that marginal types can be screened out using only the baseline. With two hidden dimensions, no scalar instrument induces a participation frontier that matches the first-best (i.e., Figure 2; Armstrong, 1996; Rochet & Choné, 1998). With common costs, the optimal baseline is the BAU conservation of the marginal type, as in vBK. With heterogeneous costs, it is a weighted average over the set of marginal participants, and that weighting depends on the baseline itself; hence, (7) is a fixed-point characterization rather than a function. Comparative statics with respect to the type distributions are less tractable, since the marginal-participant weighting in Proposition 3 depends on the baseline itself; Section 5 recovers these in a two-type economy.

4.3 When is a market worthwhile?

Proposition 4 (Market viability)**.** *Let Assumptions 1 and 2 hold, fix p, and let b^* be the optimal baseline at price p. If the participating mass is positive, then*

$$W(b^*, p) > 0 \iff \mathbb{E}\Big[\psi(\gamma; p) \mid (\gamma, \Omega) \in \mathcal{P}(b^*, p)\Big] > \mathbb{E}\Big[\Omega \mid (\gamma, \Omega) \in \mathcal{P}(b^*, p)\Big],$$

where $\psi(\gamma; p) \coloneqq N \int_0^p (\delta - r) f_r(r; \gamma)\, dr$. *If the inequality fails, shutdown weakly dominates every policy at price* p.

The market is worth opening precisely when the average participant's additional conservation is worth more than the average participant's cost of supplying it. Both expectations are taken over the participation set. Asymmetric information determines the conditioning set: the regulator cannot explicitly choose who is in the participation set, but she can shift its boundary. The transaction costs determine whether the average inside this set is positive.

As stated in Proposition 1, neither friction alone makes shutdown the dominant choice of the regulator. With $\Omega \equiv 0$, the righthand side is zero and the lefthand side is non-negative at any $p \leq \delta$, so the condition holds trivially. With observable types, the regulator admits only landowners with $\psi \geq \Omega$, and the condition holds by construction.

The market viability condition can fail even at the optimal baseline, and Example 1 demonstrates this. Screening operates only from the top of the γ distribution. That is, b^* removes types above $\bar{\gamma}^*$, but every type below γ_L remains in the participation set. Because $\mathbb{E}[\theta]$ is decreasing in γ (Assumption 1), the participation set is always a lower interval: no baseline excludes the tail below γ_L while retaining the welfare-positive interval $(\gamma_L, \bar{\gamma}^*)$. The impossibility is a consequence of monotonicity. Because a landowner's willingness to participate falls as his land becomes more valuable, any instrument that admits a high-γ type necessarily admits every lower type, so the regulator can move the participation boundary but cannot screen from the bottom. An instrument whose participation set were non-monotone in γ would not face this constraint.

These landowners conserve almost nothing beyond what they would have conserved anyway, so the gross social value of their additional conservation is near zero while their participation cost is positive. They enter the conditional expectation on the left side of the condition and drag it down, and the baseline cannot remove them. Whether the market is viable therefore depends on how much mass sits at that lower tail relative to the interval the regulator can retain while generating positive welfare. That depends on the distribution, not the mechanism.

5 Two-type economy

Section 4.3 left the regulator with an instrument that cannot reach the landowners doing the most damage (i.e., the most non-additional). At a fixed price, the participation set is a lower

interval in γ, so the baseline can only move the boundary from above, and the non-additional tail stays in the market at every welfare-positive baseline. This section explores using price as a tool to do screening that the baseline alone cannot do.

A landowner's participation threshold is a function of price. Given two types, if their thresholds have a fixed ordering at every price, no price can separate them: one is always screenable while the other is not. With a common participation cost, this ordering is fixed, and the regulator's only screening tool is the baseline (hence vBK's result that the effect of price on average cost of additional conservation is monotone). Heterogeneous participation costs break this. The cost enters the threshold as Ω_i/p, while the revenue from conservation is proportional to p. Thus, a high participation cost can outweigh the conservation revenue at low prices while this becomes less likely at higher prices—the threshold of the two types can cross, and on either side of the crossing the regulator can admit a different one of them. The price acquires a selection role not seen in vBK or Bento et al. (2015), where it governs only how much conservation a participant supplies.

Characterizing this in the continuum is not tractable. Price enters aggregate welfare through three channels simultaneously: each participant's welfare contribution, his participation threshold, and the identity of the marginal type. Proposition 3's fixed point does not admit a tractable reduced-form objective function in p. I therefore turn to a two-type economy. The structure is restrictive, but it reduces the price question to a comparison between a pooling regime and one of two price-dependent separating regimes. In doing so, I illustrate that when asymmetric information and transaction costs are both present, sub-Pigouvian pricing can dominate the Pigouvian benchmark. The comparison makes the mechanism visible, and it also permits a clean characterization of when the market should be shut down entirely, which the continuum model can state (Proposition 4) but not solve analytically.

Assumption 3 (High and low landowner types)**.** *I take the following as given:*

- *The number of parcels is normalized to unity, with a type i landowner owning a share N_i of all parcels;*
- $\gamma_\ell \ll 0 < \gamma_h < \delta$*;*
- $0 \leq \Omega_h < \Omega_\ell$*;*
- $\Omega_\ell > N_\ell \int_0^\delta (\delta - r) f_r(r; \gamma_\ell)\, dr$*: ℓ-types always contribute negative welfare;*
- $N_\ell F_r(p; \gamma_\ell) > N_h F_r(p; \gamma_h)$ *for all* $p \leq \delta$*.*

Assumption 3 splits the economy into two mutually exclusive and collectively exhaustive types. The assumptions on the welfare contributions of the two types immediately imply that h is the more desirable group of landowners—if possible, h would be the only participating type.

For a given p, the ℓ-types have a lower participation threshold than h-types, on average, if and only if $\mathbb{E}[\theta_h] > \mathbb{E}[\theta_\ell]$. This can be expressed as an inequality between the **no-cost threshold difference**

$$\begin{aligned} g(p) &\coloneqq p\,\mathbb{E}\Big[\theta(\gamma_\ell, 0; p) - \theta(\gamma_h, 0; p)\Big] \\ &\equiv p\Big[N_\ell F_r(0;\gamma_\ell) - N_h F_r(0;\gamma_h)\Big] + \int_0^p (p-r)\Big[N_\ell f_r(r;\gamma_\ell) - N_h f_r(r;\gamma_h)\Big]\,\mathrm{d}r, \end{aligned} \tag{8}$$

which is simply the difference in the price-varying aspect of each type's thresholds, and the **participation cost difference**, $D \coloneqq \Omega_\ell - \Omega_h$, which is the simple measure of how much larger ℓ's participation cost is than h's. Thus, h has a higher expected participation threshold if and only if $g(p) < D$:

$$\mathbb{E}[\theta_h] > \mathbb{E}[\theta_\ell] \iff D > g(p).$$

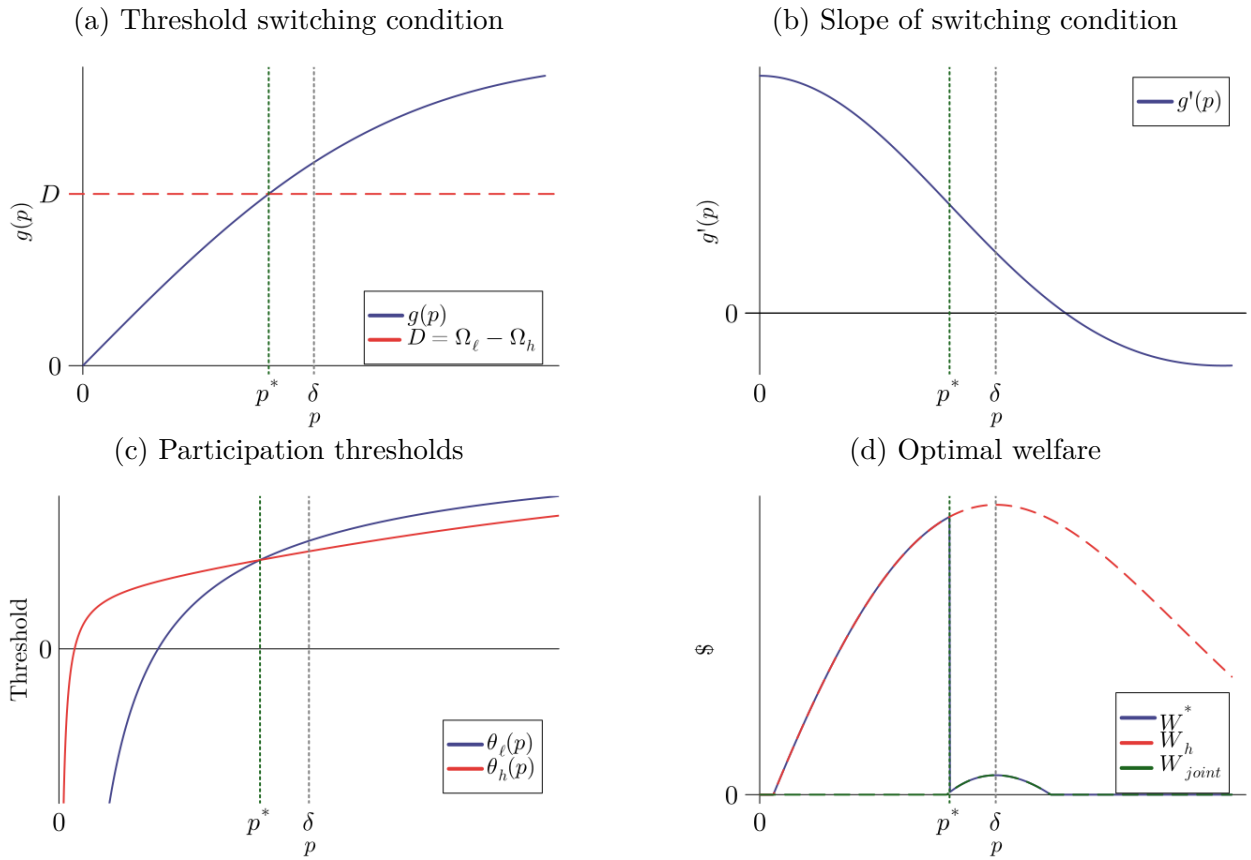
The final point in Assumption 3 states that the non-additional type conserves more parcels in expectation than the additional type at every price between 0 and δ. This alone guarantees that $g(p)$ is strictly increasing on $[0, \delta]$.

5.1 Characterizing optimal policy

When $g(p) > D$, ℓ's expected threshold exceeds h's. So any baseline low enough to admit h also admits ℓ: the two types cannot be separated in a welfare-improving direction. Thus, the regulator's only choice is to pool the two landowners or close the market. On the other hand, when $g(p) < D$, the ordering reverses and any baseline $b \in \Big(\mathbb{E}[\theta_\ell(p)], \mathbb{E}[\theta_h(p)]\Big]$ will include h while excluding ℓ. Since $g(0) = 0 < D$, such a price always exists under assumption 3. The relevant question is which price is optimal; I show here that the optimal price is the highest price such that ℓ can be screened out of the market.

Whether a crossing is useful depends on its direction at the current price (i.e., the sign of $g'(p)$). When $g(p) \in [0, D)$ and $g'(p) > 0$, raising the price narrows the distance between the two thresholds. A price just below a crossing of D leaves h with the higher threshold, while a price just above the same crossing eliminates it. These "upward" crossings are the

Figure 5: Dynamics of the two-type market



Notes: Panel (a) shows the shape of the g function, illustrating the source of p^*. Panel (b) shows that the derivative of g is positive on the Pigouvian domain, guaranteeing that an h-only regime is possible for prices below p^*. Panel (c) shows the thresholds underlying g to demonstrate that p^* is the price at which threshold switching occurs. Panel (d) illustrates the aggregate welfare under different regimes; it is clear that the optimal policy is to set price just below p^*.

ones the regulator can exploit. This is Lemma 3.[9]

Lemma 3. *Call $\bar{p}$ an **upward crossing** of D if $g(\bar{p}) = D$ and $g'(\bar{p}) > 0$. For any upward crossing $\bar{p}$ and a sufficiently small $\epsilon > 0$:*[10]

- *for $p \in (\bar{p} - \epsilon, \bar{p})$: $g(p) < D$, so $\mathbb{E}[\theta_h(p)] > \mathbb{E}[\theta_\ell(p)]$*
- *for $p \in (\bar{p}, \bar{p} + \epsilon)$: $g(p) > D$, so $\mathbb{E}[\theta_h(p)] < \mathbb{E}[\theta_\ell(p)]$*

Hence any $b \in \left(\mathbb{E}[\theta_\ell(p)], \mathbb{E}[\theta_h(p)]\right]$ enrolls h alone.

[9] A similar lemma could be written for "downward crossings," but these are not used in the analysis.

[10] "Sufficiently small" means $0 < \epsilon < p - \max\{p', 0\}$ where p' is the next-smallest upward crossing, if one exists.

Figure 5 demonstrates the key features of Proposition 5 graphically. In particular, panels (a) and (b) show the shape of the $g(p)$ function and its derivative, respectively. In panel (a), it is clear that $\bar{p}$ occurs where $g(\bar{p}) = D$; furthermore, it is the only (upward) crossing along the domain $(0, 2\delta]$ where 2δ is arbitrarily chosen. Panel (b) demonstrates that the slope of $g(p)$ is subtly changing, and eventually turns negative after δ.[11] Panel (c) shows the thresholds underlying the $g(p)$ function—again, one can see that $\bar{p}$ is determined where the two thresholds switch, which is where $g(p) = D$.

Proposition 5 (Sub-Pigouvian optimality)**.** *Suppose there exists at least one upward crossing of D in $(0, \delta)$. Let p^* be the largest such crossing. Write*

$$W_h(p) = \max\Big\{\mathbb{E}[w(\gamma_h, \Omega_h; p)],\, 0\Big\}.$$

In an h-only regime, the aggregate welfare is $W_h(p)$. Then:

(i) The crossing is the supremum of the separating regime: *$W_h(p)$ is continuous and weakly increasing on $[0, \delta]$, and strictly increasing wherever positive. Hence, the welfare attainable through h-only enrollment has supremum*

$$\sup_{\{p:\, p<\delta \,\wedge\, g(p)<D\}} W_h(p) = W_h(p^*),$$

Note that $\mathbb{E}[w(\gamma_h, \Omega_h; p^*)]$ is approached but not truly attained as a separating baseline is not possible at p^*.

(ii) Sub-Pigouvian optimality: *Pooling welfare $W_{joint}(p) = \int_0^p (\delta - r)\Big[N_\ell f_r(r; \gamma_\ell) + N_h f_r(r; \gamma_h)\Big]\, dr - \Omega_\ell - \Omega_h$ is also strictly increasing on $[0, \delta]$. As such, the best pooling policy is Pigouvian. The regulator's problem is therefore to choose the policy that maximizes over the choice set:*

$$W^* = \max\Big\{W_h(p^*),\, W_{joint}(\delta),\, 0\Big\}.$$

Furthermore, the separating regime weakly dominates if and only if

$$W_h(p^*) \geq \max\Big\{W_{joint}(\delta),\, 0\Big\},$$

with the domination strict if the inequality is strict. An equivalent statement is

$$\mathbb{E}[w(\gamma_h, \Omega_h; p^*)] \geq \max\Big\{\mathbb{E}[w(\gamma_h, \Omega_h; \delta)] + \mathbb{E}[w(\gamma_\ell, \Omega_\ell; \delta)],\, 0\Big\}.$$

[11] That $g'(p)$ turns negative implies that there may be a second (or more) crossing at higher prices. According to Proposition 5, a Pigouvian price dominates a market with $p > \delta$, so these later crossings are not relevant to the policy problem at hand.

The proposition above identifies three things. First, among all separating prices between 0 and δ, p^* is the best because welfare from h-only enrollment is increasing in price on that interval. Second, whether separation is actually optimal depends on whether the welfare gain from excluding ℓ outweighs the welfare cost of inducing less conservation from h by setting a lower price. Third, if a pooling regime dominates a separating regime, the optimal price is Pigouvian. Strictly speaking, when the separating regime is dominant, the regulator can only approach $W_h(p^*)$ because ℓ participates when price is literally p^*; she does this by choosing a price less than, but arbitrarily close to, p^*.

The non-Pigouvian result depends critically on transaction costs. Without $\Omega_\ell > \Omega_h$, the threshold ordering never switches, and separation is impossible at any price. It is the combination of $\Omega_\ell > \Omega_h$ and the difference in γs that generates the crossing, and hence the possibility of non-Pigouvian optimality in offset markets. Indeed, with common participation costs, ℓ can never be screened out. This is Proposition 6; in such a scenario, the regulator's choice set reduces to Pigouvian pooling or shutdown.

Proposition 6 (No separation without cost heterogeneity)**.** *If $\Omega_\ell = \Omega_h$, then $D = 0$ and $g(p) > 0$ for all $p \in (0, \delta]$ under Assumption 3. As such, $\mathbb{E}[\theta_\ell(p)] > \mathbb{E}[\theta_h(p)]$ at every price. No baseline admits h while excluding ℓ.*

Without asymmetric information, the regulator could simply exclude ℓ-types directly. The baseline is needed precisely because the regulator cannot observe γ_i, thus she can only use the uniform policy instruments (b, p) to induce self-selection. The upward crossing is the condition under which self-selection achieves what direct exclusion would achieve under full information (though with less overall conservation). Panel (d) in Figure 5 illustrates the case where the optimal welfare occurs at $p^* < \delta$. In particular, $W_h(p)$, $W_{\text{joint}}(p)$, and $W^*(p)$ show aggregate welfare during the h-only regime, the pooled regime, and the optimal welfare, respectively.

5.2 Comparative statics

Every parameter acts on the regulator's problem through at most two channels. The **selection channel** shifts p^*, the largest price at which ℓ can still be excluded; because W_h is strictly increasing in price, p^* caps how much conservation the regulator can achieve in a separating regime. The **conservation channel** measures the conservation given up when going from pooled Pigouvian policy to a separating h-only policy. This is all of the conservation that ℓ provides as well as the plots owned by h that have returns in the interval

Table 2: Comparative statics in two-type economy

	Parameters						
	Ω_ℓ	Ω_h	N_ℓ	N_h	γ_ℓ	γ_h	δ
	Sign of effect						
p^*	$+$	$-$	$-$	$+$	$+$	$-$	0
ΔW	$+$	$-$	?	?	?	?	$-$

Notes: This table summarizes the comparative statics reported in Propositions 7 and 8. Each cell has either a $+$, $-$, 0, or ?, which represents the sign of the effect that an increase in the column parameter has on the policy measure for that row. The row p^* is the separating price from Proposition 5. The effect of δ on p^* is 0 because the separating price is, by construction, unaffected by changes in δ. The row ΔW is the aggregate welfare gains from screening out the ℓ-type landowner, rather than the joint market which includes both types.

$[p^*, \delta]$.

The selection channel exists only because participation is costly. With $\Omega_\ell = \Omega_h$ the thresholds never cross, so p^* does not exist. In such a case, every comparative static below collapses to its conservation channel alone—which is the frictionless case studied in vBK and Bento et al. (2015). Only when participation costs are heterogeneous can the regulator open the selection channel.

5.2.1 The separating price

Because p^* is defined as the solution to $g(p^*) = D$, every comparative static comes from either raising the hurdle D or changing how fast $g(p)$ climbs toward the upward crossing. Anything that raises D or lowers $g(p)$ pushes the crossing out to a higher price; anything that brings the thresholds toward each other decreases that price. Since $W_h(p)$ is increasing in price on the separating region, p^* is a ceiling on how much conservation the regulator can incentivize while still excluding ℓ.

Let ξ be any relevant parameter. The comparative statics around p^* are found by invoking the implicit function theorem:

$$\frac{\partial p^*}{\partial \xi} = \frac{\partial D/\partial \xi - \partial g(p)/\partial \xi}{g'(p)}\Big|_{p=p^*} \tag{9}$$

and $g'(p^*) > 0$ by Assumption 3, so each sign is determined solely by the numerator.

Proposition 7 (Effects on the separating price)**.** *Under Assumptions 1–3,*

$$\frac{\partial p^*}{\partial \Omega_\ell} > 0, \qquad \frac{\partial p^*}{\partial \Omega_h} < 0, \qquad \frac{\partial p^*}{\partial N_\ell} < 0, \qquad \frac{\partial p^*}{\partial \gamma_\ell} > 0, \qquad \frac{\partial p^*}{\partial \gamma_h} < 0, \qquad \frac{\partial p^*}{\partial \delta} = 0.$$

Participation costs. The gap D is the only channel through which participation costs affect the separating price. The regulator separates at prices where ℓ's cost disadvantage outweighs his threshold advantage. Raising Ω_ℓ widens that gap but has no effect on $g(p)$. The result is that the upward crossing occurs at a higher price. The friction that makes ℓ harmful is the same friction that makes him excludable. Raising Ω_h does the reverse.

Parcel shares. Because an increase in N_ℓ comes necessarily with a decrease in N_h, the total effect of increasing the share of plots owned by ℓ is composed of not only the direct effect from ℓ owning more land, but also the indirect effect from decreasing h's holdings. When the share of parcels owned by ℓ increases, the crossing decreases so that $g(p)$ intersects D at a lower price: p^* decreases. An economy with more non-additional land therefore forces screening at a price that buys less conservation. The two pieces respond differently to price—the baseline piece scales linearly in p, the forgone-earnings piece nonlinearly—so a shift in composition cannot generally be undone by a price adjustment.

Average returns. When γ_ℓ rises, ℓ begins to incur greater opportunity costs because more of his plots' returns enter the $(0, p]$ interval. As such, some non-additional parcels become additional, increasing his opportunity cost of participation. This additional opportunity cost reduces ℓ's baseline threshold, making ℓ more likely to self-select out of the market, even at higher prices. The regulator can encourage more conservation with a higher price while still keeping ℓ out of the market. As γ_h rises, h has fewer contractible plots and his opportunity cost of participating increases. Since separation requires h to be more willing than ℓ, anything that lowers h's expected threshold forces a lower separating price.

In general, the effects of a uniform (additive) shift in both γ_ℓ and γ_h do not cancel. From the derivations in Appendix B,

$$\left|\frac{\partial p^*}{\partial \gamma_i}\right| = \left.\frac{N_i}{g'(p)}\right|_{p=p^*} \cdot \Pr\Big[r_i \in (0, p^*]\Big], \quad i \in \{\ell, h\}. \tag{10}$$

The average returns matter proportionally to the number of plots a type holds near the conservation margin. Because $\gamma_\ell \ll 0 < \gamma_h < \delta$, h holds a much greater number of plots near the margin. Hence, a uniform upward shift in both types' returns lowers p^* on net.[12]

[12] This is true so long as

$$N_\ell < \frac{\Pr[r_h \in (0, p^*]]}{\Pr[r_\ell \in (0, p^*]] + \Pr[r_h \in (0, p^*]]}.$$

The value of conservation. Neither $g(p)$ nor D involves δ, so the separating price is invariant to the value of conservation; separating depends entirely on the incentives of a policy. All of δ's effect runs through the conservation channel.

5.2.2 The gain from separating

Let $\Delta W \coloneqq W_h(p^*) - W_{\text{joint}}(\delta)$ denote the welfare gain from separating rather than pooling at the Pigouvian price:

$$\Delta W = -\underbrace{N_h \int_{p^*}^{\delta} (\delta - r) f_r(r; \gamma_h)\, \mathrm{d}r}_{h\text{-conservation sacrificed}} - \underbrace{N_\ell \int_{0}^{\delta} (\delta - r) f_r(r; \gamma_\ell)\, \mathrm{d}r}_{\ell\text{-conservation forgone}} + \Omega_\ell. \tag{11}$$

Removing ℓ forgoes all the additional conservation he would have provided. Because removing ℓ requires the separating price, it requires forgoing the conservation of all the h-owned plots with returns in the interval $(p^*, \delta]$. Avoiding Ω_ℓ is a pure welfare gain.

For any parameter ξ,

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\xi} = N_h(\delta - p^*) f_r(p^*; \gamma_h) \cdot \underbrace{\frac{\partial p^*}{\partial \xi}}_{\substack{\text{selection} \\ \text{channel}}} + \underbrace{\frac{\partial \Delta W}{\partial \xi}\Big|_{p=p^*}}_{\substack{\text{conservation} \\ \text{channel}}} \tag{12}$$

Proposition 8 (Effects on the gain from separating)**.** *Under Assumptions 1–3,*

$$\frac{\partial \Delta W}{\partial \Omega_\ell} > 0, \qquad \frac{\partial \Delta W}{\partial \Omega_h} < 0, \qquad \frac{\partial \Delta W}{\partial \delta} < 0,$$

while $\frac{\partial \Delta W}{\partial N_\ell}$ *and* $\frac{\partial \Delta W}{\partial \gamma_\ell}$ *are ambiguous in general. Moreover,* $\frac{\partial \Delta W}{\partial \gamma_h} < 0$ *whenever* $\gamma_h \leq p^*$.

Where both channels agree. Raising Ω_ℓ improves the case for separating unambiguously. It raises the deadweight loss avoided one-for-one, and it raises p^*, so the regulator sacrifices less of h's conservation. This double role is unavailable in models where non-additionality is costly through transfers, because there the harm scales with the payment rather than with a parameter the regulator can screen on.

Raising δ unambiguously reduces the welfare gains from separating. When conservation is worth more, screening ℓ out costs more because the value of the forgone conservation increases.

Where one channel vanishes. Because h pays his participation cost under both regimes, raising Ω_h has no direct impact on the welfare gains. The entire effect goes through the selection channel. Increasing Ω_h lowers p^*, decreasing the conservation in the separating regime.

Where the channels oppose. The remaining effects are ambiguous.

Raising γ_ℓ decreases ℓ's participation threshold, increasing p^*. At the same time, it converts some of his non-additional plots into additional conservation under the Pigouvian pooling regime, increasing the forgone conservation. Thus, the selection channel is positive while the conservation channel is negative. The result is that screening becomes both easier and less valuable at the same time. This trade-off arises only when selection is price-dependent.

Increasing N_ℓ (decreasing N_h) lowers p^*, creating less conservation in the separating regime. At the same time, because h now owns less land, the forgone conservation is smaller. The direct channel is $\int_{p^*}^{\delta}(\delta - r) f_r(r;\gamma_h)\,\mathrm{d}r - \int_0^{\delta}(\delta - r) f_r(r;\gamma_\ell)\,\mathrm{d}r$, which is positive only when the social value of h conserving the plots between p^* and δ outweighs the social value of all of ℓ's additional conservation in the Pigouvian pooling regime. Under Assumption 3, this is the typical case. The net effect then depends on whether that gain exceeds the negative selection channel.

5.3 Examples

Example 3 (Non-Pigouvian optimal policy)**.** Let $\delta = 0.5$. Consider an economy with two landowners ℓ and h who have returns of $[-0.1, -0.1, -0.1, -0.1]$ and $[-0.1, -0.1, 0.1, 0.4]$ and fixed costs of $\Omega_\ell = 0.5$ and $\Omega_h = 0.1$, respectively. Adverse selection can be mitigated in this case, and, in fact, the optimal policy sets price below the value of the externality.

One can verify that Assumption 3 is satisfied:

- $\gamma_\ell < 0 < \gamma_h < \delta$: $\gamma_\ell = -0.1$ and $\gamma_h = 0.075$, so this assumption is satisfied.
- $\Omega_\ell = 0.5 > 0.1 = \Omega_h$: This is satisfied.
- $\Omega_\ell > N_\ell \int_0^{\delta}(\delta - r) f_r(r;\gamma_\ell)\,\mathrm{d}r$: Instead of taking the integral over the distribution, we can calculate the welfare contribution exactly. Because there are no plots with returns between 0 and δ, the righthand side is 0. Thus, this condition is satisfied.

The two thresholds cross at $p = 0.3$. Because $p = 0.3$ will include ℓ as well as h, the price must be set just below 0.3. In this example, the welfare contribution of h is a constant $w_h(p) = 0.3$ for and $p \in [0.1, 4)$, so any price between 0.1 and 0.3 will efficiently screen the market. On the screened section of the market, the aggregate welfare is equal to the welfare contribution of h alone. At $p \geq 0.3$, the aggregate welfare is the sum of both types.

This example illustrates that if the market is in the pooling regime (which includes the Pigouvian price), it is better to shut down the market. However, welfare can be increased, and actually turned positive, with $p < \delta$.

■

The key to Example 3 is that the order of the thresholds changes at some price between δ and $\delta/2$, so halving the price results in an unambiguous decrease in additional avoided deforestation from h. In a world with no transaction costs, this always decreases welfare. When participation is costly, however, the costs enter aggregate welfare. As such forgone conservation is not unambiguously welfare decreasing if society benefits by keeping ℓ out of the market. In this way, the "cost" of keeping ℓ out is the additional environmental benefit that would have been gained by preserving h's highest-return land. In Example 3, that environmental benefit ($\delta - 0.4 = 0.1$) was less than ℓ's participation cost ($\Omega_\ell = 0.5$).

Excluding types at the cost of decreased conservation might sound like a high price. However, as price decreases, the order in which h chooses plots to conserve or deforest is increasing in their welfare contribution (assuming they are welfare positive). In other words: as the price falls, h stops conserving his lowest-welfare plots first. There is a form of social sorting occurring that naturally increases welfare at the highest rate as price decreases.

6 Conclusion

Offset markets are usually evaluated on how accurately they measure a counterfactual. This paper argues that measurement is only half of the design problem. Who participates matters as much as how much conservation they provide. Furthermore, the friction governing entry is the fixed, privately known cost of registering, validating, and verifying a project, which is not merely a drag on the gains from trade—it can reverse their sign.

The modeling framework extends van Benthem and Kerr (2013) by giving each landowner a second hidden characteristic: a participation cost that is absorbed by the certification process, benefiting no one. Under either friction alone, the market is weakly welfare-improving.

Without participation costs, every parcel induced into conservation at a price no greater than the externality value contributes non-negatively. Thus, adverse selection shrinks the gains arbitrarily but cannot make them negative. With no information asymmetry, a type-contingent baseline and an eligibility rule can deliver the first-best outcome. It is only under a combination of these frictions that aggregate welfare turns negative, allowing market closure to be the dominant choice of the regulator.

The mechanism is a division of labor between the two frictions. Asymmetric information determines who is drawn into the market while transaction costs determine how much welfare their participation creates. Because a landowner's participation threshold falls as his land becomes more commercially valuable, the participation set is always a lower interval in γ. That is the additionality problem summarized: those who are the most non-additional are the hardest to exclude. When participation is costly, these landowners act as pure welfare reducers.

Three design results follow. First, unless the price is fixed to the externality value, the optimal baseline is generally not the BAU conservation of the marginal landowner. This is because it acts as a screening mechanism rather than a transfer size governor, excluding landowners whose parcels are commercially valuable enough, or whose participation is sufficiently costly, that their participation generates less social surplus than their costs. When participation costs are heterogeneous, the baseline changes from being a type-selector to a cost intensity control that thins participation across all γ-types at once. The optimality condition here equates the average welfare contribution of the marginal crowd to zero. Second, the optimal price may lie below the Pigouvian benchmark. Participation costs enter a landowner's threshold as $\frac{\Omega_i}{p}$ while the surplus on induced conservation rises with p, so the ordering of thresholds across types can switch as price varies. Pricing just below such a crossing can separate welfare-positive from welfare-negative types, which can be welfare-improving. Third, when both frictions are severe, market viability fails at every price, and shutdown is optimal.

These results speak directly to the current policy agenda. Most of that effort—remote sensing, machine learning, stricter validation—is aimed at the counterfactual, and the model takes no position against it: predicting B_i more accurately lets the regulator place the baseline where she intends. But accuracy operates on only one of the two hidden dimensions. A regulator with a perfect counterfactual model still does not observe the joint distribution of returns and participation costs among the landowners sitting at the participation margin, and it is that distribution which determines whether the marginal crowd is worth retaining. Baseline accuracy is necessary but not sufficient. Two complements follow: lowering the

fixed cost of certification, which shrinks the pure loss and thins the welfare-negative tail directly, and treating the baseline as a screening instrument rather than as a device for setting transfers.

The assumptions here are deliberately conservative. The regulator has an unlimited budget and purchases offsets for their conservation value rather than surrendering them against a cap, which switches off both channels through which non-additional credits are conventionally harmful. Relaxing either strengthens the case for closure rather than weakening it. Several extensions remain open—a spatially varying and uncertain δ, graduated payment schedules, and competitive procurement—but the first-order question is empirical. The conditions under which an offset market is welfare-reducing are stated here as restrictions on the joint distribution of returns and participation costs among participants. Whether the world's largest offset programs satisfy them is a question this framework makes answerable, and one worth answering before the next round of reform.

References

Araujo, R., Costa, F., & Sant'Anna, M. (2026). Efficient conservation of the Brazilian Amazon: Estimates from a dynamic model. *Review of Economic Studies*, *93*(1), 72–105. https://doi.org/10.1093/restud/rdaf031 1.

Armstrong, M. (1996). Multiproduct Nonlinear Pricing. *Econometrica*, *64*(1), 51. https://doi.org/10.2307/2171924 13, 25.

Aronoff, D., & Rafey, W. (2026). Conservation priorities and environmental offsets: Markets for Florida wetlands. *American Economic Review*, *116*(5), 1723–1764. https://doi.org/10.1257/aer.20231016 1.

Aspelund, K. M., & Russo, A. (2026). Additionality and asymmetric information in environmental markets: Evidence from conservation auctions. *American Economic Review*, *116*(8), 2765–2805. https://doi.org/10.1257/aer.20250295 4.

Badgley, G., Freeman, J., Hamman, J. J., Haya, B., Trugman, A. T., Anderegg, W. R. L., & Cullenward, D. (2022). Systematic over-crediting in California's forest carbon offsets program. *Global Change Biology*, *28*(4), 1433–1445. https://doi.org/10.1111/gcb.15943 5.

Bagnoli, M., & Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, *26*(2), 445–469. https://doi.org/10.1007/s00199-004-0514-4 18.

Baudry, M., Faure, A., & Quemin, S. (2021). Emissions trading with transaction costs. *Journal of Environmental Economics and Management*, *108*, 102468. https://doi.org/10.1016/j.jeem.2021.102468 6.

Bento, A. M., Kanbur, R., & Leard, B. (2015). Designing efficient markets for carbon offsets with distributional constraints. *Journal of Environmental Economics and Management*, *70*, 51–71. https://doi.org/10.1016/j.jeem.2014.10.003 5, 6, 13, 27, 32.

Calel, R., Colmer, J., Dechezleprêtre, A., & Glachant, M. (2025). Do carbon offsets offset carbon? *American Economic Journal: Applied Economics*, *17*(1), 1–40. https://doi.org/10.1257/app.20230052 1.

Cesa-Bianchi, N., Colomboni, R., & Kasy, M. (2025). Adaptive Maximization of Social Welfare. *Econometrica*, *93*(3), 1073–1104. https://doi.org/10.3982/ECTA22351 13.

Dionne, G., & Doherty, N. (1992). Adverse selection in insurance markets: A selective survey. In G. Dionne (Ed.). J. D. Cummins (Ser. ed.), *Contributions to Insurance Economics* (pp. 97–140, Vol. 13). Springer Netherlands. https://doi.org/10.1007/978-94-017-1168-5_4 6.

Einav, L., & Finkelstein, A. (2011). Selection in Insurance Markets: Theory and Empirics in Pictures. *Journal of Economic Perspectives*, *25*(1), 115–138. https://doi.org/10.1257/jep.25.1.115 6.

Ferraro, P. J. (2008). Asymmetric information and contract design for payments for environmental services. *Ecological Economics*, *65*(4), 810–821. https://doi.org/10.1016/j.ecolecon.2007.07.029 6, 7, 13.

Fowlie, M., Cullenward, D., Burtraw, D., Holt, B., Roedner Sutter, K., & Kerstein, H. (2024). *Annual report of the independent emissions market advisory committee*. California Environmental Protection Agency. https://calepa.ca.gov/2024-iemac-annual-report/ 1.

Hahn, R. W., Hendren, N., Metcalfe, R. D., & Sprung-Keyser, B. (2026). A welfare analysis of policies impacting climate change. *American Economic Review*, *116*(7), 2368–2421. https://doi.org/10.1257/aer.20250166 2.

Jaraitė-Kazukauskė, J., & Kazukauskas, A. (2015). Do transaction costs influence firm trading behaviour in the European Emissions Trading System? *Environmental and Resource Economics*, *62*(3), 583–613. https://doi.org/10.1007/s10640-014-9831-7 2, 6.

Kim, S. M., & Sanford, L. (2026). Evaluating carbon offset projects using machine learning. *Working paper*. https://www.dropbox.com/scl/fi/xqapd7tshm6pxiyjt0a64/Offset_Ver241211.pdf?rlkey=8ntm7f7b5xydthnewenkg31pm&st=9tavnndg&e=1&dl=0 5.

Mason, C. F., & Plantinga, A. J. (2013). The additionality problem with offsets: Optimal contracts for carbon sequestration in forests. *Journal of Environmental Economics and Management*, *66*(1), 1–14. https://doi.org/10.1016/j.jeem.2013.02.003 6, 7, 13.

Naegele, H. (2018). Offset credits in the EU ETS: A quantile estimation of firm-level transaction costs. *Environmental and Resource Economics*, *70*(1), 77–106. https://doi.org/10.1007/s10640-017-0111-1 2, 6, 25.

Pearson, T. R. H., Brown, S., Sohngen, B., Henman, J., & Ohrel, S. (2014). Transaction costs for carbon sequestration projects in the tropical forest sector. *Mitigation and Adaptation Strategies for Global Change*, *19*(8), 1209–1222. https://doi.org/10.1007/s11027-013-9469-8 2, 6, 25.

Robinson, W. (2026). Transaction costs of emissions offsets: Evidence from California. *No. 26-003*. https://go.ncsu.edu/cenrep.wp.26.003 2, 25.

Rochet, J.-C., & Choné, P. (1998). Ironing, Sweeping, and Multidimensional Screening. *Econometrica*, *66*(4), 783. https://doi.org/10.2307/2999574 13, 25.

Singh, R., & Weninger, Q. (2017). Cap-and-trade under transactions costs and factor irreversibility. *Economic Theory*, *64*(2), 357–407. https://doi.org/10.1007/s00199-016-0991-2 6.

Stavins, R. N. (1995). Transaction costs and tradeable permits. *Journal of Environmental Economics and Management*, *29*(2), 133–148. https://doi.org/10.1006/jeem.1995.1036 6.

van Benthem, A., & Kerr, S. (2013). Scale and transfers in international emissions offset programs. *Journal of Public Economics*, *107*, 31–46. https://doi.org/10.1016/j.jpubeco.2013.08.004 1–3, 13, 36.

West, T. A. P., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., Frey, G. P., & Kontoleon, A. (2023). Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science*, *381*(6660), 873–877. https://doi.org/10.1126/science.ade3535 1.

World Bank. (2026). *State and Trends of Carbon Pricing 2026*. https://doi.org/10.1596/978-1-4648-2348-0 1.

Wunder, S., Börner, J., Ezzine-de-Blas, D., Feder, S., & Pagiola, S. (2020). Payments for environmental services: Past performance and pending potentials. *Annual Review of Resource Economics*, *12*(1), 209–234. https://doi.org/10.1146/annurev-resource-100518-094206 6.

A Proofs

Proof of Proposition 1.

Proof. (i). Fix any (b, p) with $0 < p \leq \delta$ and take $i \in \mathcal{P}(b, p)$. With $\Omega_i = 0$,

$$w_i(p) = \sum_{j:\, r_{ij} \in (0,p]} (\delta - r_{ij}).$$

Each plot in the sum satisfies $r_{ij} \leq p \leq \delta$, so $\delta - r_{ij} \geq \delta - p \geq 0$. Hence, $w_i(p) \geq 0$ is a sum over non-negative numbers, so 0 is a lower bound for $W(b, p)$. Note that this invokes no properties of the participation set, so it holds whether or not returns are privately known; hence $\sup_{(b,p)} W(b, p) \geq 0$.

(ii). Set $p = \delta$ and let the regulator admit i if and only if $w_i(\delta) \geq 0$. Because she observes $\left(\{r_{ij}\}_i, \Omega_i\right)$ she computes each $w_i(\delta)$, and this eligibility rule is implementable. The type-specific baselines allow her to implement the eligibility rule as $b = B_i$ and $p = \delta$ leave an eligible landowner with a payoff of exactly $\psi_i(\delta) - \Omega_i$, his welfare contribution. Transaction costs enter the eligibility rule only through the sign of $w_i(\delta)$, and, therefore, a welfare-negative type is never incentivized to enter the market. Aggregate welfare is then $\sum_{i:\, w_i(\delta) \geq 0} w_i(\delta) \geq 0$, with equality when no type qualifies (which coincides with shutdown).

(iii). The final clause is simply the contrapositive of (i) and (ii). That is, if $\sup_{(b,p)} W(b, p) < 0$ then *both* transaction costs and asymmetric information must be present. Existence is witnessed by example 1, in which every (b, p) that admits the welfare-positive types also admits a mass of low-γ, high-Ω landowners whose negative contributions dominate, so that $\sup_{(b,p)} W(b, p) < 0$. □

Proof of Lemma 1.

Proof. Continuity of the expected value follows from the continuity of f_r. The limit as $\gamma \to \pm\infty$ follows because as γ moves away from $(0, p]$ in either direction, the mass of plots in $(0, p]$ goes toward zero. Thus

$$N \int_0^p (\delta - r) f_r(r; \gamma)\, \mathrm{d}r \to 0.$$

Hence, $\mathbb{E}[w(\gamma, \Omega; p)] \to -\Omega$.

If $\max_\gamma \mathbb{E}[w(\gamma, \Omega; p)] > 0$, continuity and the limits together imply that there is at least one point where $\mathbb{E}[w(\gamma, \Omega; p)] = 0$. The unimodality that comes from the log-concavity of f

means there is a single peak, which must be above zero. Call this peak $\tilde{\gamma}$. $\mathbb{E}[w(\gamma, \Omega; p)]$ is thus monotonic, continuous, and increasing to the left of $\tilde{\gamma}$ and decreasing to the right. Hence, there is a unique zero-crossing on either side; this means there are only two zero-crossings, with one strictly less than the other. □

Proof of Lemma 2.

Proof. Define

$$\Psi(\gamma; b, p) \coloneqq \psi(\gamma; p) F_\Omega(\bar{\Omega}(\gamma; b, p)) - \int_0^{\bar{\Omega}(\gamma;b,p)} \Omega f_\Omega(\Omega)\, \mathrm{d}\Omega.$$

Then, the aggregate welfare function is

$$W(b, p) = I \int_{-\infty}^{\bar{\gamma}} \Psi(\gamma; b, p) f_\gamma(\gamma)\, \mathrm{d}\gamma,$$

where Ψ is the expected welfare contribution weighted by participation probability, minus the total cost paid. Both Ψ and the upper limit $\bar{\gamma}$ depend on b; the baseline enters $\bar{\Omega}$ linearly through (5), so $\frac{\partial \bar{\Omega}}{\partial b} = -p$ for every γ.

Differentiation under the integral sign is justified as follows. Since $F_\Omega \leq 1$ and $\psi(\gamma; p)$ is bounded uniformly in γ—the integral defining it runs over the compact interval $[0, p]$, on which f_r is bounded under Assumption 1—we have $|\Psi(\gamma; b, p)| \leq \sup_\gamma |\psi(\gamma; p)| + \Omega_{\max}$ and $|\partial\Psi/\partial b| \leq p\Big(\sup_\gamma |\psi(\gamma; p)| + \Omega_{\max}\Big) \max f_\Omega$, both independent of b. Multiplied by f_γ these are integrable envelopes, so dominated convergence applies.

$$\frac{\partial W(b, p)}{\partial b} = I\Psi(\bar{\gamma}; b, p) f_\gamma(\bar{\gamma}) \frac{\partial \bar{\gamma}}{\partial b} + I \int_{-\infty}^{\bar{\gamma}} \frac{\partial \Psi(\gamma; b, p)}{\partial b} f_\gamma(\gamma)\, \mathrm{d}\gamma.$$

However, observe that the expected welfare contribution is zero for the highest-γ types because, by definition, $\bar{\Omega}(\bar{\gamma}; b, p) = 0$. Thus, $\Psi(\bar{\gamma}; b, p)$ vanishes:

$$\Psi(\bar{\gamma}; b, p) = \psi(\bar{\gamma}; p) \cdot F_\Omega(0) - \int_0^0 \omega\, f_\Omega(\omega)\, \mathrm{d}\omega = \psi(\bar{\gamma}; p) \cdot 0 - 0 = 0.$$

Assumption 2 ensures that $F_\Omega(0) = 0$, and the cost integral is over a degenerate interval. Intuitively, no mass at the cutoff type actually participates, so shifting $\bar{\gamma}$ has no first-order effect.

Substituting $\frac{\partial \bar{\Omega}}{\partial b} = -p$, we have

$$\begin{aligned}\frac{\partial W(b,p)}{\partial b} &= I \int_{-\infty}^{\bar{\gamma}} \frac{\partial \Psi(\gamma; b, p)}{\partial b} f_\gamma(\gamma)\,\mathrm{d}\gamma \\ &= I \int_{-\infty}^{\bar{\gamma}} \Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big] f_\Omega\Big(\bar{\Omega}(\gamma;b,p)\Big) \frac{\partial \bar{\Omega}(\gamma;b,p)}{\partial b} f_\gamma(\gamma)\,\mathrm{d}\gamma \\ &= -pI \int_{-\infty}^{\bar{\gamma}} \Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big] f_\Omega\Big(\bar{\Omega}(\gamma;b,p)\Big) f_\gamma(\gamma)\,\mathrm{d}\gamma,\end{aligned}$$

which is (6). □

Proof of Corollary 2.1.

Proof. By Lemma 2, $\frac{\partial W}{\partial b} = -pI\,Z\,\mathbb{E}_h\Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big]$, where $Z = \int_{-\infty}^{\bar{\gamma}} f_\Omega\Big(\bar{\Omega}(\gamma;b,p)\Big)\, f_\gamma(\gamma)\,\mathrm{d}\gamma > 0$ is the normalizer defining h. Since $pI\,Z > 0$, the sign of $\partial W/\partial b$ is the opposite of the sign of $\mathbb{E}_h\Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big]$. Hence increasing the baseline is aggregate welfare-improving if and only if the marginal participants are welfare-negative on average. □

Proof of Proposition 3.

Proof. **First, I show that a maximum exists.**
Write $W(b,p) = I \int_{-\infty}^{\bar{\gamma}(b,p)} \Psi(\gamma;b,p) f_\gamma(\gamma)\,\mathrm{d}\gamma$ with Ψ as in the proof of Lemma 2. Because $\psi(\gamma;p)$ is bounded, $F_\Omega \le 1$, and $\int_0^{\bar{\Omega}} \Omega f_\Omega(\Omega)\,\mathrm{d}\Omega$ is also bounded. Since $f_\gamma \in [0,1]$, $|\Psi| f_\gamma$ is a bounded function whose bounds are determined independent of b. At the same time, both Ψ and the limit $\bar{\gamma}(b,p)$ are continuous in b on $(0,N)$, and the limits as $b \to 0$ and $b \to N$ exist: $W(0,p)$ and 0, respectively. Hence, $W(b,p)$ is continuous on the compact interval $[0,N]$ and attains a maximum there by Weierstrass's theorem.

I now characterize the maximizer.
At an interior maximizer $\frac{\partial W}{\partial b} = 0$. It follows from Lemma 2 that this satisfies

$$\int_{-\infty}^{\bar{\gamma}} \Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big] f_\Omega\Big(\bar{\Omega}(\gamma;b,p)\Big) f_\gamma(\gamma)\,d\gamma = 0. \tag{FOC}$$

Note that $\int_{-\infty}^{\bar{\gamma}} f_\Omega(\bar{\Omega}(\gamma';b,p)) f_\gamma(\gamma')\,\mathrm{d}\gamma'$ is finite, since f_Ω is bounded and the integral of f_γ is bounded above by one and is strictly positive. Because $\mathbb{E}[\theta(\gamma,0;p)]$ is continuous and strictly decreasing on $(0,N)$, the set of types with $\bar{\Omega}(\gamma;b,p) \in (0,\Omega_{\max})$ is nonempty, and $f_\Omega(\bar{\Omega}) > 0$ by Assumption 2. Dividing FOC by this integral therefore preserves the condition. Now it

is expressed in terms of the marginal-participant's density h of Corollary 2.1:

$$\int_{-\infty}^{\bar{\gamma}} \Big[\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p)\Big]\, h(\gamma;b,p)\,\mathrm{d}\gamma = 0, \tag{$*$}$$

where h integrates to 1 over its support in γ, and is the density of landowners sitting exactly at the participation margin. Substituting the definition of $\psi(\gamma;p)$ and $\bar{\Omega}(\gamma;b,p)$, the difference inside the brackets becomes

$$\psi(\gamma;p) - \bar{\Omega}(\gamma;b,p) = \mathbb{E}_h\left[N(\delta-p)\int_0^p f_r(r;\gamma)\,\mathrm{d}r - pNF_r(0;\gamma)\right] + pb. \tag{$**$}$$

Because pb is constant in γ, substituting $(**)$ into $(*)$ gives

$$\begin{aligned}
0 &= \int_{-\infty}^{\bar{\gamma}} \left[N(\delta-p)\int_0^p f_r(r;\gamma)\,\mathrm{d}r - pNF_r(0;\gamma)\right] h(\gamma;b,p)\,\mathrm{d}\gamma + pb\underbrace{\int_{-\infty}^{\bar{\gamma}} h(\gamma;b,p)\,\mathrm{d}\gamma}_{=1} \\
-pb &= \int_{-\infty}^{\bar{\gamma}} \left[N(\delta-p)\int_0^p f_r(r;\gamma)\,\mathrm{d}r - pNF_r(0;\gamma)\right] h(\gamma;b,p)\,\mathrm{d}\gamma \\
b^* &= N\int_{-\infty}^{\bar{\gamma}} \left[F_r(0;\gamma) - \left(\frac{\delta-p}{p}\right)\int_0^p f_r(r;\gamma)\,\mathrm{d}r\right] h(\gamma;b,p)\,\mathrm{d}\gamma \\
b^* &= N\,\mathbb{E}_h\left[F_r(0;\gamma) - \left(\frac{\delta-p}{p}\right)\int_0^p f_r(r;\gamma)\,\mathrm{d}r\right]
\end{aligned}$$

which is (7). Because $h(\gamma;b,p)$ depends on b through the cost cutoff $\bar{\Omega}(\gamma;b,p)$, this is a fixed point characterization. That is, the optimal heterogeneous-cost baseline equals *its own* h-weighted average. At the Pigouvian price, the wedge $(\delta-p)/p$ vanishes, leaving $b^*(\delta) = N\,\mathbb{E}_h[F_r(0;\gamma)] = \mathbb{E}_h[B(\gamma)]$ by (1). □

Proof of Lemma 3.

Proof. Recall that

$$g(p) = p\;\mathbb{E}\Big[\theta(\gamma_\ell,0;p) - \theta(\gamma_h,0;p)\Big] \quad \text{and} \quad D = \Omega_\ell - \Omega_h.$$

Since $p > 0$, this gives the pointwise equivalence

$$\mathbb{E}[\theta_\ell(p)] \gtrless \mathbb{E}[\theta_h(p)] \iff g(p) \gtrless D, \tag{$*$}$$

where $\theta_i(p) = \theta(\gamma_i,\Omega_i;p)$.

Since F_r is continuous in p, the derivative $g'(p) = N_\ell F_r(p;\gamma_\ell) - N_h F_r(p;\gamma_h)$ is continuous.

Hence $g'(\bar{p}) > 0$ implies $g' > 0$ on a neighborhood of $\bar{p}$. Choose

$$0 < \epsilon < \bar{p} - \max\{p', 0\},$$

where p' is the largest upward crossing strictly below $\bar{p}$ (and $\epsilon < \bar{p}$ if no such p' exists). This guarantees g is strictly increasing on $(\bar{p} - \epsilon, \bar{p} + \epsilon)$ and that $\bar{p}$ is the only crossing of D in this interval.

Because g is strictly increasing on this interval and $g(\bar{p}) = D$:

- for $p \in (\bar{p} - \epsilon, \bar{p})$: $g(p) < g(\bar{p}) = D$, so by $(*)$ $\mathbb{E}[\theta_h(p)] > \mathbb{E}[\theta_\ell(p)]$; and
- for $p \in (\bar{p}, \bar{p} + \epsilon)$: $g(p) > g(\bar{p}) = D$, so by $(*)$ $\mathbb{E}[\theta_h(p)] < \mathbb{E}[\theta_\ell(p)]$.

Thus, the threshold ordering switches at $\bar{p}$ from h having the higher expected threshold just below to ℓ having it just above. In particular, for $p \in (\bar{p} - \epsilon, \bar{p})$ the interval $\Big(\mathbb{E}[\theta_\ell(p)], \mathbb{E}[\theta_h(p)]\Big]$ is nonempty, and any baseline b in it satisfies $b \leq \mathbb{E}[\theta_h(p)]$ (so h participates) and $b > \mathbb{E}[\theta_\ell(p)]$ (so ℓ does not), separating the two types. At $p = \bar{p}$, $(*)$ with $g(\bar{p}) = D$ gives $\mathbb{E}[\theta_h(\bar{p})] = \mathbb{E}[\theta_\ell(\bar{p})]$, so the separating interval collapses and no baseline separates the types. □

Proof of Proposition 5.

Proof. **(i) The crossing is the supremum of the separating regime.** Differentiating,

$$W_h'(p) = N_h(\delta - p) f_r(p; \gamma_h).$$

For $p \in (0, \delta)$ both $\delta - p > 0$ and $f_r(p; \gamma_h) > 0$ (Assumption 1), so $W_h'(p) > 0$: W_h is continuous and strictly increasing on $[0, \delta)$. The separating prices form the set $\{p < \delta : g(p) < D\}$. Because $\bar{p}$ is the largest upward crossing, $g(\bar{p}) = D$, so $\bar{p}$ is a boundary point of this set and is not contained in it. A continuous, strictly increasing function attains its supremum over such a set only in the limit, so

$$\sup_{\{p < \delta :\, g(p) < D\}} W_h(p) = \lim_{p \uparrow \bar{p}} W_h(p) = W_h(\bar{p}),$$

which is approached but not attained: there is no separating price at which welfare equals $W_h(\bar{p})$.

(ii) Sub-Pigouvian optimality. Pooling welfare is

$$W_{\text{joint}}(p) = N_\ell \int_0^p (\delta - r) f_r(r; \gamma_\ell)\, \mathrm{d}r + N_h \int_0^p (\delta - r) f_r(r; \gamma_h)\, \mathrm{d}r - \Omega_\ell - \Omega_h,$$

with $W'_{\text{joint}}(p) = (\delta - p)\Big[N_\ell f_r(p;\gamma_\ell) + N_h f_r(p;\gamma_h)\Big] > 0$ on $[0,\delta)$, so pooling welfare is maximized at the Pigouvian price $p = \delta$. Shutdown yields 0. The three mutually exclusive regimes—separate, pool, shut down—yield welfare $W_h(\bar{p})$, $W_{\text{joint}}(\delta)$, and 0 respectively, so the value of the regulator's problem is

$$W = \max\Big\{W_h(\bar{p}),\, W_{\text{joint}}(\delta),\, 0\Big\}.$$

The separating regime strictly dominates if and only if $W_h(\bar{p}) > \max\{W_{\text{joint}}(\delta), 0\}$, equivalently

$$\mathbb{E}[w(\gamma_h, \Omega_h; \bar{p})] > \max\Big\{\mathbb{E}[w(\gamma_h, \Omega_h; \delta)] + \mathbb{E}[w(\gamma_\ell, \Omega_\ell; \delta)],\, 0\Big\}.$$

When this strict inequality holds, $W = W_h(\bar{p})$ is a supremum attained by no policy: for every $\eta > 0$ there is $\epsilon_\eta > 0$ such that every $p \in (\bar{p} - \epsilon_\eta, \bar{p})$ yields $W_h(p) > W - \eta$, while the shortfall of $\max\{W_{\text{joint}}(\delta), 0\}$ below $W_h(\bar{p})$ is a fixed positive constant. Hence the optimal price is sub-Pigouvian. □

B Comparative statics derivations

Under Assumption 1, the following identities hold:

$$\frac{\partial f_r(r;\gamma)}{\partial \gamma} \equiv -\frac{\partial f_r(r;\gamma)}{\partial r}, \qquad \frac{\partial F_r(x;\gamma)}{\partial \gamma} \equiv -f_r(x;\gamma). \tag{B.1}$$

Lemma 4 (Integration-by-parts identities)**.** *For any $0 \le a < c$ and any γ, Assumption 1 implies*

$$-\int_a^c (c-r)\,\frac{\partial f_r(r;\gamma)}{\partial r}\,dr = (c-a)f_r(a;\gamma) - \int_a^c f_r(r;\gamma)\,dr, \tag{B.2}$$

$$\int_a^c (c-r)\,\frac{\partial f_r(r;\gamma)}{\partial r}\,dr = \int_a^c f_r(r;\gamma)\,dr - (c-a)f_r(a;\gamma). \tag{B.3}$$

Proof. Integrating by parts, $\int_a^c (c-r) f'_r(r;\gamma)\,\mathrm{d}r = \Big[(c-r)f_r(r)\Big]_a^c + \int_a^c f_r(r;\gamma)\,\mathrm{d}r = -(c-a)f_r(a;\gamma) + \int_a^c f_r(r;\gamma)\,\mathrm{d}r$. The boundary term is $0 - (c-a)f_r(a;\gamma)$, which gives (B.3); (B.2) is its negative. □

Proof of Proposition 7

Proof. Each sign follows from (9) with $g'(p^*) > 0$.

Ω_ℓ **and** Ω_h**.** Neither cost enters g, and $D = \Omega_\ell - \Omega_h$, so $\partial D/\partial\Omega_\ell = 1$ and $\partial D/\partial\Omega_h = -1$. Hence

$$\frac{\partial p^*}{\partial \Omega_\ell} = \frac{1}{g'(p^*)} > 0, \qquad \frac{\partial p^*}{\partial \Omega_h} = \frac{-1}{g'(p^*)} < 0.$$

Parcel shares. Using $N_h = 1 - N_\ell$ in (8),

$$\left.\frac{\partial g(p)}{\partial N_\ell}\right|_{p=p^*} = \underbrace{p^*\Big[F_r(0;\gamma_\ell) + F_r(0;\gamma_h)\Big]}_{\text{baseline channel}} + \underbrace{\int_0^{p^*} (p^* - r)\Big[f_r(r;\gamma_\ell) + f_r(r;\gamma_h)\Big]\,\mathrm{d}r}_{\text{forgone-earnings channel}} > 0,$$

and $\frac{\partial D}{\partial N_\ell} = 0$, so $\frac{\partial p^*}{\partial N_\ell} < 0$. Since $\frac{\partial g(p)}{\partial N_h} = -\frac{\partial g(p)}{\partial N_\ell}$, the reverse sign holds for N_h. The two bracketed channels scale differently in p^*. The first channel increases linearly while the second increases only through the truncated integral, which is the claim in the text.

Average returns. Differentiating (8) with respect to γ_ℓ and applying (B.1) and (B.2) with $a = 0$, $c = p^*$:

$$\left.\frac{\partial g(p)}{\partial \gamma_\ell}\right|_{p=p^*} = -p^* N_\ell f_r(0;\gamma_\ell) + N_\ell\left[p^* f_r(0;\gamma_\ell) - \int_0^{p^*} f_r(r;\gamma_\ell)\,\mathrm{d}r\right] = -N_\ell \int_0^{p^*} f_r(r;\gamma_\ell)\,\mathrm{d}r < 0.$$

The two terms in $p^* f_r(0;\gamma_\ell)$ cancel exactly: the baseline effect of a shift in γ is offset by its effect on the marginal plots. Hence $\partial p^*/\partial\gamma_\ell > 0$. The γ_h terms enter (8) with the opposite sign, so $\partial g/\partial\gamma_h = +N_h \int_0^{p^*} f_r(r;\gamma_h)\,\mathrm{d}r > 0$ and $\partial p^*/\partial\gamma_h < 0$. Dividing each by $g'(p^*)$ and writing $\int_0^{p^*} f_r(r;\gamma_i)\,\mathrm{d}r = \Pr[r_i \in (0, p^*]] + F_r(0;\gamma_i) - F_r(0;\gamma_i)$ gives (10).

The externality. δ appears in neither $g(p)$ nor D. □

Proof of Proposition 8

Proof. The selection channel is common to all parameters. Differentiating (11) with respect to p^*, the boundary term of the first integral gives

$$\frac{\partial \Delta W}{\partial p^*} = N_h(\delta - p^*) f_r(p^*;\gamma_h) > 0 \quad \text{for } p^* < \delta.$$

Ω_ℓ**.** The direct effect is $+1$ from (11). Adding the selection channel,

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\Omega_\ell} = 1 + \frac{N_h(\delta - p^*) f_r(p^*;\gamma_h)}{g'(p^*)} > 0.$$

Ω_h. Ω_h enters $W_h(p^*)$ and $W_{\text{joint}}(\delta)$ identically, so the direct effects cancel and

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\Omega_h} = N_h(\delta - p^*) f_r(p^*; \gamma_h) \cdot \left(\frac{-1}{g'(p^*)}\right) \; < \; 0.$$

δ. Since $\partial p^*/\partial \delta = 0$, only the direct channel operates. The integrand $(\delta - r)$ vanishes at $r = \delta$, so both boundary terms drop and

$$\frac{\mathrm{d}\Delta W}{\mathrm{d}\delta} = -N_h \int_{p^*}^{\delta} f_r(r; \gamma_h)\, \mathrm{d}r - N_\ell \int_0^{\delta} f_r(r; \gamma_\ell)\, \mathrm{d}r \; < \; 0.$$

N_ℓ. With $N_h = 1 - N_\ell$, the direct channel is

$$\left.\frac{\partial \Delta W}{\partial N_\ell}\right|_{p^*} = \int_{p^*}^{\delta} (\delta - r) f_r(r; \gamma_h)\, \mathrm{d}r \; - \; \int_0^{\delta} (\delta - r) f_r(r; \gamma_\ell)\, \mathrm{d}r,$$

which is positive whenever h's mass above p^* exceeds ℓ's mass in $(0, \delta)$. The selection channel is $N_h(\delta - p^*) f_r(p^*; \gamma_h) \cdot \partial p^*/\partial N_\ell < 0$. The total is therefore ambiguous.

γ_ℓ. Applying (B.1) then (B.3) with $a = 0$, $c = \delta$, the direct channel is

$$\left.\frac{\partial \Delta W}{\partial \gamma_\ell}\right|_{p^*} = N_\ell \int_0^{\delta} (\delta - r) \frac{\partial f_r(r; \gamma_\ell)}{\partial r}\, \mathrm{d}r = N_\ell \left[\int_0^{\delta} f_r(r; \gamma_\ell)\, \mathrm{d}r - \delta f_r(0; \gamma_\ell)\right] \; < \; 0,$$

where the sign follows because $\gamma_\ell < 0$ implies $f_r(\cdot; \gamma_\ell)$ is decreasing on $[0, \delta]$, so its average over the interval is below its value at the left endpoint. The selection channel is positive, so the total is ambiguous.

γ_h. Applying (B.3) with $a = p^*$, $c = \delta$, the direct channel is

$$\left.\frac{\partial \Delta W}{\partial \gamma_h}\right|_{p^*} = N_h \left[\int_{p^*}^{\delta} f_r(r; \gamma_h)\, \mathrm{d}r - (\delta - p^*) f_r(p^*; \gamma_h)\right],$$

which is negative if and only if $f_r(\cdot; \gamma_h)$ is decreasing on $[p^*, \delta]$, i.e. if and only if $\gamma_h \leq p^*$. The selection channel is $N_h(\delta - p^*) f_r(p^*; \gamma_h) \cdot \partial p^*/\partial \gamma_h < 0$. Under $\gamma_h \leq p^*$ both channels are negative, making the overall effect of increasing γ_h negative. Otherwise the total is ambiguous. □