Reallocation of Stakeholders: Exit, Engagement and the Measurement of Impact *

Briana Chang† and Harrison Hong‡

April 2026

Abstract

A purpose-driven stakeholder can leave a harmful firm for a cleaner one (exit) or stay and pay for mitigation (engagement), depending on what the market offers. We model these outside options with a matching model, where a rise in purpose reallocates stakeholders. Multiple stakeholder groups and harm that scales with firm size are needed to rationalize the coexistence of exit and engagement in the data. Engagement impacts a firm directly; exit spills over across firms. Because exit registers as noncompliance, a firm-level difference-in-differences understates the per-stakeholder mitigation effect of a rise in purpose; a compliance-adjusted LATE recovers it despite spillovers.

*We thank Yeon-Koo Che, Oliver Hart, Marcus Opp, Jan Starmans, Lasse Pedersen, David Lando, Navin Karthik, Qingmin Liu, Lily Fang, Alminas Zaldokas, Wenxi Jiang, Lou Dong, Kai Li, Adam Zhang, Deeksha Gupta, Morris Davis, Kerry Back, Bruce Carlin, Gustavo Grullon, Stephanie Johnson, Kunal Sachdeva, Yuhang Xing, and seminar participants at the Stockholm School of Economics, Copenhagen Business School, INSEAD, Columbia University, Rice University, SKKU, Peking University HSBC Business School Inaugural Finance Colloquium, HEC-HKUST Sustainable Finance Workshop 2024, Minnesota Corporate Finance Conference, the Baruch Sustainable Finance Conference, American Finance Association Meetings, Rutgers Business School, University of Zurich, FIRN Tasmania Conference and University of New South Wales for helpful comments. Hong acknowledges support from a Guggenheim Fellowship.

†Hong Kong University of Science and Technology

‡Columbia University

1 Introduction

When purpose-driven stakeholders press firms to reduce social harm, two different responses appear in the data. Some *exit*: banks divest from fossil-fuel borrowers, investors screen out offenders, employees leave. Others *engage*: they remain and subsidize firm mitigation. What determines which stakeholders do which—and why do both occur at once, in the same market, in response to the same rise in purpose? Intuitively, the choice between staying to reform a firm and leaving for a better-aligned one depends on what the rest of the market offers.[1] That outside option, however, is not exogenously given. It ought to be determined in equilibrium, set by how heterogeneous stakeholders sort across heterogeneous firms.

To make this precise, we develop a multi-sided matching model in which firms assemble teams from several stakeholder groups—workers, banks, suppliers—who differ in both productivity and purpose. Purpose-driven stakeholders derive disutility from the harm their firm generates,[2] so firms mitigate harm to attract and retain them. Each stakeholder's response is governed by an outside option that the matching market sets, and a shift in preferences moves those outside options all at once. The exit-versus-engagement margin therefore cannot be located firm by firm; it must be solved for in general equilibrium, in the market that prices stakeholders' alternatives. In our model, exit and engagement are not competing responses but two outcomes of a single sorting equilibrium.

Two features of this environment govern whether a rise in purpose produces exit, engagement, or both. The first is whether harm scales with firm size: when it does not, purpose operates through engagement alone. The second is whether there are multiple groups of stakeholders: with a single group, a preference shock simply re-ranks the affected stakeholders downward and all of them exit; a second group is what lets a firm retain a newly purpose-driven stakeholder at unchanged size. Remove either feature and the model collapses to a corner—engagement only, or exit only. The multi-sided matching structure is thus the minimal environment in which exit and engagement coexist.

Consider first the benchmark in which harm is independent of firm size. Sorting on productivity and sorting on purpose then separate: stakeholders and firms match on productivity, and, holding productivity fixed, a firm is indifferent between a profit-driven and a purpose-driven stakeholder, since the cost of mitigation does not depend on output. A rise in the share of purpose-driven stakeholders thus leaves the matching intact—affected stakeholders remain with firms of the same size while compensating differentials adjust—so impact operates entirely

[1] The choice echoes the classic tension between exit and voice (Hirschman 1970).

[2] We model purpose as disutility over firm-level harm, encompassing both warm-glow and non-consequentialist moral motives. This is the workhorse assumption in the literature on prosocial behavior (Andreoni 1990, Bénabou and Tirole 2006, Besley and Ghatak 2005) and has strong empirical support (Hong and Kacperczyk 2009, Riedl and Smeets 2017, Bonnefon et al. 2025).

through engagement, with no spillovers across groups.

The general case, where harm scales with production—as it does for pollution and many other social harms—is different. Firms are no longer indifferent between stakeholder types: because output generates harm, productivity is worth less inside a team with purpose-driven members, where additional output must be offset by costly mitigation. Sorting on productivity and on purpose become interdependent, and the matching problem becomes substantially harder to characterize.

Importantly, the engagement region exists only because stakeholders come in multiple groups. With a single group, a rise in purpose lowers every affected stakeholder's effective productivity by the same amount, so all of them re-rank downward and rematch with smaller firms: the response is exit, with no engagement anywhere. It is the second group that supplies the margin—the mixed teams—through which a firm retains a newly purpose-driven stakeholder. The multi-stakeholder structure is therefore the source of engagement itself.

We thus provide a sequential algorithm that constructs the equilibrium for an arbitrary number of stakeholder groups, reducing the multi-sided problem to a nested sequence of bilateral matches. The equilibrium sorts stakeholders on both productivity and purpose. Above a productivity threshold, large firms outbid one another for profit-driven stakeholders to avoid the mitigation their output would require, so high-productivity teams separate cleanly into "clean" firms staffed by purpose-driven stakeholders and "dirty" firms staffed by profit-driven ones. Below the threshold, full separation is infeasible, and some firms adopt moderate mitigation to attract productive purpose-driven stakeholders, forming "mixed" teams. These regions are what set outside options: at the top, a purpose-driven stakeholder's alternative is another already-clean firm, so persuading her own firm to mitigate is expensive; in the mixed region, a firm can absorb her at unchanged size, so it is cheap.

These outside options determine the response to a rise in purpose. For high-productivity stakeholders, inducing a large firm to mitigate would require prohibitive transfers, so they exit, reallocating toward smaller, purpose-aligned firms. For lower-productivity stakeholders, whose firms are nearly indifferent across purpose types, the relationship is stable, so they engage—remaining in place and sharing the cost of additional mitigation. Exit occurs at the top of the productivity distribution, engagement at the bottom.

Exit and engagement have the same per-firm impact on harm but land on different firms: engagement lowers harm at the stakeholder's own firm, while exit lowers it elsewhere, as displaced stakeholders reallocate and induce mitigation at smaller firms — that is, the impact trickles down the productivity ladder. When harm scales with size, spillovers run not only down the productivity ladder but across stakeholder groups.[3] A rise in purpose-driven banks,

[3]Such reallocation spillovers echo those studied elsewhere in economics—agglomeration (Ellison and Glaeser 1997), firm entry (Greenstone, Hornbeck, and Moretti 2010), production networks (Acemoglu, Akcigit, and Kerr

for instance, raises the welfare of purpose-driven workers, who can now match with more productive purpose-driven banks, while profit-driven banks earn rents as productive firms bid for them. Such cross-group spillovers are absent under independence and are the second reason multiple stakeholder types must be modeled together.

This has a direct consequence for measurement. A firm-level difference-in-differences design to capture the per-stakeholder mitigation effect of a rise in purpose registers engagement but not exit, so it is unbiased when stakeholders engage, understates impact when they exit, and understates in proportion to the exit share when both occur. Exit is noncompliance in the language of the literature on local average treatment effects (LATE; Imbens and Angrist 1994, Angrist, Imbens, and Rubin 1996). We show that the per-stakeholder mitigation effect is recovered as a LATE whose first stage is the take-up of a purpose-driven member—an estimator robust to the reallocation spillover, unlike the interference that ordinarily contaminates such estimands (Chaisemartin and D'Haultfœuille 2018), precisely because here a spillover *is* the treatment relocating, entering first stage and outcome alike.

We validate our measurement implications using a calibrated model of the wave of bank decarbonization commitments following the 2015 Paris Agreement. Empirical work has found a null impact using a difference-in-differences design (Kacperczyk and Peydró 2022). Simulations confirm that LATE recovers the true effect across most of the firm-size distribution, failing only in the large-firm tail where exit relocates impact entirely off the treated set.

Related literature. The literatures closest to ours study stakeholder responses to social harm one group at a time, and so treat exit and engagement as alternatives—modeling one margin, or asking which of the two is more effective—never as outcomes that coexist. Modeling multiple groups together is what makes them coexist, and that coexistence is what turns a standing puzzle in the empirical divestment literature into a measurement result.

Socially responsible finance. This literature models a single stakeholder group, shareholders, whose only lever is exit: by divesting, purpose-driven investors raise a harmful firm's cost of capital. Its central question is how far arbitrageurs limit this effect by buying the very stocks that purpose-driven investors exit, absorbing the price pressure that divestment is meant to create (Berk and Binsbergen 2025, Heinkel, Kraus, and Zechner 2001, Hong, Wang, and Yang 2021, Pástor, Stambaugh, and Taylor 2021, Pedersen, Fitzgibbons, and Pomorski 2021, Hong and Kacperczyk 2009). Because there is only one group, there is no counterparty a firm can retain by mitigating, so engagement never arises as a distinct equilibrium margin. These models describe divestment but are silent on why, in the same market and facing the same harm, many stakeholders instead stay and pay for change.

2016), and macro-financial shocks (Huber 2023)—though here they operate through purpose-driven sorting in stakeholder–firm matching.

Exit versus engagement. A second, primarily normative literature adds engagement, arguing that exit, on its own, is an inferior or even counterproductive tool. Broccardo, Hart, Zingales, et al. 2022 show that voice dominates exit once shareholders are consequentialist and voting is cheap; Edmans, Levit, and Schneemeier 2022 show that blanket divestment can backfire, because it removes a firm's incentive to improve, so that holding and rewarding improvers dominates outright exit; Green and Roth 2025 argue that simply excluding brown firms and tilting toward green ones need not create impact, and can forgo the opportunities where capital matters most; and Oehmke and Opp 2023 show that impact-oriented capital can steer firms toward clean production even where commercial capital is abundant, because firm-level financing frictions let responsible capital fund a scale increase that profit-seeking investors will not. Common to all four is a single stakeholder group: each studies one lever, or ranks one against the other, but none has exit and engagement arise together in a single equilibrium. So none confronts the question that coexistence forces: how to measure the impact of a rise in purpose when some stakeholders exit and others engage at once.

Coexistence and the measurement of impact. Our contribution is positive rather than normative: it asks not which lever is better but what a rise in purpose does, and how to measure it. Answering that question inverts the received conclusion—exit is not ineffective but *mismeasured*, since its impact reallocates to other firms and registers as noncompliance at the firm left behind. This reconciles the null and mixed findings that firm-level designs report for divestment and stakeholder pressure (Kacperczyk and Peydró 2022, Duchin, Gao, and Xu 2022, Akey and Appel 2019, Hartzmark and Shue 2022), and, as we show in Section 6, the per-stakeholder effect is then recovered as a compliance-adjusted local average treatment effect that the reallocation spillover leaves intact.

Matching and sorting. Methodologically, our environment is a multi-sided matching model with transferable utility and two-dimensional heterogeneity—stakeholders differ in both productivity and purpose—and we use the optimal-transport equivalence between stable matching and a planner's problem (Galichon 2016). Most models collapse multidimensional traits into a single index (Sattinger 1979, Tervio 2008, Gabaix and Landier 2008) or, as in Boerma, Tsyvinski, and Zimin 2025, obtain multi-sided sorting from a submodular technology in one dimension; we instead retain a standard supermodular technology and let productivity and purpose interact endogenously through firm-level mitigation, which is what makes sorting on the two dimensions interdependent. Relative to bilateral multidimensional matching (Dupuy and Galichon 2014, Lindenlaub 2017, Chiappori, McCann, and Pass 2016, Chiappori, Oreffice, and Quintana-Domeque 2018), our methodological contribution is a tractable sequential algorithm that reduces the multi-sided problem to a nested sequence of bilateral matches, letting us trace how a purpose shock propagates across groups.

2 Model

2.1 Environment

Production and harm. There are N types of stakeholders and one firm, indexed by $\ell \in L \equiv \{1, 2, \ldots, N+1\}$. Each stakeholder type $\ell \leq N$ has skill $x_\ell \in X_\ell$, while the firm has productivity $x_{N+1} \in X_{N+1}$. All types have unit mass, with smoothly distributed skills on compact supports.

Output depends multiplicatively on all agents' characteristics:

$$y(\mathbf{x}) = \prod_{\ell=1}^{N+1} x_\ell,$$

where $\mathbf{x} = (x_1, \ldots, x_{N+1})$. As discussed in Tervio (2008), the linear assumption on the arguments does not preclude different stakeholders having different skills contributing to their ability to affect output.[4]

Relationship between production and harm. The relationship between harm and production is denoted by

$$\phi(y) = \phi_0 + \sigma y,$$

where $\sigma \geq 0$. When $\sigma = 0$, harm does not increase with production. When $\sigma > 0$, more production generates more environmental or social harm. For example, when harm takes the form of emissions, σ thus represents the emission rate of the firm per unit of production.

Stakeholder preferences. Each stakeholder matches to exactly one firm, capturing the notion of a bilateral relationship (e.g. a bank lending to one firm, or a worker employed at one firm). Stakeholders' utilities differ in whether they are purpose-driven. Purpose-driven stakeholders are modeled as having disutility over the harm, denoted by e, generated by the firm with which they have a relationship.

This warm-glow preference creates incentives for the stakeholder's matched firm to mitigate.[5] Let $\theta_\ell \in \{0, 1\}$ denote type, and the utility is given by

$$u(p, e \mid \theta_\ell) = p - \theta_\ell \psi(e), \tag{1}$$

where p is the transfer received, e is harm created by the hiring firm, and disutility $\psi(\cdot)$ is

[4]Specifically, one could interpret $x_\ell = b_\ell(\hat{x}_\ell)$, which represents the effective ability of some underlying skill $\hat{x}_\ell$, where b_ℓ is an increasing transformation of the scale of measurement for a factor quality. For example, a Cobb-Douglas production function $x_0 \hat{x}_1^{\alpha} \hat{x}_2^{(1-\alpha)}$ can be nested as $x_1 = \hat{x}_1^{\alpha}$ and $x_2 = \hat{x}_2^{1-\alpha}$.

[5]Given the preferences are on firm harm within the team, the matching problem is not subject to externalities.

increasing and convex in harm, $\psi'(e) > 0$ and $\psi''(e) > 0$. That is, profit-driven stakeholders ($\theta_\ell = 0$) care only about compensation. purpose-driven stakeholders ($\theta_\ell = 1$) also dislike the harm produced by their firm. Thus stakeholder types are $a_\ell = (x_\ell, \theta_\ell)$, distributed with measure μ_ℓ on $X_\ell \times \{0, 1\}$. The share of purpose-driven stakeholders of type ℓ is denoted by λ_ℓ.

Firms themselves are profit-maximizing and do not have non-profit-driven preferences, i.e. $a_{N+1} = (x_{N+1}, 0)$ and $\lambda_{N+1} = 0$. That is, firms are special in the sense that their characteristics are always one-dimensional ($\lambda_{N+1} = 0$).[6] Throughout, skills and purpose-driven preferences are assumed to be independently and identically distributed. We use this assumption to highlight that, if there is any correlation between productivity and mitigation, it will arise endogenously through the equilibrium matching.

Remark on disutility over harm. We have specified stakeholder disutility over the level of harm. An alternative formulation would have stakeholders experience disutility over the *rate* of harm—for instance, emissions per unit of output. In that case, one would replace $\psi(e)$ with $\psi(e/y)$, and the relevant condition for our results would be whether the harm rate is correlated with productivity. Our main results go through under either specification: what matters is that the dimension of harm stakeholders care about varies systematically with firm size. We adopt the level specification for tractability.

Assumption 1. *Skill and purpose are independently distributed. Conditional on skill x_ℓ, a stakeholder of type ℓ is purpose-driven with probability λ_ℓ.*

Lastly, we label stakeholder types such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. That is, types are ordered by the prevalence of purpose-driven stakeholders.

Mitigation. While firms produce harm, they can reduce harm by $m \geq 0$ at linear cost cm. Hence, the harm generated by a firm in our model is given by $\phi(y) - m$, which depends on two endogenous objects: his productivity (which depends on which stakeholders he employed) and his abatement policy.

Given any match, let $\mathbf{a} = (a_1, \ldots, a_{N+1})$ represent the characteristic of all agents within the team, and $\Lambda(\mathbf{a})$ represent the total surplus of the team. Since we have transferrable utilities, the mitigation within any match must be chosen to maximize the total surplus, which yields

$$\Lambda(\mathbf{a}) = \max_{m \geq 0} \; y(\mathbf{x}) - cm - n(\mathbf{a})\psi(\phi(y(\mathbf{x})) - m), \tag{2}$$

[6] We use this assumption to highlight that firms only mitigate when matching with purpose-driven stakeholders, and it must be profitable for them to do so. Our setup can easily be extended to allow firms to also care about harm.

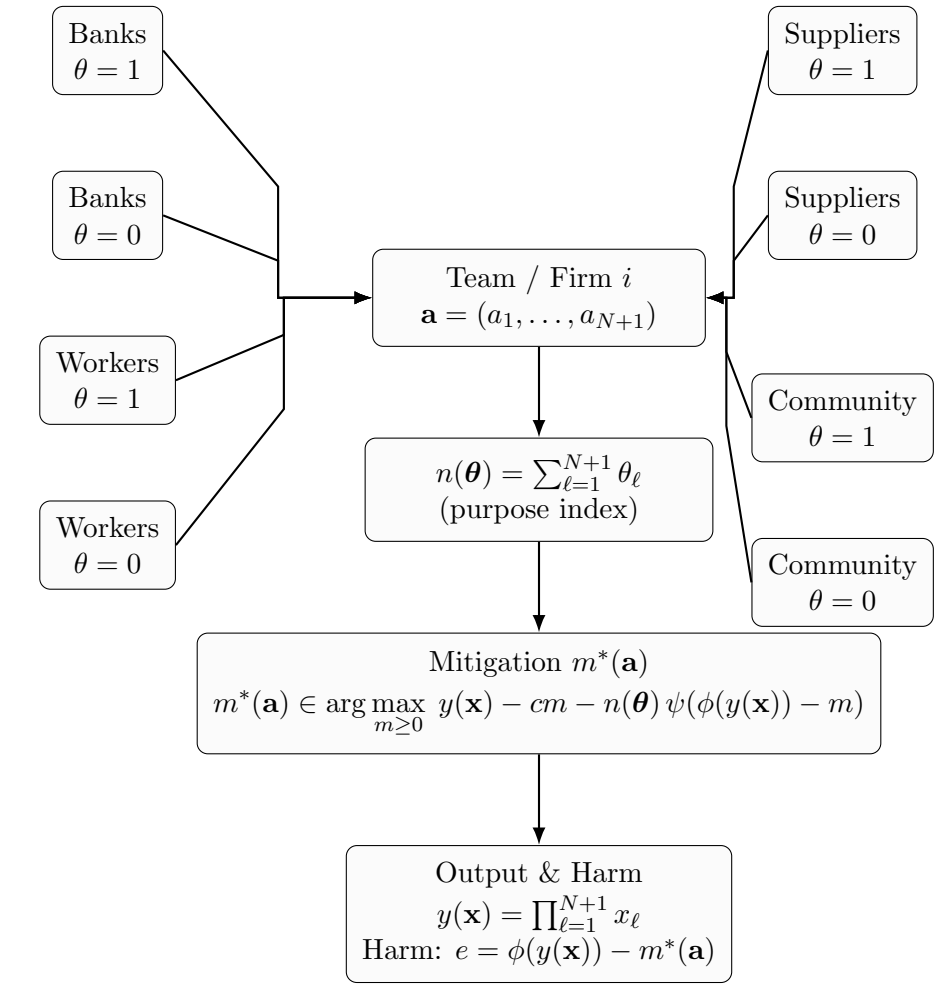


Figure 1: Extended schematic: banks and workers (left), suppliers and community (right) match with a firm, which determines the purpose index, mitigation, and resulting output and harm.

where $n(\mathbf{a}) \equiv \sum_{\ell=1}^{N+1} \theta_\ell$ denotes the *stakeholder-purpose index* of the team, i.e. the number of purpose-driven members within the team.

Stakeholder payoffs. For any stakeholder a_ℓ, let the utility be

$$U_\ell(a_\ell) = \max_{\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}} \Lambda(\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}, a_\ell) - \sum_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'}). \tag{3}$$

That is, the payoff to stakeholder a_ℓ is the team surplus minus the equilibrium utilities of all other members. The firm's problem is a special case of (3) with $\ell = N+1$. Figure 1 summarizes the flow of our model.

Competitive equilibrium. A competitive equilibrium consists of: (1) an allocation function γ that represents the matching between firm and all types of stakeholders, which is defined as a probability measure on the set of $(X_\ell \times \{0,1\})^{N+1}$, (2) a mitigation policy $m^*(\mathbf{a})$ for each team, and (3) equilibrium utilities $\{U_\ell(a_\ell)\}$ for all types $\forall \ell \in L$. These must satisfy:

- (Optimal mitigation) The mitigation policy $m^*(\mathbf{a})$ solves Equation (2).
- (Stable match) No matched agent can increase their utilities by forming a new team: for any $\mathbf{a} = (a_1, a_2 ... a_{N+1})$, $\Sigma_{\ell \in L} U_\ell(a_\ell) \geq \Lambda(\mathbf{a})$, where an equality obtains on the support of γ.
- (Market clearing) The marginal of γ should equal the marginal distribution of F_ℓ.

Matching is stable if no coalition can profitably deviate to form a new team. Accordingly, for any assignment $\mathbf{a}'$ outside the support of γ, the joint surplus generated by such a deviation, $\Lambda(\mathbf{a}')$, must be weakly lower than the sum of the members' equilibrium utilities. Following the matching literature with transferable utility, any stable outcome solves the planner's problem (see Boerma, Tsyvinski, and Zimin 2025). Therefore, the equilibrium maximizes aggregate surplus. Finally, given equilibrium utilities and the mitigation policy, transfers are uniquely pinned down for each stakeholder by Equation 1.

Remark on efficiency. Our setting does not explicitly feature externalities, since stakeholders suffer disutility over the harm generated by their matched firm. However, one could further assume that stakeholders also suffer disutility from aggregate emissions. This extension would not affect the equilibrium allocation, as stakeholders are infinitesimal and take aggregate emissions as given. However, it would generally imply that the equilibrium is no longer efficient.

2.2 General Properties

Observe that a team can be characterized by two sufficient statistics: its productivity and its stakeholder-purpose index. Specifically, from Equation 2, the surplus of any team can be written as $\Lambda(\mathbf{a}) = \Omega\left(\prod_{\ell=1}^{N+1} x_\ell, \sum_{\ell=1}^{N+1} \theta_\ell\right)$,where

$$\Omega(y, n) = \max_{m \geq 0} \Big\{ y - cm - n\psi(\phi(y) - m) \Big\}. \tag{4}$$

We impose the following assumptions to guarantee that mitigation is interior whenever $n \geq 1$, and that production is always valuable despite mitigation costs. Let $y_{min} \equiv y(\underline{\mathbf{x}})$ denote the lowest production, which is composed of the lowest skill of type ℓ,denoted by $\underline{x_\ell}$.

Assumption 2. *(i) (Interior mitigation)* $\psi'(\phi(y_{min})) > c$. *(ii) (Production valuable)* $1 - c\sigma > 0 \, \forall y$ *and* $\Omega(y_{min}, N) > 0$.

The first condition represents that, at the lowest productivity (and thus lowest possible harm), it is nevertheless optimal to have positive mitigation as long as there is one purpose-driven stakeholder. The second condition guarantees that $\Omega_y(y, n) > 0$, as for any $n \geq 1$,

$$\Omega_y(y, n) = 1 - n\psi'(\phi(y) - m^*(y, n))\phi'(y) = 1 - c\sigma, \tag{5}$$

where we use the fact the optimal mitigation $m^*(y, n)$ must satisfy the FOC condition, $n\psi'(\phi(y) - m^*(y, n)) = c$. Hence, taking into account the additional abatement cost generated by the production, production is always valuable.

Sorting on purpose. Observe that adding a purpose-driven stakeholder is costly ($\Omega_n(y, n) < 0$), but the marginal cost is decreasing ($\Omega(y, n)$ is strictly convex in n). This is because mitigation is non-rival and has an economic scale.

This implies clustering: purpose-driven agents prefer to join each other. Let $n^*_{-\ell}(x_\ell, \theta)$ be the index of the matching team of stakeholder (x_ℓ, θ), which consists of other types of stakeholders.

Lemma 1. *(Concentration of purpose-driven stakeholders) Consider two stakeholders of type* ℓ *with the same ability* ($x_\ell = x_\ell$)*, then*

$$n^*_{-\ell}(x_\ell, 1) \;\geq\; n^*_{-\ell}(x_\ell, 0).$$

That is, conditional on the skill, purpose-driven stakeholders must join a team with a higher purpose-index than the equivalent profit-driven stakeholders. Otherwise, one can switch these two agents, which results in a more extreme distribution of n but does not affect the output in each team, and such a deviation generates higher payoff as $\Omega(y, n)$ is convex in n.

Productivity–purpose interdependence. Next we study how productivity and purpose interact. Output always raises surplus ($\Omega_y(y, n) > 0$), but the marginal value of productivity can depend on the team's purpose index. Specifically, from Equation 5, for any $n \geq 1$,

$$\Omega_y(y, n) = 1 - c\sigma \geq \Omega_y(y, 0) = 1,$$

where the equality holds only when $\sigma = 0$. Intuitively, as long as higher output creates more harm ($\sigma > 0$), productivity becomes less valuable to teams with purpose-driven members than to purely profit-driven teams.

Let $(y_{-\ell}, n_{-\ell})$ denote the productivity and purpose-driven index of the team that excludes agent of type ℓ, and $V(y_{-\ell}, n_{-\ell})$ represent the total utility accruing to all members in the team.

The matching between the team $(y_{-\ell}, n_{-\ell})$ and the stakeholder (x_ℓ, θ_ℓ) can thus be rewritten as

$$U(x_\ell, \theta_\ell) = \max_{(y_{-\ell}, n_{-\ell})} \Omega(x_\ell y_{-\ell},\, n_{-\ell} + \theta_\ell) - V(y_{-\ell}, n_{-\ell}), \tag{6}$$

As standard, all else equal, a more productive stakeholder always generates higher team surplus and thus must obtain a higher market utility (i.e., $U(x_\ell, \theta)$ increases in x_ℓ). Under the market competition, it is thus the team that is most willing to pay for higher skill will end up hiring a higher skilled stakeholder. Specifically, the marginal contribution of a more productive profit-driven stakeholder $(x_\ell, 0)$ to the team $(y_{-\ell}, n_{-\ell})$ is given by

$$\frac{\partial}{\partial x_\ell}\Omega(y_{-\ell}x_\ell,\, n_{-\ell}) = \begin{cases} (1 - c\sigma)\, y_{-\ell} & n_{-\ell} \geq 1 \\ y_{-\ell} & n_{-\ell} = 0. \end{cases} \tag{7}$$

First of all, due to the standard complementary in production, fixing $n_{-\ell}$, a more productivity team $y_{-\ell}$ must value a high skilled worker more. Moreover, note that for any $\sigma > 0$, fixing the productivity $y_{-\ell}$, a profit-driven stakeholder's skill is relatively more valuable to a purely profit-driven team ($n_{-\ell} = 0$) than to the team that consists of purpose-driven stakeholders. This is because a higher productivity also means more abatement to the team with $n_{-\ell} \geq 1$.

The joint effect of productivity and purpose composition on the marginal return to profit-driven skill can thus be summarized by the index $z(y_{-\ell}, n_\ell)$ below,

$$z(y_{-\ell}, n_{-\ell}) \equiv \begin{cases} (1 - c\sigma) y_{-\ell} & n_{-\ell} \geq 1 \\ y_{-\ell} & n_{-\ell} = 0. \end{cases} \tag{8}$$

The z-index captures the effective productivity of a team from the perspective of a profit-driven stakeholder: it equals the team's raw productivity when the team is entirely profit-driven, but is discounted by the factor $(1 - c\sigma)$ when the team contains purpose-driven members, reflecting the additional mitigation cost that higher output entails.

Lemma 2. *(Positive assortative matching) (i) Among profit-driven stakeholders ($\theta_\ell = 0$), higher skill x_ℓ implies matching with a team with a higher effective productivity (z-index). (ii) Among purpose-driven stakeholders ($\theta_\ell = 1$), higher skill x_ℓ implies matching with a team with higher productivity $y_{-\ell}$.*

By monotone comparative statics, a higher-skill profit-driven stakeholders must be matched with the team with z-index (i.e., with higher willingness to pay for profit-driven stakeholder's skill), which thus explains result (i).

On the other hand, for a purpose-driven stakeholder $(x_\ell, 1)$, productivity is always dis-

counted regardless of purpose-driven index of the team $n_{-\ell}$, as

$$\frac{\partial}{\partial x_\ell}\Omega(y_{-\ell}x_\ell,\, n_{-\ell}+1) = (1-c\sigma)y_{-\ell}. \tag{9}$$

As a result, a team's willingness to pay for purpose-driven stakeholder's skill can be reduced to the ranking of their productivity $y_{-\ell}$. Thus, higher-skill purpose-driven stakeholders therefore must match with more productive teams.

Lemma above thus suggests that, for any $\sigma > 0$, the more productive teams will not necessarily match with more productive stakeholders. Intuitively, since teams without purpose-driven members value profit-driven stakeholders relatively more, as they can avoid mitigation, equilibrium matching generally distorts productivity. Hence, equilibrium sorting must jointly account for productivity and purpose.

3 Two Benchmarks

We isolate the two structural forces behind exit and engagement by shutting each down in turn. The first benchmark removes the harm–size correlation ($\sigma = 0$); the second removes multiplicity (only a single stakeholder group with purpose-driven preference). Each collapses the exit–engagement margin to a single response, and together they motivate the general model of Section 4, in which both can arise.

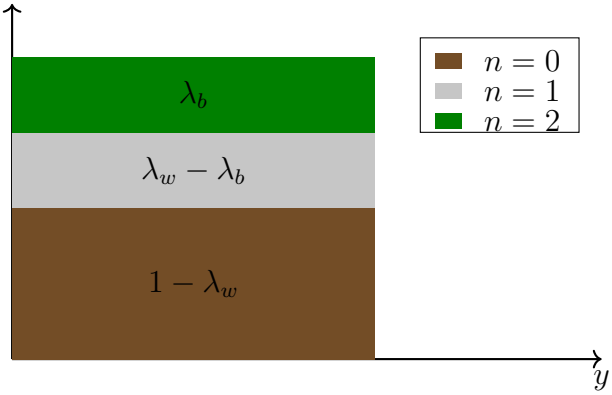
3.1 Independence Benchmark: Engagement Only

We begin with independence. When $\sigma = 0$, the model delivers two sharp results: purpose shocks operate exclusively through engagement, and there are no spillovers across stakeholder groups.

Sorting and Compensating Differentials under Independence When $\sigma = 0$, surplus becomes separable in y and n, so sorting on ability and sorting on purpose can be solved independently. From Equation (5), the reason is that the marginal value of productivity does not depend on the team's purpose-driven index, since $\Omega_y(y, n) = 1 \forall n \geq 0$ when $\sigma = 0$. There is therefore positive assortative sorting on x_ℓ for all ℓ. More productive stakeholders are matched with one another, and consequently, with more productive firms. Hence, the productivity of a team with ranking i is thus given by $y[i] = \prod_{\ell=1}^{N+1} x_\ell[i]$.

Given productivity, how then are the distribution of the index n and compensation determined? For expository purposes, let us first consider the case of $N = 2$. Figure 2 illustrates the simple case in which firms form relationships with two stakeholder types ($N = 2$, with

Figure 2: Distribution of purpose-driven index across y, where $\lambda_w \geq \lambda_b$



$\ell \in \{w, b\}$), interpreted as workers and banks, and where $\lambda_w \geq \lambda_b$. Because purpose-driven banks are relatively scarce, all such banks can be matched with purpose-driven workers (according to Lemma 1). Hence, the measure of firms with exactly two purpose-driven stakeholders, $n = 2$, is given by λ_b. By market clearing, some purpose-driven workers must then be matched with profit-driven banks, thereby forming firms with $n = 1$. The measure of firms with exactly one purpose-driven stakeholder is therefore $\lambda_w - \lambda_b$. The remaining firms are purely brown, with measure $1 - \lambda_w$.

Compensating differentials (due to equilibrium variation in mitigation across firms) are pinned down by the following indifference conditions. First, a firm must be indifferent between hiring a $n = 0$ team (profit-driven worker and bank) and hiring a $n = 1$ team (purpose-driven worker and profit-driven bank since $\lambda_w > \lambda_b$). Hence, the difference in compensation between profit-driven versus purpose-driven workers must reflect the difference in the firm's mitigation cost:

$$p_w(x_w, 0) - p_w(x_w, 1 \mid 1) = c\left(\xi_0^* - \xi_1^*\right),$$

where ξ^* represents the equilibrium level of emissions.

Second, the difference in compensation between profit-driven and purpose-driven banks is pinned down by the condition that a firm is indifferent between hiring a $(n = 1)$ team (purpose-driven worker and profit-driven bank) and hiring a $(n = 2)$ team (purpose-driven worker and bank). This yields

$$p_b(x_b, 0) - p_b(x_b, 1 \mid 2) = c\left(\xi_1^* - \xi_2^*\right) - \left[\psi(\xi_1^*) - \psi(\xi_2^*)\right].$$

The first term represents the reduction in mitigation cost. The second term captures the fact that the firm must also pay more to its purpose-driven worker when harm is higher, because that worker's disutility rises.

These results generalize immediately to any N, where types are ranked based on λ_ℓ.

Proposition 1. *When $\sigma = 0$:*

1. *(Sorting on productivity) Two stakeholders with the same productivity but different preferences work for teams with the same productivity, and more productive stakeholders work for more productive teams;*

2. *(Sorting on preferences) For any given y, the measure of firms with stakeholder-purpose index ℓ is $\lambda_\ell - \lambda_{\ell+1}$, while the measure of purely profit-driven firms is $1 - \lambda_1$*

3. *(Compensating differentials) The higher compensation for profit-driven compared to purpose-driven stakeholders is due to the savings in the cost of mitigation:*

$$\Delta_\ell(x_\ell) \equiv p_\ell(x, 0) - p_\ell(x, 1 \mid \ell) = c\left(\xi^*_{\ell-1} - \xi^*_\ell\right) - (\ell - 1)\left[\psi(\xi^*_{\ell-1}) - \psi(\xi^*_\ell)\right] \qquad \forall \ell \geq 1,$$

where $\xi^*_0 = \phi_0$.

Note that the compensating differential again has two terms. The first term captures the firm's benefit from a lower mitigation cost. At the same time, greater harm imposes additional disutility on existing purpose-driven stakeholders and therefore requires higher compensation to those stakeholders.

Engagement Only We now consider the effect of a change in λ_ℓ, an increase in the measure of purpose-driven stakeholders of type ℓ. Our goal is to determine whether the effect operates through engagement or through exit. Formally, we say that a stakeholder *engages* if his equilibrium team, summarized by $z^*(x_\ell, \theta)$, remains unchanged, and *exits* if it changes. [7]

Because sorting on productivity is independent of preferences or purpose, any change in λ_ℓ affects only the distribution of the purpose-driven index n, holding fixed productivity. For example, consider again the $N = 2$ case of banks and workers with $\lambda_w > \lambda_b$. Increasing the measure of purpose-driven banks, λ_b, by δ_b raises the measure of firms with two purpose-driven stakeholders by δ_b and lowers the measure of firms with one purpose-driven stakeholder by the same amount. In this sense, when banks become purpose-driven, they simply induce their firms to adopt more mitigation. These treated banks therefore effectively subsidize their firms by giving up their premium, since the difference in their compensation is exactly $p_b(x_b, 0) - p_b(x_b, 1|2)$. Intuitively, equilibrium prices ensure that firms are indifferent between profit-driven and purpose-driven stakeholders. When a stakeholder becomes purpose-driven, the firm is

[7] Conditional on $z^*(x_\ell, \theta)$, agents may be indifferent across purpose-driven indices n, so such variation is payoff-irrelevant. For this reason, we define engagement as the case in which $z^*(x_\ell, \theta)$ remains unchanged. Equivalently, one may interpret the preference shock as operating at the marginal firm, namely the firm with index $n = \ell - 1$, since the aggregate effect must occur there.

therefore willing to engage and saves the premium it had previously been paying for profit-driven stakeholders. Therefore, treated stakeholders will not exit.

Corollary 1 generalizes this intuition to any N.

Corollary 1. *(Impact Via Engagement Only) Suppose the share of purpose-driven type-ℓ stakeholders increases by a small amount $\delta_\ell > 0$, with $\lambda_\ell + \delta_\ell < \lambda_{\ell-1}$. When $\sigma = 0$, all treated stakeholders work for firms of the same size as before the shock, compensating differentials adjust, and for any y, a measure δ_ℓ of firms reduces harm by $\xi^*_{\ell-1} - \xi^*_\ell > 0$.*

The intuition is straightforward: when harm is independent of firm size, a purpose-driven stakeholder's disutility does not depend on which firm she works for—only on whether her firm mitigates. There is no incentive to leave a large firm for a smaller, cleaner one: the stakeholder can induce mitigation wherever she is. Engagement weakly dominates exit.

No Spillover Effect on Stakeholders' Utilities We now analyze how such shocks affect stakeholders' utilities. According to Equation (6), the return to skill x_ℓ is determined by the stakeholder's equilibrium team, $z^*(x_\ell, \theta)$. By the envelope theorem, we have

$$\frac{\partial U_\ell(x_\ell, \theta)}{\partial x_\ell} = z^*(x_\ell, \theta), \tag{10}$$

where $z^*(x_\ell, \theta)$ denotes the stakeholder's equilibrium matching team. Hence, the utility of stakeholder (x_ℓ, θ) is given by

$$U_\ell(x_\ell, \theta) = \int_{\underline{x}_\ell}^{x_\ell} z^*(\tilde{x}_\ell, \theta)\, d\tilde{x}_\ell + U_\ell(\underline{x}_\ell, \theta). \tag{11}$$

Note that, for any ℓ, the least valuable stakeholder is the least productive purpose-driven stakeholder. This stakeholder must therefore obtain his outside option, which we normalize to zero. Hence, $U_\ell(\underline{x}_\ell, 1) = 0$. In contrast, the least productive pecuniary stakeholder earns an additional premium due to compensating differentials.

Therefore, given the matching outcome $z^*(x_\ell, \theta)$, equilibrium utilities for all stakeholders are uniquely pinned down. It follows immediately that if such shocks do not affect the matching outcome, then stakeholders' utilities remain unchanged.

Corollary 2. *(No Reallocation and No Spillover) A preference or purpose shock of size δ_ℓ does not affect the equilibrium allocation or the utilities of any agents. That is, $z^*(x_{\ell'}, \theta)$ and $U_{\ell'}(x, \theta)$ remain unchanged for all ℓ', x, and θ. Conditional on skills, the utility gain of a profit-driven relative to a purpose-driven stakeholder remains the same, which is given by*

$$U_\ell(x_\ell, 0) - U_\ell(x_\ell, 1) = \Delta_\ell + \psi(\xi^*_\ell), \tag{12}$$

and decreases in ℓ.

In this sense, the payoffs of different stakeholder types are isolated: an increase in the prevalence of purpose-driven preferences among type-ℓ stakeholders does not affect the utilities of other agents. This no-spillover result has a direct empirical interpretation: a commitment by banks to become purpose-driven does not affect the welfare of workers at those same firms, or at any other firm. The commitment is absorbed locally through compensating differentials without propagating across stakeholder types or across firms.

Note that while utilities remain the same, the payment to existing purpose-driven stakeholders would decrease if the firm were to reduce the harm so that the purpose-driven stakeholder remains indifferent. Thus, the compensation of a purpose-driven stakeholder of type ℓ in a firm with index $n \geq \ell$ is

$$p_\ell(x, 1 \mid n) = U_\ell(x, 1) + \psi(\xi_n^*) \qquad \forall n \geq \ell, \tag{13}$$

where ξ_n^* denotes equilibrium harm under optimal mitigation, characterized by $n\psi'(\xi) = c$, which is strictly decreasing in n.

3.2 The Single-Group Benchmark: Exit Only

The reason for the engagement-only result above is that preference shocks do not affect a stakeholder's productivity ranking; hence, there is no reallocation. We now consider the other extreme case, where the effective productivity ranking of a treated stakeholder (i.e., a pecuniary stakeholder who becomes purpose-driven) must decrease. As a result, these treated stakeholders must reallocate (i.e., exit).

PAM using the z-index. This can be seen in the simple case when there is only one group of stakeholders that has purpose-driven preferences (i.e., $\lambda_1 > 0$ and $\lambda_{\ell'} = 0 \; \forall \ell' \neq 1$). That is, consider again the case of $N = 2$, where both banks and firms are purely pecuniary, while only some workers have purpose-driven preferences. The sorting in this special case can be solved in the following way. First, re-rank workers based on their own z-index, where $z_w = (1 - c\sigma)\, x_w$ if $\theta_w = 1$ and $z_w = x_w$ otherwise. Then, there is positive assortative matching between the worker's z-index z_w, the bank's size, and the firm's size. This is because the other groups of stakeholders (banks and firms) are purely profit-driven and all value workers based on their effective productivity; hence the larger banks and firms must hire the workers with a higher z-index.

Exit only, affecting all stakeholders' utilities. When a pecuniary worker x_w becomes purpose-driven, his z-index must decrease from x_w to $(1 - c\sigma)\, x_w$. As a result of PAM, a treated worker must now move to a smaller firm, which is also matched with a smaller bank.

Since such a shock changes the distribution of z_w, all stakeholders' matching partners $z^*(x_\ell, \theta)$ change through this channel, which in turn affects $U_\ell(x_\ell, \theta)$. Specifically, since an increase in λ_w results in a worse distribution of z_w, all banks and firms will hire workers with lower z_w, and thus their profits must decline, according to Equation (11). On the other hand, the untreated pecuniary workers become more scarce, and thus their z_w-ranking improves; they move to larger firms and banks, enjoying higher utilities.

Proposition 2 (Single-Group Benchmark: Exit Only)**.** *In the environment with $\sigma > 0$ and only one group of purpose-driven stakeholders ($\lambda_1 > 0$ and $\lambda_{\ell'} = 0\ \forall \ell' \neq 1$), every treated stakeholder exits their original firm and moves to a smaller firm when there is an increase in λ_1. An increase in in λ_1 also results in lower profits for all other groups of stakeholders.*

3.3 On the Coexistence of Exit and Engagement

Together with the independence benchmark of Section 3, this delineates the two structural forces behind the exit–engagement nexus. When $\sigma = 0$, preference shocks do not affect ranking, and thus the matching remains stable, and impact operates through engagement alone (Corollary 1). When $\sigma > 0$ and there is only a single group of purpose-driven stakeholders, their preference shock must change the ranking of their original team, which therefore results in exit.

These two benchmarks highlight that the coexistence of exit and engagement requires both that harm scale with size and that there be multiple groups of purpose-driven stakeholders. In such a general environment, whether the ranking of the team changes depends on who else is in the team, which is an endogenous object. Specifically, according to Equation (8), if a treated stakeholder's original team members already consist of other purpose-driven stakeholders, then the preference shock need not change the ranking of the team; hence, both exit and engagement are possible. We characterize that general case next, and establish that exit (engagement) arises at the top (bottom) of the distribution.

4 General Characterization

We now consider the general setting where environmental or social harm scales with production, i.e. $\sigma > 0$. Intuitively, larger firms now have stronger incentives to avoid mitigation and are willing to match with less productive profit-driven stakeholders to do so. Hence, unlike the independence benchmark of Section 3, the most productive firms will compete for relatively less productive profit-driven stakeholders. Sorting on the dimensions of productivity and purpose are no longer independent. For this reason, the multi-dimensional matching problem is much more complex.

We first analyze a simple environment, where firms form relationships with two types of stakeholders. We then extend the construction to general N via a sequential algorithm.

4.1 Illustrative example with $N = 2$

To fix ideas, consider again that firms form relationships with two types of stakeholders (say workers and banks), with purpose-driven probability λ_w and λ_b.

We first focus on the matching between workers and banks. Since firms are all profit-driven, the match between the firm and the team consisting of workers and banks can then be characterized by PAM on the firm's productivity x_{N+1} and the team's z-index. Thus, workers and banks understand that if their team has a higher z-index, they will be able to match with a larger firm.

Balanced supply: full segmentation. When the shares of purpose-driven workers and banks are identical, $\lambda_w = \lambda_b$, equilibrium matching between workers and banks is again very simple, as it must be fully segmented. Because of clustering (Lemma 1), purpose-driven workers match exclusively with purpose-driven banks, and profit-driven workers match exclusively with profit-driven banks. Within the purpose-driven or profit-driven segment, matching is positive assortative in skills by Lemma 2.

Lemma 3. *(Balanced Supply) When $\lambda_w = \lambda_b$, equilibrium matching is fully segmented, with positive assortative matching within each segment. Moreover, there will be a positive sorting between firms and the z-index of the team.*

We can visualize the equilibrium matching outcomes in Figure 3. Consider two banks of the same talent x_b — one profit-driven $(x_b, 0)$ and one purpose-driven $(x_b, 1)$. The profit-driven bank's matching worker is given by the dashed brown line, while the purpose-driven bank's is given by the dashed green line. They both match with a worker of the same quality x_w but the profit-driven bank matches with the worker with a higher z-index, which in this case is just x_w while the purpose-driven bank matches with a worker with a z-index of $(1 - c\sigma)x_w$ (following Equation 8). The positive slope of the two lines reflects positive assortative matching of bank and worker quality.

Unbalanced supply: segmentation at the top, mixing at the bottom. When $\lambda_w \neq \lambda_b$, full segmentation is no longer feasible due to the need for market clearing. Suppose $\lambda_w > \lambda_b$, so that workers have relatively more purpose-driven stakeholders. Market clearing then requires some purpose-driven workers to match with profit-driven banks. We show that such mixing occurs only at the bottom but not at the top of the skill distribution.

Lemma 4. *(Full Separation At The Top) When $\lambda_w \geq \lambda_b$, there exists a cutoff $\hat{x}_b$ such that, for all $x_b \geq \hat{x}_b$, profit-driven (purpose-driven) banks match only with profit-driven (purpose-driven) workers.*

Relative to the balanced benchmark, profit-driven workers are scarce, so maintaining pure profit-driven teams requires sacrificing productivity. At high skill levels, profit-driven banks continue to prefer pure profit-driven matches, while at lower skill levels they optimally accept purpose-driven workers, generating mixing.

Productivity loss versus mitigation cost. A profit-driven bank faces a fundamental trade-off: it can match with a lower-skill profit-driven worker to avoid mitigation costs, or match with a higher-skill purpose-driven worker and incur those costs. Figure 3 illustrates how this trade-off resolves under unbalanced supply ($\lambda_w > \lambda_b$).

The dashed lines represent the balanced supply benchmark ($\lambda_w = \lambda_b$). The solid lines represent the imbalanced case. To understand the shift, consider two banks of identical skill x_b, one profit-driven $(x_b, 0)$ and one purpose-driven $(x_b, 1)$. Because the cost of mitigation is proportional to total output, high-talent profit-driven banks have the strongest incentive to avoid purpose-driven partners. Consequently, they aggressively out-bid other stakeholders for the now-scarce pool of high-talent profit-driven workers.

However, because these profit-driven workers are scarce relative to the bank population, these high-talent banks must "settle" for workers of lower skill than they would in a balanced market. This is why the solid brown line lies below the dashed brown line—it represents a downward displacement in the quality of profit-driven matches. Conversely, since there is a surfeit of talented purpose-driven workers, the solid green line sits above the dashed green line, as purpose-driven banks find it easier to secure top-tier talent.

For "middling" profit-driven banks (those below the cutoff $\hat{x}_b$), the scarcity of profit-driven workers becomes prohibitive. Having been out-competed for the remaining pool of reasonable profit-driven talent, these banks find that the productivity loss of matching with an untalented profit-driven worker exceeds the cost of mitigation. At this threshold, they are willing to match with high-talent purpose-driven workers. This transition is represented by the solid gray line in the mixing region. Here, the productivity of the purpose-driven match is high enough that the bank is willing to internalize the mitigation tax rather than match with the low-skill profit-driven workers left at the bottom of the distribution. The profit-driven bank is indifferent between profit-driven or purpose-driven workers now as long as these workers have the same effective productivity z-index.

PAM between banks (x_b) and workers' effective productivity (z_w) in the mixing region. In the mixing region, we show that matching can be constructed based on the positive

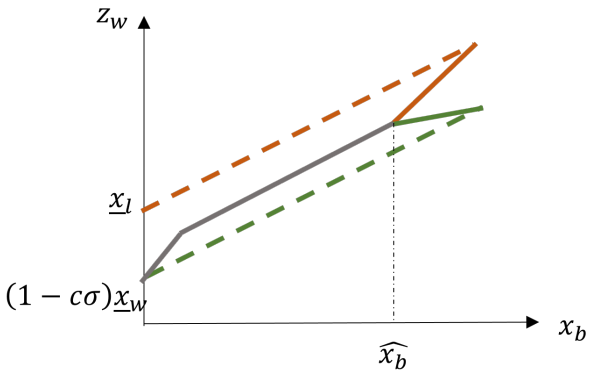


Figure 3: Dashed lines represent balanced supply, where the z-index of purpose-driven matches is lower due to the discount factor $(1 - c\sigma)$. The solid lines represent unbalanced supply, where full separation occurs the top. The dashed purpose-driven (profit-driven) line represents the z-index of the matching workers under the balanced supply for purpose-driven (profit-driven) bank.

sorting between the effective productivity of workers z_w and banks ability x_b

Because profit-driven banks rank potential workers according to their effective productivity, we define a cumulative measure of "effective supply" among workers. Let $g_w^1(y)$ and $g_w^0(y)$ denote the probability density functions of the productivity distribution for teams with purpose-driven members ($n \geq 1$) and purely profit-driven members ($n = 0$), respectively. We define $\Psi_w(z)$ as the total mass of worker teams whose effective productivity is less than or equal to z:

$$\Psi_w(z) \equiv \int_{\underline{y}}^{z/(1-c\sigma)} g_w^1(\tilde{y})\, d\tilde{y} + \int_{\underline{y}}^{z} g_w^0(\tilde{y})\, d\tilde{y}. \tag{14}$$

In this expression, the first integral accounts for purpose-driven teams. Because their productivity is discounted by $(1 - c\sigma)$, a purpose-driven team must have a higher raw productivity ($\tilde{y} = z/(1 - c\sigma)$) to provide the same effective z as a profit-driven team.

Market clearing then requires that the rank of a bank in their skill distribution $F_b(x_b)$ matches the rank of their partner in the effective productivity distribution $\Psi_w(z)$. Let $\phi_m(x_b)$ denote the matching function that assigns a stakeholder of skill x_b to a team with a specific z-index in the mixing region. The equilibrium matching is determined by:

$$\Psi_w(\phi_m(x_b)) = F_b(x_b), \tag{15}$$

as illustrated by the solid gray line in Figure 3. This construction implies that two banks with the same rank but different preferences $(x_b, 0)$ and $(x_b, 1)$ will be matched with workers with the same z-index of $\phi_m(x_b)$.

Clustering within the mixing region. While banks with a given x_b are matched with workers of the same z-index, according to Lemma 1, stakeholders with the same preference

are matched with each other as much as they can. Specifically, conditional on x_b and z, all profit-driven workers with index z and $n = 0$, which are relatively scarce, will be matched with profit-driven banks, forming a team index with $n = 0$. On the other hand, since profit-driven banks are relatively abundant, some of them will then be matched with purpose-driven workers with the same index z, forming teams with $n = 1$. Purpose-driven banks, which again are relatively scarce, will all match with purpose-driven workers, yielding teams with $n = 2$.

Summary. High-skill stakeholders remain segmented by their purpose, as the cost of mixing is prohibitive at the top of the productivity distribution. Mixing occurs at the bottom, where productivity losses from matching with a less-skilled partner are small. This is the origin of the trickle-down effect that we will analyze in Section 5: exit occurs at the top, engagement at the bottom. Competition endogenously determines which stakeholders bear the cost of mitigation.

4.2 General case: sequential algorithm for $N \geq 2$

We now construct the equilibrium with N types of stakeholders via a recursive sequence of bilateral matches. Types are ordered by the prevalence of purpose-driven stakeholders. Label stakeholder types such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N,$$

where λ_ℓ denotes the share of purpose-driven stakeholders of type ℓ. Type 1 is thus the most purpose-driven, while later types are increasingly profit-driven. For instance, workers might be λ_1, suppliers might be λ_2, while banks might be λ_N, and so on.

While the multi-sided matching problem with N types is high-dimensional, the following sequential algorithm simplifies it into a series of nested bilateral decisions.

1. **Initial Match:** Match type 1 and type 2 stakeholders using the bilateral results from Section 3.2. This generates a distribution of teams with type 1 and 2 stakeholders with productivity $y_2 = x_1 x_2$ and a purpose-count $n_2 \in \{0, 1, 2\}$.

2. **Iteration:** For each subsequent stage $\tau \geq 3$, the team formed at stage $\tau - 1$, is treated as a single partner with characteristics $(y_{\tau-1}, n_{\tau-1})$ and is matched with stakeholders of type τ.

   The matching construction follows the same patterns shown in Figure 2: full separation holds at the top of the distribution, while mixing arises at the bottom. The x-axis now represents the ability of type τ, x_τ, and the y-axis is the effective z-index of the team, given by $z_{\tau-1}(y_{\tau-1}, n_{\tau-1})$.

   Let $x^*_\tau(y_{\tau-1}, n_{\tau-1})$ and $\theta^*_\tau(y_{\tau-1}, n_{\tau-1})$ denote the characteristics of the stakeholder of type τ matched to a team $(y_{\tau-1}, n_{\tau-1})$. After matching with type τ, team characteristics then

evolve according to

$$y_\tau = y_{\tau-1} \cdot x^*_\tau(y_{\tau-1}, n_{\tau-1}), \qquad n_\tau = n_{\tau-1} + \theta^*_\tau(y_{\tau-1}, n_{\tau-1}). \tag{16}$$

After the matching at stage τ, the team now comprises types $1, \ldots, \tau$, with characteristics (y_τ, n_τ). Let $g^1_\tau(y_\tau)$ and $g^0_\tau(y_\tau)$ denote the densities of teams with $n_\tau \geq 1$ and $n_\tau = 0$, respectively.

3. **Final Firm Allocation:** The completed N-stakeholder teams, characterized by their final effective productivity z_N, are matched with firms (type $N+1$) based on the firm's productivity x_{N+1}.

Monotonicity of scarcity condition. Note that one key assumption for constructing PAM based on x_τ (e.g. banks) and the effective productivity index of the team, $z_{\tau-1}$, is that type τ has relatively abundant profit-driven stakeholders and they are the ones mixing under our construction. The measure of the profit-driven stakeholder conditional on x_τ is constant under the i.i.d assumption, which is given by $(1-\lambda_\tau)$. On the other hand, the measure of the pure profit-driven stakeholder conditional on z_τ is given by $\Pr(n_\tau = 0|z_\tau) = \frac{g^0_\tau(z)}{g^0_\tau(z)+g^1_\tau(z/(1-c\sigma))}$, which is generally not constant. Hence, the following condition below is needed to ensure that the measure of the profit-driven stakeholder of type τ is higher than the measure of pure profit-driven team everywhere under the mixing region.

Assumption 3. $(1-\lambda_\tau) \geq \Pr(n_{\tau-1} = 0|z_{\tau-1})$ *for* $\forall z_{\tau-1} \leq (1-c\sigma)\bar{z}$.

Proposition 3. *(Sequential Construction of Equilibrium) Suppose Assumption 3 holds at every stage τ. Then the equilibrium matching outcome can be constructed sequentially as follows. At each stage τ, the matching between teams $(y_{\tau-1}, n_{\tau-1})$ and stakeholders (x_τ, θ_τ) is characterized by a cutoff $\hat{x}_\tau$ such that:*

1. Full separation at the top: *profit-driven (purpose-driven) stakeholders above the cutoff match exclusively with teams with $n_\tau = 0$ ($n_\tau \geq 1$).*

2. Mixing at the bottom: *below the cutoff, matching is characterized by positive assortative matching between stakeholder skill x_τ and the team z-index $z_{\tau-1} = z(y_{\tau-1}, n_{\tau-1})$. Conditional on $(z_{\tau-1}, x_\tau)$, purpose-driven stakeholders are matched first with teams with higher purpose-driven indices.*

Team characteristics evolve according to (16).

Evolution of purpose-driven teams. The sequential structure has two implications. First, a purely profit-driven team ($n_\tau = 0$) at period τ remains profit-driven in all subsequent stages, since later stakeholder types are weakly more profit-driven. As a result, the total mass of profit-driven teams is $(1 - \lambda_1)$, the mass of profit-driven type-1 stakeholders. Second, in later stages, purpose-driven stakeholders are increasingly scarce. Consequently, only teams with higher purpose-driven indices continue to match with purpose-driven stakeholders, so that purpose-driven intensity becomes progressively concentrated among a shrinking subset of teams.

4.3 Rents

Under the independence benchmark of Section 3, a treated stakeholder can simply subsidize their original firm so that the firm is willing to pay for additional mitigation costs. When $\sigma > 0$, differences in compensation between profit-driven and purpose-driven stakeholders are no longer due solely to compensating differentials. An additional component emerges: scarce profit-driven stakeholders at the top of productivity distribution earn rents because productive firms compete aggressively to avoid mitigation. This avoidance rent drives a wedge beyond compensating differentials and explains why treated stakeholders who are highly productive cannot afford to engage—no transfer is large enough to compensate the firm.

Specifically, for type 1 stakeholders (the scarcest profit-driven type), our equilibrium construction implies that they remain in purely profit-driven teams and facilitate the avoidance of mitigation costs. They therefore match with more productive firms and earn higher wages relative to their purpose-driven counterparts. For type $\ell \geq 2$ stakeholders, the picture is more nuanced. In the separation region ($x_\ell > \hat{x}_\ell$), profit-driven stakeholders match with teams of higher z-index than their purpose-driven counterparts and earn rents beyond compensating differentials. In the mixing region ($x_\ell \leq \hat{x}_\ell$), profit-driven and purpose-driven stakeholders of the same skill match with teams of the same z-index, so the premium reflects only compensating differentials.

Proposition 4 formalizes this intuition.

Proposition 4. *(Rent for Scarce Profit-Driven Stakeholders) Profit-driven stakeholders of type* 1 *and for* $\forall \ell \geq 2$ *earn rents (beyond compensating differentials), where* $z^*(x_\ell, 0) > z^*(x_\ell, 1)$*, and* $p_\ell(x, 0) - p_\ell(x, 1|\ell) > \Delta_\ell(x_\ell)$ *for* $x_\ell > \hat{x}_\ell$*. In contrast, profit-driven stakeholders in the mixing region earn no additional rent. That is, for* $\forall \ell \geq 2$*,* $p_\ell(x, 0) - p_\ell(x, 1|\ell) = \Delta_\ell(x_\ell)$ $\forall x_\ell \leq \hat{x}_\ell$*.*

5 Exit vs. Engagement in General Equilibrium

We now use the equilibrium characterization to study how an exogenous increase in the share of purpose-driven stakeholders affects firm–stakeholder relationships and firm harm. Consider

a shock that increases the measure of purpose-driven stakeholders of type ℓ (i.e., λ_ℓ). Following the shock, do affected stakeholders optimally remain with their current firms (*engagement*) or reallocate to new firms (*exit*)? And what happens to aggregate harm and mitigation?

In contrast to the independence benchmark of Section 3, we show that aggregate outcomes now manifest very differently. First, both exit and engagement can arise. Second, even though the preference shocks are i.i.d. across the productivity distribution, little direct impact or mitigation materializes at the top of that distribution. Instead, aggregate impact is concentrated in the middle region of firm productivity, through equilibrium spillovers—a trickle-down effect. Lastly, an increase in the measure of type-ℓ stakeholders now affects the utilities of other stakeholders.

5.1 Prevalence of Exit versus Engagement

We now show that both exit and engagement can arise when the shock increases the measure of stakeholders of type $\ell \geq 2$.

Illustration: the $N = 2$ case. To illustrate the intuition, consider our $N = 2$ case with banks (λ_b) and workers (λ_w) from Section 4.1. In Figure 4, we show what happens to the equilibrium matching outcomes when λ_b increases by δ—that is, the $t = 0$ equilibrium versus the $t = 1$ equilibrium. The solid lines represent the $t = 0$ equilibrium, as illustrated in Figure 3. The dotted lines represent the $t = 1$ equilibrium. More purpose-driven banks reduce competition for profit-driven workers. Hence, the remaining profit-driven banks can now match with more productive profit-driven workers, which is reflected in the dotted brown line lying above the solid brown line. Similarly, purpose-driven banks are now matched with less productive purpose-driven workers, which is reflected in the dotted green line lying below the solid green line.

As a result, the separation threshold falls (i.e., $\hat{x}_{b,1} < \hat{x}_{b,0}$), so the separation region expands, which can be interpreted as profit-driven banks becoming less willing to mix. This means that treated banks below the new cutoff $\hat{x}_{b,1}$ remain with their previous matches—that is, they engage. For treated banks above this new cutoff, there is in general reallocation, i.e. exit.

We now show that this illustration holds for general N.

Proposition 5. *(Exit vs. Engagement) Suppose $\ell \geq 2$ and the share of purpose-driven type-ℓ stakeholders increases by a small amount $\delta_\ell > 0$, with $\delta_\ell < \lambda_{\ell-1} - \lambda_\ell$. Then there exists a cutoff firm size such that: (i) treated stakeholders at firms above the cutoff exit to smaller firms with higher stakeholder-purpose indices; (ii) treated stakeholders below the cutoff engage the firm of the same size.*

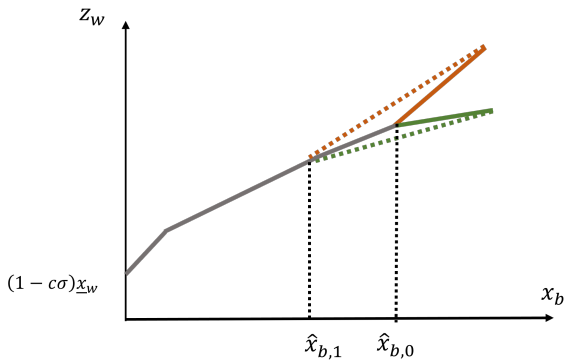


Figure 4: The solid (dotted) line represents the matching before (after) the preference shock that increases the measure of purpose-driven banks.

Exit at the top. At the top of the distribution, full separation implies that the stakeholder-purpose index is binary, $n \in \{0, N\}$. Treated stakeholders previously matched with large firms must now reallocate to firms with higher purpose-driven indices. Moreover, once a stakeholder becomes purpose-driven, she no longer enjoys an avoidance rent. As a result, treated high-ability stakeholders optimally exit to smaller firms after the shock.

Engagement at the bottom. Consider a treated stakeholder with ability $x_\ell < \hat{x}_{\ell,1}$. By construction, both before and after the shock, stakeholders $(x_\ell, 0)$ and $(x_\ell, 1)$ are matched with teams of the same z-index.[8] Therefore, when a low-ability stakeholder becomes purpose-driven, the ranking of her equilibrium match does not change: $z^*(x_\ell, 1) = z^*(x_\ell, 0)$. She can thus remain with a firm of the same size. Specifically, let $\hat{x}_{N+1,t}$ be the firm matched to the cutoff stakeholder $\hat{x}_{\ell,t}$ at period t, which is given by $F_{N+1}(\hat{x}_{N+1,t}) = F_\ell(\hat{x}_{\ell,t})$. Then treated stakeholders at firms below the cutoff $\hat{x}_{N+1,1}$ work for a firm of the same size.

Remark: exit only for type-1 stakeholders. The logic here also explains why engagement cannot happen when the shock hits type-1 stakeholders, the type with the highest measure of purpose-driven stakeholders. The reason is that profit-driven stakeholders are then relatively scarce; hence, in equilibrium, every profit-driven stakeholder of type 1, $(x_\ell, 0)$, matches with a team of higher z-index than the purpose-driven stakeholder of the same ability, $(x_\ell, 1)$. All treated stakeholders in this case must therefore reallocate, so no engagement occurs.

5.2 Impact on Emissions

We now study how preference shocks affect firm harm. We focus on stakeholders of type $\ell \geq 2$, since both exit and engagement can arise for them. Recall that firm harm depends on

[8] In the mixing region, the sorting function $\phi_m(x_\ell)$ that solves (15) is independent of λ_ℓ.

productivity y and the stakeholder-purpose index n:

$$e^*(y,n) = \begin{cases} \phi(y), & n = 0 \quad \text{(purely profit-driven team)}, \\ \xi_n^*, & n \geq 1 \quad \text{(purpose-driven team)}. \end{cases}$$

When $n \geq 1$, harm depends only on n and not on productivity. Hence, the first-order impact of preference shocks operates through changes in the distribution of the stakeholder-purpose index across firms.

Aggregate impact. As in the independence case, because purpose-driven stakeholders cluster, the distribution of firms by stakeholder-purpose index is determined by $(\lambda_1, \ldots, \lambda_N)$. The key difference, however, is that the distribution of y is no longer independent of the distribution of n. Nevertheless, at the aggregate level, an increase in λ_2 by δ raises the measure of fully purpose-driven firms by δ and reduces the measure of partially purpose-driven firms by δ, leaving purely profit-driven firms unchanged.

This logic generalizes to any N: conditional on the ordering of λ_ℓ remaining unchanged (i.e., $\lambda_{\ell-1} \geq \lambda_\ell + \delta_\ell$), increasing λ_ℓ by δ_ℓ raises the share of firms with index ℓ by δ_ℓ and lowers the share with index $\ell - 1$ by δ_ℓ. In other words, the change in aggregate emissions is $\delta_\ell \Delta\xi_\ell$, where $\Delta\xi_\ell \equiv \xi_{\ell-1}^* - \xi_\ell^* > 0$. That is, at the aggregate level, a δ_ℓ measure of firms lower their emissions from $\xi_{\ell-1}^*$ to ξ_ℓ^*.

Note that, while the measure of purely profit-driven firms stays the same—since we focus on shocks to type $\ell \geq 2$—which firms are profit-driven can affect emissions, as harm depends on productivity for firms with $n = 0$. However, this channel is quantitatively minor, since we have a continuum of firms and consider a small δ_ℓ. We therefore focus on changes in the distribution of n alone throughout the rest of the paper when analyzing the impact on emissions.

Trickle-down effect among middle-sized firms. We now show that, even though the preference shock is spread uniformly across the ability distribution, aggregate impact is realized disproportionately in the middle of the firm-size distribution. This result arises precisely because of the coexistence of exit and engagement: exiting stakeholders at the top move to firms that already mitigate, so some deliver no direct reduction, while engagement at the bottom delivers the per-treated amount directly, at the treated firms themselves.

The shortfall, however, is made up in the middle. This is because treated high-ability stakeholders crowd out lower-ability untreated ones, who then join smaller firms and make those firms more purpose-driven. Importantly, this spillover effect is largest in the middle and does not extend all the way down to the bottom of the distribution. The reason is that, according to Proposition 5, there is no reallocation below the cutoff; hence, any spillover effect

can arise only above the cutoff. As a result, this displacement generates additional harm reduction in the middle of the distribution.

Proposition 6. *(Trickle-Down Effect) For any small shock $\delta_\ell > 0$ that preserves the ordering of types, the aggregate impact on emissions is $\delta_\ell \Delta\xi_\ell$ $\forall \ell \geq 2$. (i) At the bottom ($x_\ell \leq \hat{x}_{\ell,1}$), the average harm reduction equals $\delta_\ell \Delta\xi_\ell$. (ii) In the middle region ($x_\ell \in [\hat{x}_{\ell,0}, \hat{x}_{\ell,1}]$), the average harm reduction exceeds $\delta_\ell \Delta\xi_\ell$. (iii) At the top ($x_\ell \geq \hat{x}_{\ell,1}$), the average harm reduction is lower than $\delta_\ell \Delta\xi_\ell$.*

Firm-level effect. We now analyze the micro-level effect. The impact on an individual firm is given by the change in its purpose-driven index before and after the shock, $n_{i,1} - n_{i,0}$. Given any firm size x_i, the model uniquely pins down the z-index of its matching team in each period. However, firms are generally indifferent across teams with different indices n but the same z-index. To rule out payoff-irrelevant reshuffling, we impose the following minimal-adjustment rule at the individual-firm level.

Assumption 4. *(Minimal-adjustment rule). For any two firms with the same size, $(x_i, n_{i,t})$ and $(x_j, n_{j,t})$, the ranking of their purpose-driven indices remains the same before and after the shock: if $n_{i,0} \geq n_{j,0}$, then $n_{i,1} \geq n_{j,1}$.*

In other words, conditional on firms of the same size, it cannot be the case that a brown and a greener firm switch their ranking. This assumption can be justified when there is an $\epsilon > 0$ adjustment cost of abatement technology, which generates persistence of abatement at the firm level. This natural assumption allows us to establish predictions about the impact at the individual-firm level, which is useful for quantifying the impact empirically.

Proposition 7. *(Firm-Level Impact) Consider a δ_ℓ shock to stakeholders of type ℓ. A treated stakeholder engages her original firm (x, n_0) if and only if the firm is relatively small ($x < \hat{x}_{N+1}$) and has an index of $n_0 = \ell - 1$. The direct effect of engagement on each treated firm is $\Delta\xi_\ell \equiv \xi^*_{\ell-1} - \xi^*_\ell$.*

5.3 Spillover Effects on Payoffs

We finally examine how preference shocks affect equilibrium payoffs. Recall that in Section 3, where $\sigma = 0$, there are no spillovers across types of stakeholders. When $\sigma > 0$, this is no longer true. Although only type-ℓ stakeholders are directly treated, the induced reallocation alters competition and matching economy-wide.

An increase in λ_ℓ intensifies competition among purpose-driven type-ℓ stakeholders, lowering the z-index of their equilibrium matches at the top of the distribution (the green line in Figure 3

shifts downward). Conversely, profit-driven type-ℓ stakeholders become scarce and match with higher-ranked teams. Thus, purpose-driven (profit-driven) type-ℓ stakeholders at the top are worse off (better off).

For stakeholders of other types $\ell' \neq \ell$, the effect is reversed. Due to complementarities in production, improvements in the skill distribution of purpose-driven teams raise productivity and benefit purpose-driven stakeholders of other types. In contrast, profit-driven stakeholders of other types face less productive matches.

Proposition 8. *(Spillover Effects on Payoffs) An increase in the share of purpose-driven type-ℓ stakeholders results in a lower (higher) payoff for purpose-driven (profit-driven) stakeholders of type ℓ that are relatively productive. On the other hand, it results in a higher (lower) payoff for purpose-driven (profit-driven) stakeholders of any other type $\ell' \neq \ell$ that are relatively productive. Payoffs are unchanged for all low-ability stakeholders.*

Engagement does not generate payoff spillovers, since matching at the bottom remains unchanged. However, because engagement reduces firm harm, the monetary payments to other purpose-driven stakeholders must decline.

6 Implications of the Coexistence of Exit and Engagement for Quantifying Impact

The coexistence of exit and engagement shapes not only equilibrium outcomes but also what an empirical researcher can recover from them. Our model implies that firm mitigation per treated stakeholder is identical whether she engages or exits; the two channels, however, land on different firms, and only engagement registers at the firm the researcher treats. When a treated stakeholder exits, her previously connected firm may fail to comply with its assignment. Symmetrically, firms that were never connected to a treated stakeholder—the control group—can nevertheless change their emissions when they form a new relationship. The consequence for measurement is stark and quantifiable: in our calibration to post-Paris bank decarbonization below, a firm-level difference-in-differences recovers only about a fifth of the true per-stakeholder effect, while a compliance-adjusted LATE recovers it in full.

Because of this general-equilibrium reallocation, the setting is best understood as a two-sided noncompliance problem. Firms that are matched with profit-driven stakeholders even though they were assigned to the treatment group are "never-takers"; conversely, firms that are matched with purpose-driven stakeholders even though they were assigned to the control group are "always-takers."

Noncompliance framework. We now formalize the mapping from the model-implied reallocation triggered by the preference shock to an empirical setting with noncompliance. Each firm i is characterized by its size x_i and purpose-driven index $n_{i,t}$, where $t = 0$ ($t = 1$) denotes the period before (after) the preference shock δ_ℓ that randomly turns a δ_ℓ measure of profit-driven type-ℓ stakeholders purpose-driven (i.e., these are the treated stakeholders). The treated (untreated) firms are those connected to a treated (untreated) stakeholder at $t = 0$, denoted by $Z_i = 1$ ($Z_i = 0$). That is, $Z_i \in \{0, 1\}$ is the binary treatment indicator.

Let $D_i \in \{0, 1\}$ indicate whether firm i is connected to a purpose-driven stakeholder at $t = 1$. In our framework, whether a firm ends up with a purpose-driven stakeholder decomposes into an own-engagement channel and a reallocation channel:

$$D_i \;=\; e_i\, Z_i \;+\; (1 - e_i)\, \eta_i,$$

where $e_i \in \{0, 1\}$ indicates whether the firm and its stakeholder preserve their existing relationship. Hence, the first term is the direct effect of engagement, where $e_i Z_i = 1$ if and only if the treated stakeholder engages (i.e., $e_i = 1$ and $Z_i = 1$). On the other hand, if the firm reallocates to a new stakeholder ($e_i = 0$), then $\eta_i \in \{0, 1\}$ indicates whether a firm whose relationship breaks is re-matched with a purpose-driven stakeholder. The probability of matching with a purpose-driven stakeholder for each firm i depends on the equilibrium matching outcome, which we elaborate on below.

6.1 The Bias in Firm-Level Estimates

First, since a standard firm-level difference-in-differences recovers only the engagement component, one can immediately see that it understates the per-stakeholder mitigation effect. Formally, let $D_T(x) \equiv \mathbb{E}[D \mid Z = 1, x]$ and $D_C(x) \equiv \mathbb{E}[D \mid Z = 0, x]$ denote the size-conditional take-up probabilities for the treatment and control groups, i.e. the probability that a firm is matched to a purpose-driven stakeholder of type ℓ at $t = 1$. The intention-to-treat contrast in any size cell is defined as $\mathrm{ITT}(x) \equiv \mathbb{E}[Y \mid Z = 1, x] - \mathbb{E}[Y \mid Z = 0, x]$, where Y is the change in the firm's emissions.

Lemma 5 (Exit and the Bias in Firm-Level Estimates)**.** *The firm-level difference-in-differences understates the per-stakeholder mitigation effect:*

$$\widehat{DiD} = \mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0] < \Delta\xi_\ell. \tag{17}$$

Moreover, $ITT_\ell(x) = 0$ for relatively large firms ($x > \hat{x}_\ell$). And, under Assumption 4, for $x < \hat{x}_\ell$, $ITT_\ell(x) = \{D_T(x) - D_C(x)\}\, \Delta\xi_\ell \leq \xi_\ell$.

6.2 Recovering the Per-Stakeholder Effect via LATE

Lemma 5 is not only a warning but also a prescription. Because the setting can be understood as a randomized assignment with two-sided noncompliance, a compliance-adjusted estimator recovers the per-stakeholder effect exactly—but only on the support where the instrument actually moves stakeholder composition.

Where the instrument is silent: no compliers above the cutoff. For $x > \hat{x}$ the engagement rate is zero, since treated stakeholders always exit; both the treatment and control groups thus reach purpose-driven stakeholders only through reallocation. Because both groups face the same equilibrium distribution when reallocating, Z does not move D; hence $D_C(x) = D_T(x)$. In other words, there are no compliers and $\text{ITT}(x) = 0$. The Wald ratio $\text{ITT}(x)/0$ is therefore undefined, and no conditional LATE can be formed in this region. This is not a technical nuisance but the empirical signature of exit: among large firms the instrument is silent, so they lie outside any identified set.

Where the instrument bites: conditional LATE below the cutoff. Fix a size cell $x < \hat{x}$. The key observation is that $\text{ITT}(x) = \{D_T(x) - D_C(x)\}\,\Delta\xi_\ell$ is an *intention-to-treat* object: it is the per-stakeholder effect scaled by the take-up gap. Dividing the reduced form by the first stage undoes precisely this attenuation, provided the first stage is nonzero.

Specifically, according to Proposition 7, a treated firm has $D_i = 1$ if and only if it already hires $\ell - 1$ purpose-driven stakeholders; otherwise, it drops the treated stakeholder and replaces her with an identical purpose-driven one. Hence, the take-up rate equals the engagement rate for the treatment group, which is the probability that a firm has an index of $n_0 = \ell - 1$ conditional on size x, denoted $\pi_{\ell-1}(x)$.

By market clearing, a treated bank dropped by a firm in the treatment group must be hired by another firm that also has $\ell - 1$ stakeholders in the control group. Let $\rho_\ell \equiv \frac{\delta_\ell}{1-\lambda_\ell}$ denote the share of treated firms. The take-up rate in the control group, $D_C(x)$, must then satisfy

$$\rho_\ell D_T(x) + (1 - \rho_\ell) D_C(x) = \rho_\ell, \tag{18}$$

given that the take-up rates in both groups must sum to ρ_ℓ in this region (Proposition 6).

Hence, using $D_T(x) = \pi_{\ell-1}(x)$ and Equation (18), the first stage is strictly positive:

$$D_T(x) - D_C(x) = \frac{\pi_{\ell-1}(x) - \rho_\ell}{1 - \rho_\ell} > 0, \tag{19}$$

so compliers exist and the Wald estimator is well-defined:

$$\text{LATE}(x) \;=\; \frac{\mathbb{E}[Y \mid Z=1,x]-\mathbb{E}[Y \mid Z=0,x]}{\mathbb{E}[D \mid Z=1,x]-\mathbb{E}[D \mid Z=0,x]} \;=\; \frac{\text{ITT}(x)}{D_T(x)-D_C(x)} = \Delta\xi \qquad x<\hat{x}. \tag{20}$$

Thus, on the complier-relevant support the compliance-adjusted estimator recovers the per-stakeholder mitigation effect exactly.

The primary estimate: pooled LATE. Because the conditional estimate is undefined for $x > \hat{x}$, we take the *pooled* Wald estimator as the headline object. Its appeal is that it handles the restricted support automatically: the zero-first-stage cells contribute zero to both the numerator and the denominator, and hence drop out of the weighting without any hand-selection of the support. Aggregating, the pooled LATE identifies the per-stakeholder effect on exactly the firms whose stakeholder composition the instrument moves:

$$\text{LATE} = \frac{\mathbb{E}[Y \mid Z=1]-\mathbb{E}[Y \mid Z=0]}{\mathbb{E}[D \mid Z=1]-\mathbb{E}[D \mid Z=0]} = \Delta\xi. \tag{21}$$

It is instructive to see why the pooled first stage isolates the engagement margin. Both groups are exposed to the same reallocation draw, so the spillover term appears in the take-up rate of the treatment and control groups alike. The treatment group differs only in that its engagers also comply, adding the engagement term on top. Because the spillover term is common to both groups, it cancels in the difference, and the compliance rate collapses to the pure engagement rate. In other words, the treatment group's take-up is "engagement plus spillover" while the control group's take-up is "spillover only," so the compliance adjustment in the denominator of the pooled Wald estimator is effectively the engagement rate alone.

Proposition 9 (Compliance-Adjusted Identification)**.** *The pooled estimator* (21) *recovers the per-stakeholder effect* $\Delta\xi$*. The Wald estimator conditional on size recovers the per-stakeholder effect* $\Delta\xi$ *only for relatively small firms.*

Why LATE works despite the spillover. This is the methodological core of the result, and it is where our analysis departs from the standard treatment of interference in program evaluation. That literature treats spillovers as a nuisance that contaminates the estimand and must be bounded or designed away; here, by contrast, the spillover is derived from the matching equilibrium and turns out to be of a benign type that leaves the LATE estimand exactly intact. The key is that reallocation acts on the *treatment* rather than on the *outcome*: when a purpose-driven stakeholder exits a treated firm and re-matches with a control firm, that control firm does not merely have its emissions perturbed from the outside—it becomes a

genuinely treated unit, connected to the same stakeholder and therefore delivering the identical per-firm reduction $\Delta\xi_\ell$.

In short, the spillover surfaces entirely in the first stage, showing up as a nonzero take-up rate in the control group. Because the reduced form is then merely a scaling of this first stage by the common effect, $\mathrm{ITT}(x) = [D_T(x) - D_C(x)]\,\Delta\xi$ over the relevant region, the reallocation term enters numerator and denominator through the same channel and cancels in the ratio, regardless of the magnitude of the spillover. The spillover is thus *treatment-propagation* rather than *outcome-externality*,[9] and it is absorbed by the IV ratio rather than contaminating it.

6.3 Illustration with Calibrated Data

We now apply the estimator to data generated by the model itself, where the true per-stakeholder effect $\Delta\xi$ is known.

6.3.1 Calibration

We first calibrate the model to the wave of bank decarbonization commitments following the 2015 Paris Agreement and use the calibrated economy as a data-generating process. Although the model accommodates general N stakeholder types, a disciplined calibration requires observable data on green shares, compensating differentials, and a well-identified purpose shock for each group. We focus on two stakeholder types—banks and workers—for which such data are available. Direct evidence on green shares comes from bank commitment data (Kacperczyk and Peydró (2022)) and worker surveys (Krueger, Metzger, and Wu (2021)), while compensating differentials are observable in both interest-rate spreads and wage discounts. Data on other stakeholder groups—such as suppliers and customers—are currently lacking in comparable form. We calibrate to the 500 most carbon-intensive publicly listed firms, primarily in the power sector.

Green shares. Green preferences are substantially more prevalent among workers than banks. Survey evidence indicates that 33% of workers accept green jobs at an average wage discount of 28% (Krueger, Metzger, and Wu (2021)). By contrast, Kacperczyk and Peydró (2022) report that only 7% of bank lending flows to green firms prior to the shock. This asymmetry in green shares is reflected in compensating differentials: workers accept large wage discounts while banks show only small interest-rate differentials—as the model predicts when green workers are relatively more abundant than green banks.

[9] What would break LATE—a direct path from a neighbor's treatment to a firm's own emissions, $Y_i = Y_i(D_i, \mathbf{D}_{-i})$—is precisely absent here, because homogeneity of $\Delta\xi$ means a re-matched firm carries the full treatment effect with it rather than exporting an externality onto an untreated one.

We set $\lambda_1 = 33\%$ (workers) and $\lambda_2 = 7\%$ (banks) at $t = 0$. Bank talent parameters are calibrated on $[1.1, 1.4]$ to match asset and debt distributions from bank loan data. Following Branikas et al. (2022), worker talent is distributed with $\gamma_1 = -0.4$ on $[0.08, 0.3]$, and firm productivity follows $\gamma_3 = 5$ on support [5,000, 100,000].

Emissions and mitigation. The harm–productivity link is set at $\sigma = 5{,}000$, calibrated from the ratio of carbon emissions to firm revenues in Trucost data. The mitigation cost $c = 0.00008$ is based on carbon-capture cost surveys, implying $c\sigma = 0.04$. The remaining parameters $(\kappa, \rho, \xi_1^*, \xi_2^*)$ are chosen to fit the observed production–emissions relationship (Figure 5). Table 1 collects the calibration.

Table 1: Calibration to post-Paris bank decarbonization. The green shares, wage discount, and harm–productivity ratio are matched directly to external data; the shock size matches the observed rise in green-bank commitments; the remaining production parameters are fit to the production–emissions relation.

Parameter	Value	Source / target
Green worker share λ_1	0.33	Krueger, Metzger, and Wu (2021) survey
Green bank share $(t = 0)$ λ_2	0.07	Kacperczyk and Peydró (2022)
Worker wage discount	28%	Krueger, Metzger, and Wu (2021)
Harm–productivity link σ	5,000	Trucost emissions/revenue
Mitigation cost c	0.00008	carbon-capture cost surveys
Implied product $c\sigma$	0.04	implied
Worker talent γ_1	-0.4 on $[0.08, 0.3]$	Branikas et al. (2022)
Bank talent	$[1.1, 1.4]$	bank-loan asset/debt distributions
Firm productivity γ_3	5 on [5,000, 100,000]	production–emissions fit
Purpose shock δ_2	0.08 (7%→15%)	Paris commitment wave

6.3.2 Results

Given the calibrated economy at $t = 0$, we model a purpose shock that increases the green-bank share from $\lambda_2 = 7\%$ to $\lambda_2 = 15\%$ (i.e., $\delta_\ell = 8\%$) to capture the wave of bank commitments following the Paris Agreement. We then simulate the matching outcome for both periods, subject to the minimal-adjustment rule.

The distribution of the n-index before and after the shock. The shock reshapes sorting unevenly across the size distribution—engagement at the bottom, exit at the top, and reallocation spillovers in the middle—and this unevenness is precisely what will bias the firm-level design. Figure 6 displays the aggregate distribution of (x, n_t) before and after the shock. Conditional on firm size, the share of each color (brown, grey, and green) represents the share of firms with index $n_\ell = 0, 1, 2$ in each period t. Observe that, at the bottom of the firm distribution, exactly 8% of firms transition from $n_0 = 1$ to $n_1 = 2$, equal to the treated share.

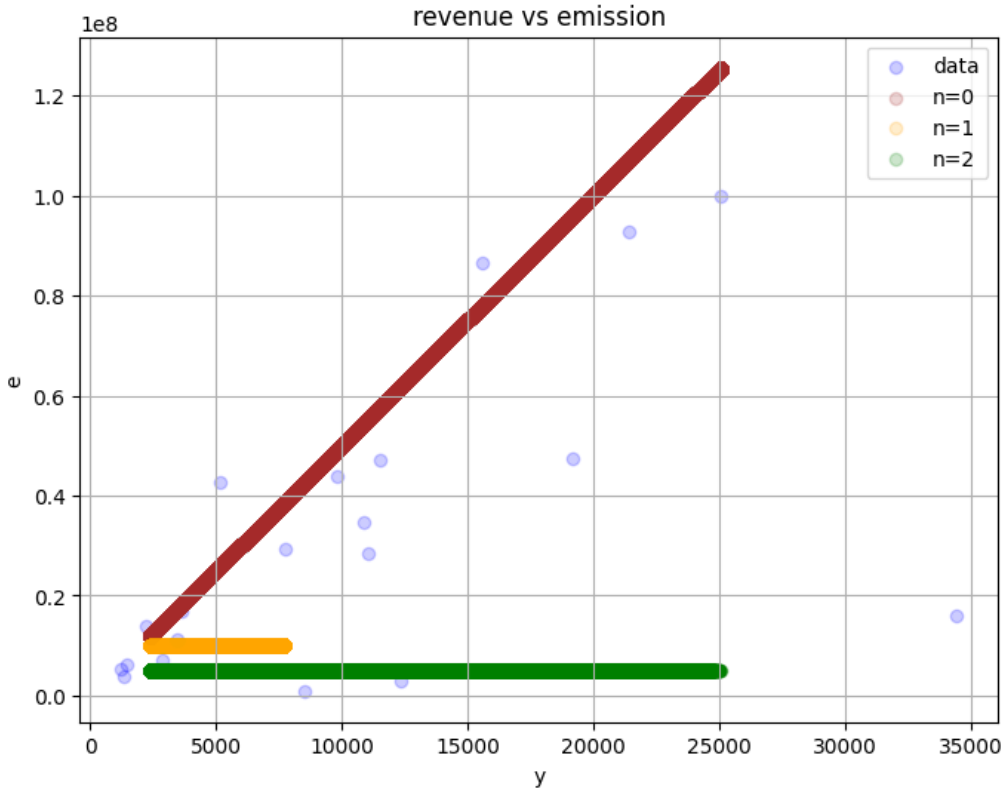


Figure 5: Calibrated production–emission relation.

At the top, however, most brown firms (above 800) remain brown. One can also see that in the middle of the distribution (firms ranked 400–600), more than 8% of firms move from $n_0 = 1$ to $n_1 = 2$.

We now apply the estimators of Section 6.2 to the simulated panel at the firm level, grouping firms into fine size bins.

Take-up rate across groups. The first stage makes the trickle-down visible: impact is limited at the top and reallocates to the middle of the size distribution. Figure 7 reports the take-up rate—the probability that a firm is matched with a green bank after the shock—for both groups. This distributional effect can be seen clearly from the relationship between the average take-up rate across groups (the blue line) and the total take-up rate (the grey dashed line), which is $\frac{\delta}{1-\lambda_\ell} = \frac{0.08}{1-7\%} = 0.086$.[10] Specifically, the average take-up is close to zero and below 8.6% at the bottom, but increases sharply to more than 20% in the middle region (buckets 7 to 11). As expected, since there is no spillover in the lower region, the blue line converges to 8.6%.

Second, the figure also displays the compliance rate, $D_T(x) - D_C(x)$. As expected, in the

[10]We know that there will be exactly 8% more firms matched with green banks. Hence, the total take-up rate is the probability that a firm turns green at $t = 1$ conditional on not having a green bank at $t = 0$.

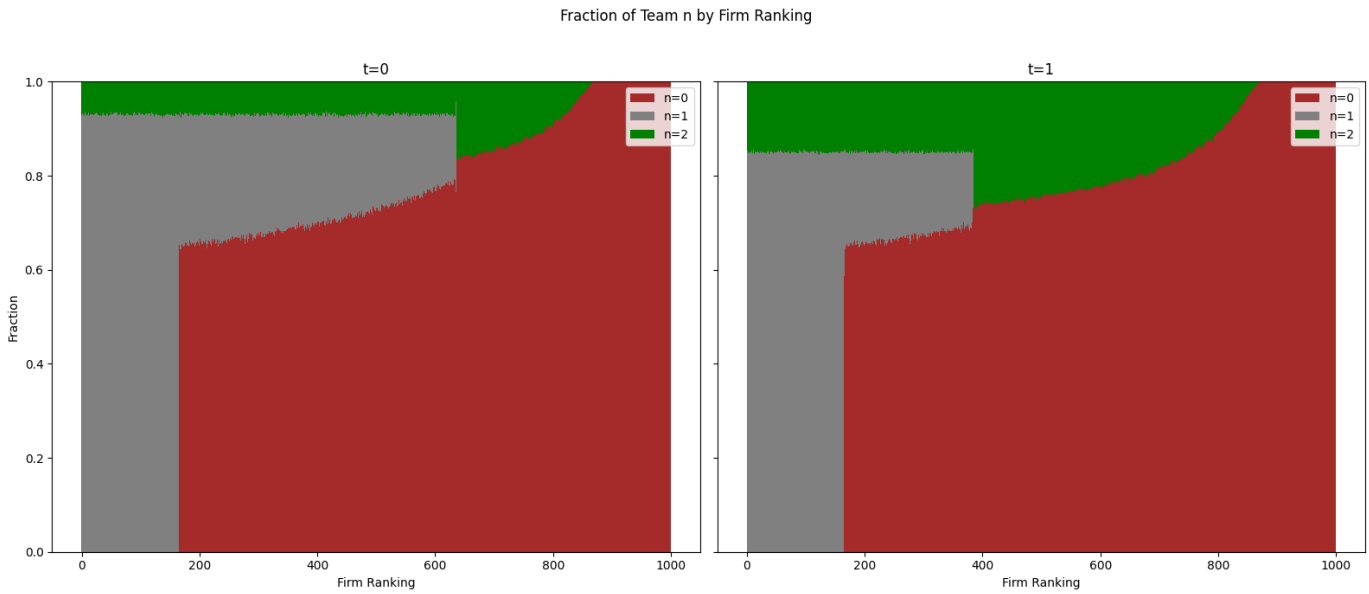


Figure 6: Firm matching outcomes before (left panel) vs. after the shock (right panel).

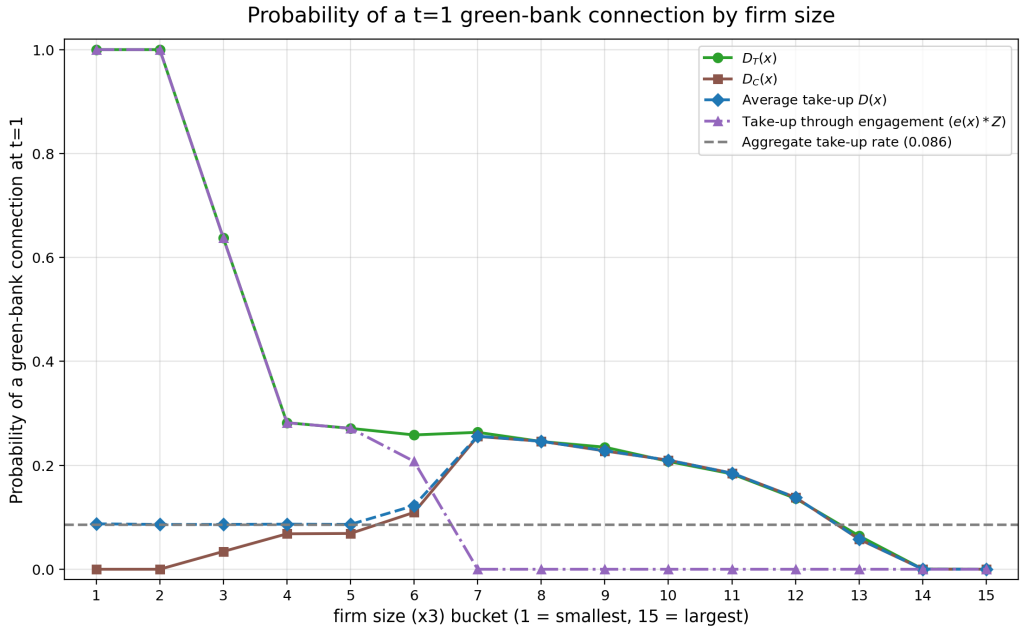


Figure 7: First stage: take-up of a green bank by firm size. The average take-up rate (blue) is near zero at the bottom, rises above 20% in the middle—where exit relocates green banks down the size distribution—and averages the aggregate rate $\delta/(1-\lambda_\ell) = 8.6\%$ (grey dashed). The compliance rate $D_T(x) - D_C(x)$, the treatment-minus-control take-up gap, is zero at the top (no compliers, since treated stakeholders all exit) and positive below the cutoff, where it coincides with the engagement rate (purple).

exit region there is no gap at the top, so $D_T(x) = D_C(x)$. Below the cutoff, one can see that the gap—i.e., the compliance rate $D_T(x) - D_C(x)$—is generally positive, as given by Equation (19). Moreover, the take-up rate $D_T(x)$ is also the engagement rate (represented by the purple line) in this region. Note that, since the very smallest firms must be matched with green workers

before the shock (i.e., there are no brown firms at the bottom in Figure 6), all treated firms in this region engage, as $\pi_{\ell-1}(x) = 1$ and $D_T(x) = 1$. According to Equation (19), the compliance rate is also one, since all treated stakeholders are matched with treated firms and thus the take-up rate for the control group is zero.

The bias and its correction. Figure 8 summarizes the compliance-adjusted estimation; Appendix A.12 details the data-generating process, the noise layers, and the estimators.

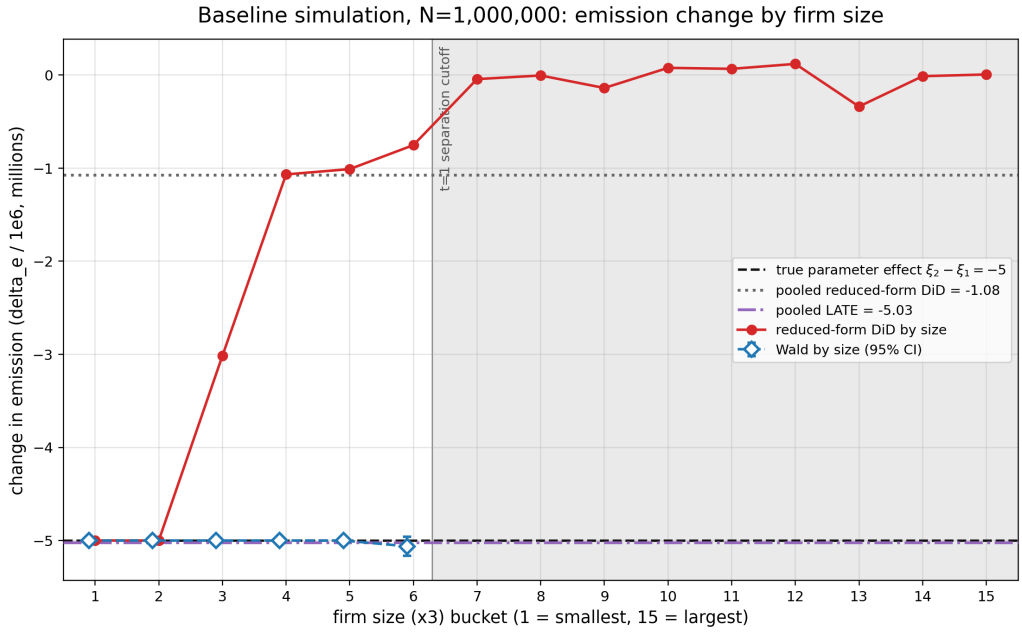


Figure 8: Recovering the per-stakeholder effect from the calibrated model. *Top*: the reduced-form difference-in-differences (red) is attenuated. The compliance-adjusted pooled LATE recovers the true per-stakeholder effect $\Delta\xi_\ell$. The Wald estimator by size, LATE(x), recovers the true per-stakeholder effect $\Delta\xi_\ell$ only for relatively small firms and is unidentified for larger firms.

The central quantitative result is that the standard design badly understates impact: the pooled firm-level difference-in-differences recovers only about *one-fifth* of the true per-stakeholder effect $\Delta\xi_\ell$, so a researcher using it would conclude the effect is roughly 80% smaller than it is. The attenuation is not uniform—it is greatest at the top of the size distribution, where exit relocates impact entirely off the treated firms, and vanishes at the bottom, where stakeholders engage. Because the bias scales with the exit share—the fraction of impact that reallocation moves off the treated set—its magnitude is an economic object rather than an artifact of the calibration: the more top-heavy the size distribution, the larger the understatement.

This reframes the empirical literature. Kacperczyk and Peydró (2022) find that post-Paris bank decarbonization commitments had a null firm-level effect on borrowers' emissions; our calibration to that same episode shows that such a null is precisely the footprint a genuine effect leaves once exit is accounted for—not evidence that the effect is absent, but evidence

of where it went. The compliance-adjusted pooled LATE recovers the true per-stakeholder effect $\Delta\xi_\ell$ in full, implying an underlying effect several times the naive difference-in-differences estimate, while the size-conditional Wald LATE(x) recovers it only for relatively small firms and is unidentified above the cutoff, where the compliance rate collapses to zero.

Taking the estimator to data. The estimator maps directly onto observables. The instrument Z is a firm's pre-period exposure to the purpose shock—for instance, a lending relationship at $t = 0$ with a bank that subsequently makes a decarbonization commitment; the endogenous treatment D is whether the firm is connected to a committed (green) bank at $t = 1$; and the outcome Y is the change in firm-level emissions. Lending relationships and their breaks are observable in syndicated-loan data (Kacperczyk and Peydró (2022)), commitments in banks' net-zero pledges, and emissions in EU ETS or EPA facility reporting. The pooled Wald ratio of the emissions change on the change in green-bank connection, instrumented by pre-period exposure, then recovers the per-stakeholder mitigation effect that a firm-level difference-in-differences on the same panel understates.

7 Conclusion

A purpose-driven stakeholder who dislikes a firm's social harm can leave for a cleaner firm or stay and pay for mitigation, and which she does depends on what the market offers. We have modeled that choice as the outcome of a matching market in which a rise in purpose reallocates stakeholders across firms, and shown that exit and engagement coexist, as in the data, only when stakeholders come in multiple groups and harm scales with firm size. Remove either feature and the economy collapses to a corner—engagement alone, or exit alone. When both features are present, large stakeholders exit while smaller ones engage in the same equilibrium. Engagement changes a firm directly; exit changes it only indirectly, as displaced stakeholders reallocate and induce mitigation at the firms they move to, spilling across the size distribution and across stakeholder groups.

This coexistence has a direct consequence for measurement. Because exit registers as non-compliance, a firm-level difference-in-differences absorbs the spillover into its control group and understates the per-stakeholder mitigation effect of a rise in purpose—so exit looks ineffective when it is only mismeasured. We show that a compliance-adjusted local average treatment effect recovers the effect despite the spillover. A calibration to the wave of bank decarbonization commitments following the 2015 Paris Agreement illustrates these dynamics, rationalizing at once the prevalence of exit by large banks, the null firm-level results in the data, and the viability of LATE to recover the per-stakeholder mitigation effect.

References

Acemoglu, Daron, Ufuk Akcigit, and William R Kerr (2016). "Innovation network". In: *Proceedings of the National Academy of Sciences* 113.41, pp. 11483–11488.

Akey, Pat and Ian Appel (2019). "Environmental externalities of activism". In: *Available at SSRN 3508808*.

Andreoni, James (1990). "Impure altruism and donations to public goods: A theory of warm-glow giving". In: *The economic journal* 100.401, pp. 464–477.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables". In: *Journal of the American Statistical Association* 91.434, pp. 444–455.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". In: *American economic review* 96.5, pp. 1652–1678.

Berk, Jonathan B. and Jules H. van Binsbergen (2025). "The Impact of Impact Investing". In: *Journal of Financial Economics*. SSRN 3909166.

Besley, Timothy and Maitreesh Ghatak (2005). "Competition and incentives with motivated agents". In: *American economic review* 95.3, pp. 616–636.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin (2025). "Sorting with Teams". In: *Journal of Political Economy* 133.2, pp. 421–454.

Bonnefon, Jean-François et al. (2025). "The moral preferences of investors: Experimental evidence". In: *Journal of Financial Economics* 163, p. 103955.

Branikas, Ioannis et al. (2022). *Sustainability Preferences of Talented Employees*. Tech. rep. SSRN Working Paper.

Broccardo, Eleonora, Oliver Hart, Luigi Zingales, et al. (2022). "Exit versus voice". In: *Journal of Political Economy* 130.12, pp. 3101–3145.

Chaisemartin, Clément de and Xavier D'Haultfœuille (2018). "Fuzzy Differences-in-Differences". In: *The Review of Economic Studies* 85.2, pp. 999–1028.

Chiappori, Pierre-André, Robert McCann, and Brendan Pass (2016). "Multidimensional matching". In: *arXiv preprint arXiv:1604.05771*.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2018). "Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market". In: *Journal of the European Economic Association* 16.1, pp. 161–198.

Duchin, Ran, Janet Gao, and Qiping Xu (2022). "Sustainability or greenwashing: Evidence from the asset market for industrial pollution". In: *Available at SSRN 4095885*.

Dupuy, Arnaud and Alfred Galichon (2014). "Personality traits and the marriage market". In: *Journal of Political Economy* 122.6, pp. 1271–1319.

Edmans, Alex, Doron Levit, and Jan Schneemeier (2022). *Socially Responsible Divestment*. ECGI Finance Working Paper 823/2022. CEPR Discussion Paper DP17262; SSRN 4093518. European Corporate Governance Institute.

Ellison, Glenn and Edward L Glaeser (1997). "Geographic concentration in US manufacturing industries: a dartboard approach". In: *Journal of political economy* 105.5, pp. 889–927.

Gabaix, Xavier and Augustin Landier (2008). "Why has CEO pay increased so much?" In: *The Quarterly Journal of Economics* 123.1, pp. 49–100.

Galichon, Alfred (2016). *Optimal transport methods in economics*. Princeton University Press.

Green, Daniel and Benjamin N. Roth (2025). "The Allocation of Socially Responsible Capital". In: *The Journal of Finance* 80.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings". In: *Journal of political economy* 118.3, pp. 536–598.

Hartzmark, Samuel M and Kelly Shue (2022). "Counterproductive sustainable investing: The impact elasticity of brown and green firms". In: *Available at SSRN 4359282*.

Heinkel, Robert, Alan Kraus, and Josef Zechner (2001). "The effect of green investment on corporate behavior". In: *Journal of financial and quantitative analysis* 36.4, pp. 431–449.

Hirschman, Albert O (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard university press.

Hong, Harrison and Marcin Kacperczyk (2009). "The price of sin: The effects of social norms on markets". In: *Journal of financial economics* 93.1, pp. 15–36.

Hong, Harrison, Neng Wang, and Jinqiang Yang (2021). *Welfare consequences of sustainable finance*. Tech. rep. National Bureau of Economic Research.

Huber, Kilian (2023). "Estimating general equilibrium spillovers of large-scale shocks". In: *The Review of Financial Studies* 36.4, pp. 1548–1584.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, pp. 467–475.

Kacperczyk, Marcin T and José-Luis Peydró (2022). "Carbon emissions and the bank-lending channel". In: *Available at SSRN 3915486*.

Krueger, Philipp, Daniel Metzger, and Jiaxin Wu (2021). "The sustainability wage gap". In: *Swedish House of Finance Research Paper* 20-14, pp. 21–17.

Lindenlaub, Ilse (2017). "Sorting multidimensional types: Theory and application". In: *The Review of Economic Studies* 84.2, pp. 718–789.

Oehmke, Martin and Marcus M Opp (2023). "A theory of socially responsible investment". In: *Swedish House of Finance Research Paper* 20-2.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor (2021). "Sustainable investing in equilibrium". In: *Journal of Financial Economics* 142.2, pp. 550–571.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski (2021). "Responsible investing: The ESG-efficient frontier". In: *Journal of Financial Economics* 142.2, pp. 572–597.

Riedl, Arno and Paul Smeets (2017). "Why do investors hold socially responsible mutual funds?" In: *the Journal of Finance* 72.6, pp. 2505–2550.

Sattinger, Michael (1979). "Differential rents and the distribution of earnings". In: *Oxford Economic Papers* 31.1, pp. 60–71.

Tervio, Marko (2008). "The difference that CEOs make: An assignment model approach". In: *American Economic Review* 98.3, pp. 642–68.

A Appendix

A.1 Proof of Lemma 1

Proof. Given that $y - cm - n\psi\big(\phi(y) - m\big)$ is affine in n,the surplus

$$\Omega(y, n) = \max_{m \geq 0} y - cm - n\psi\big(\phi(y) - m\big) \tag{A.1}$$

is decreasing and convex in n. Because $\Omega(y, n)$ is convex in n, it satisfies increasing differences in (n, θ) when evaluated at $n + \theta$. That is, for any $n' > n$ and $\theta' > \theta$,

$$\Omega(y, n' + \theta') - \Omega(y, n + \theta') \geq \Omega(y, n' + \theta) - \Omega(y, n + \theta). \tag{A.2}$$

Hence, $\Omega(y, n_{-\ell} + \theta)$ has increasing differences in $(n_{-\ell}, \theta)$. According to Equation 6, by monotone comparative statics (Topkis' theorem), the optimal choice $n^*_{-\ell}(x, \theta)$ must be weakly increasing in θ. Therefore, conditional on productivity x, agents with higher preference type θ are matched with the team with weakly higher n. □

A.2 Proof for Lemma 2

Proof. Since Equation 9 implies complementarity between green agent $(x_\ell, 1)$ and $y_{-\ell}$, hence, by the monotonic comparative statics, a green agent with higher ability must choose a team with a higher productivity than a green agent with lower ability. Similarly, Equation 7 implies complementarity between brown agent $(x_\ell, 0)$ and $z(y_{-\ell}, n_{-\ell})$; hence, a more skilled brown agent must choose a team with a higher z-index. □

A.3 Proof for Proposition 1

Proof. Let γ denote a feasible assignment on A^{N+1}, with $a_\ell = (x_\ell, \theta_\ell)$. Output and the number of purpose-driven stakeholders are given by $y(\boldsymbol{x}) = \prod_{\ell=1}^{N+1} x_\ell$, and $n(\boldsymbol{\theta}) = \sum_{\ell=1}^{N+1} \theta_\ell$.

The planner solves

$$\max_{\gamma} \int \big[y(\boldsymbol{x}) - C(n(\boldsymbol{\theta}))\big] d\gamma, \tag{A.3}$$

where

$$C(n) \equiv \min_{m} \{cm + n\psi(\phi_0 - m)\},$$

which is independent of y when $\sigma = 0$. Under independence, $F(x, \theta) = F_x(x)F_\theta(\theta)$. Hence the feasible set factorizes as

$$\Upsilon(F, \ldots, F) = \Upsilon(F_x, \ldots, F_x) \times \Upsilon(F_\theta, \ldots, F_\theta). \tag{A.4}$$

Because the surplus is additively separable, the problem decomposes:

$$\max_{\gamma} \int y(\boldsymbol{x})\, d\gamma^x - \min_{\gamma^\theta} \int C(n(\boldsymbol{\theta}))\, d\gamma^\theta. \tag{A.5}$$

Since $y(\boldsymbol{x})$ is supermodular in $\boldsymbol{x}$, positive assortative matching in productivity follows from standard matching arguments. Since $-C(n)$ is convex, the matching is positively assortative in θ. Let λ_ℓ denote the mass of stakeholders with type weakly above ℓ. Positive sorting implies that firms with at least ℓ purpose-driven stakeholders have measure λ_ℓ. Therefore firms with index ℓ have measure $\lambda_\ell - \lambda_{\ell+1}$, and firms with index 0 have measure $1 - \lambda_1$.

Compensating differentials. Firms must be indifferent between a team with index ℓ and one with index $\ell - 1$. Recall that type ℓ stakeholders must work for firms with at least $n \geq \ell$ purpose-driven stakeholders. Hence, a team with index ℓ (respectively, $\ell - 1$) consists of purpose-driven stakeholders with types $k \leq \ell$ (respectively, $k \leq \ell - 1$) and profit-driven stakeholders with types $k > \ell$ (respectively, $k > \ell - 1$).
Conditional on y, indifference between a team of index ℓ and $\ell - 1$ is therefore equivalent to the firm being indifferent between matching a purpose-driven and a profit-driven stakeholder of type ℓ. This implies

$$\Omega(y, \ell-1) - \left\{ \sum_{k=1}^{\ell-1} U_k(x,1) + \sum_{k=\ell}^{N} U_k(x,0) \right\} = \Omega(y,\ell) - \left\{ \sum_{k=1}^{\ell} U_k(x,1) + \sum_{k=\ell+1}^{N} U_k(x,0) \right\}. \tag{A.6}$$

Rearranging yields

$$U_\ell(x,0) - U_\ell(x,1) = \Omega(y, \ell-1) - \Omega(y,\ell) \tag{A.7}$$

Since $p_\ell(x,0) = U_\ell(x,0)$ and $p_\ell(x,1 \mid \ell) = U_\ell(x,1) + \psi(\xi^*_\ell)$, it follows that

$$p_\ell(x,0) - p_\ell(x,1 \mid \ell) = c(\xi^*_{\ell-1} - \xi^*_\ell) - (\ell - 1)\left[\psi(\xi^*_{\ell-1}) - \psi(\xi^*_\ell)\right]. \tag{A.8}$$

□

A.4 Proof for Stable Matches at period ℓ

Let $G_\ell(z_\ell, n_\ell)$ denote the cumulative distribution function for teams with characteristics (z_ℓ, n_ℓ) at period ℓ, and let $G^0_\ell(z)$ and $G^1_\ell(z)$ denote the measure of teams with index $n_\ell = 0$ and $n_\ell \geq 1$, respectively, having productivity no larger than z.

Let $\eta^\theta_s(x_{\ell'})$ denote the productivity of the team matched with stakeholder $(x_{\ell'}, \theta)$ in the

full-separation region. The market-clearing conditions yield

$$G_\ell^0(\bar{y}_\ell) - G_\ell^0\big(\eta_s^0(x_{\ell'})\big) = (1-\lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x_{\ell'}))\,, \tag{A.9}$$
$$G_\ell^1(\bar{z}_\ell) - G_\ell^1\big(\eta_s^1(x_{\ell'})\big) = \lambda_{\ell'}(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x_{\ell'}))\,.$$

Cutoff characterization. The cutoff $\hat{x}_{\ell'}$ can be determined in two ways. For relatively small $\lambda_{\ell'} < \hat{\lambda}_{\ell'}$, the cutoff is given by

$$\eta_s^0(\hat{x}_{\ell'}) = \eta_s^1(\hat{x}_{\ell'}). \tag{A.10}$$

Otherwise, the cutoff solves

$$G_\ell^0(\bar{z}_\ell) - G_\ell^0(z_{\ell,L}^0) = (1-\lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(\hat{x}_{\ell'}))\,, \tag{A.11}$$

where $z_{\ell,L}^0$ denotes the lowest productivity among teams with index $n=0$. Define $z_{\tau+1} = \eta_s^0(\hat{x}_{\tau+1})\cdot\hat{x}_{\tau+1}$.

Lemma 6. *Assume that $G_{\ell-1}(z,n)$ satisfies Assumption 3, and there exists $\hat{z}_{\ell-1}$ such that for all $z \geq \hat{z}_{\ell-1}$, $n \in \{0,\, \ell-1\}$. The stable matching between (z,n) and (x,θ) can be characterized by a cutoff $\hat{x}_\ell$ with the following structure.*

1. ***Full separation above the cutoff.*** *For $x_\ell > \hat{x}_\ell$, teams with $n_\ell = 0$ are matched exclusively with stakeholders of type $\theta = 0$, and teams with $n_\ell = \ell - 1$ are matched exclusively with stakeholders of type $\theta = 1$. Conditional on θ, the matching is positively assortative (PAM) in team productivity $z_{\ell-1}$ and stakeholder ability x_ℓ. Let $\eta_s^\theta(x_\ell)$ denote the z-index of the team matched with stakeholder (x_ℓ,θ) in this region. The market-clearing conditions are*

$$G_{\ell-1}^0(\bar{z}) - G_{\ell-1}^0\big(\eta_s^0(x_\ell)\big) = (1-\lambda_\ell)(F_\ell(\bar{x}_\ell) - F_\ell(x_\ell))\,, \tag{A.12}$$
$$G_{\ell-1}^1((1-c\sigma)\bar{z}) - G_{\ell-1}^1\big(\eta_s^1(x_\ell)\big) = \lambda_\ell(F_\ell(\bar{x}_\ell) - F_\ell(x_\ell))\,. \tag{A.13}$$

2. ***Mixing below the cutoff.*** *For $x_\ell \leq \hat{x}_\ell$, the matching is positively assortative in $z_{\ell-1}$ and x_ℓ. Conditional on $(z_\ell, x_{\ell+1})$, every value-driven stakeholder $(x_\ell, 1)$ is matched with a team of index $n = \ell - 1$. Let $\phi_m(x_\ell)$ denote the assignment function. Then*

$$G_{\ell-1}^0(\phi_m(x_\ell)) + G_{\ell-1}^1(\phi_m(x_\ell)) = F_\ell(x_\ell).$$

Proof. Let ξ_n^* solve $c = n\psi'(\xi_\ell^*)$, and thus $m^*(y,n) = \phi(y) - \xi_\ell^*$. For $n \geq 1$, the surplus can be

rewritten as

$$\begin{aligned}\Omega(y,n) &= y - c(\phi(y) - \xi_\ell^*) - n\psi(\xi_n^*) \\ &= (1-c\sigma)y - \{c(\phi_0 - \xi_\ell^*) + n\psi(\xi_n^*)\} \\ &= (1-c\sigma)y - C(n)\end{aligned} \tag{A.14}$$

The planner's objective can be written as

$$W = \max_{\gamma\in\Gamma} \int [\tilde{z}(y,n,\theta) - C(\theta+n)]\, d\gamma,$$

where Γ represents the permissible set and

$$\tilde{z}(y,n,\theta) = \begin{cases} y & \text{if } n+\theta = 0, \\ (1-c\sigma)y & \text{if } n+\theta \geq 1, \end{cases}$$

The proof proceeds in three steps.

Step 1: Value-driven stakeholders are never matched with $n = 0$ teams. We claim that $\gamma(x, 1, (y, 0)) = 0$ for all x and y. Because profit-driven stakeholders $(x, 0)$ are in excess supply, it is always feasible to match every $n = 0$ team with a $\theta = 0$ stakeholder. Any matching with $\gamma(x, 1, (y, 0)) > 0$ would yield a lower effective productivity $\tilde{z}(y, 0, 1) = (1 - c\sigma)y < y = \tilde{z}(y, 0, 0)$, thereby shifting the distribution of effective productivity downward in the sense of first-order stochastic dominance. This strictly reduces welfare.

Step 2: Reduction to constrained one-dimensional sorting. Since every $n = 0$ team must be matched with a $\theta = 0$ stakeholder, the effective productivity within each match simplifies to $z(y, n)\, x$. Indeed, for $n \geq 1$ we have $\tilde{z}(y, n, \theta) = (1 - c\sigma)y = z(y, n)$ regardless of θ, and for $n = 0$ we have $\tilde{z}(y, 0, 0) = y = z(y, 0)$. The objective thus reduces to

$$W = \max_{\gamma\in\Gamma_C} \int [z(y,n)\, x - C(n+\theta)]\, d\gamma, \tag{A.15}$$

subject to the constraint $\gamma(\{\theta = 1,\, n = 0\}) = 0$.
The constructed solution minimizes $\int C(\theta + n)\, d\gamma$, since all value-driven stakeholders $(x, 1)$ are matched with teams of index $n = \ell - 1$, yielding the most extreme feasible distribution of $n + \theta$. It therefore remains to verify optimality with respect to the first term, $\int z(y, n)\, x\, d\gamma$.

The supermodularity of $z \cdot x$ implies that the admissible optimum preserves PAM wherever the type-exclusion constraint is slack and departs from it minimally where the constraint binds. By the standard theory of optimal transport, the optimal assignment satisfies monotonicity and no-mass splitting conditional on θ. That is, letting $z^*(x, \theta)$ denote the equilibrium z-index

matched with stakeholder (x, θ), the function $z^*(\cdot, \theta)$ is strictly increasing in x, and, conditional on θ, each stakeholder type is matched to a unique z-index with no mass split across multiple partners.

Moreover, for $n \geq 1$ the constraint does not bind, so there exists a unique ability level x matched to each team (y, n). Hence, if $\gamma((x, 0), (y, n)) > 0$ for some $n \geq 1$, then $\gamma((x, 1), (y, n)) > 0$ as well: both stakeholder types at the same ability level x are assigned to the same team whenever $n \geq 1$.

Step 3: Separation at the top. The optimal matching features full separation at the top of the productivity distribution. Specifically, the highest-ability profit-driven stakeholder must be matched with the most productive $n = 0$ team. To see this, note that for $\sigma > 0$ there exists $\epsilon > 0$ such that $z(y, 0) > z(\bar{y}, n')$ for all $y > \bar{y} - \epsilon$ and $n' \geq 1$. Matching these top $n = 0$ teams with top $\theta = 0$ stakeholders preserves PAM without violating the type-exclusion constraint. The highest-ability value-driven stakeholders, on the other hand, are then matched with the most productive teams having $n' \geq 1$.

Since profit-driven stakeholders are in excess supply, some of them must be matched with teams of index $n \geq 1$. The key decision reduces to identifying which profit-driven stakeholders join such teams. If stakeholder $(x, 0)$ is matched with a team (y, n) with $n \geq 1$, then optimality requires that it be paired with the team currently assigned to $(x, 1)$, which has productivity index $z^*(x, 1)$. The payoff difference between matching x to its $n = 0$ team (with index $z^*(x, 0)$) versus the $n \geq 1$ team (with index $z^*(x, 1)$) is $\left[z^*(x, 0) - z^*(x, 1)\right] x$. Under Assumption 3,

$$\frac{(1 - \lambda_\ell)\, f_\ell(x)}{\lambda\, f_\ell(x)} > \frac{g^0_{\ell-1}(z)}{g^1_{\ell-1}(z)},$$

and therefore, at any point where $\eta^0_s(x_\ell)$ and $\eta^1_s(x_\ell)$ coincide at some productivity level z, the derivatives satisfy

$$\frac{d\eta^0_s(x_\ell)}{dx} = \frac{(1 - \lambda_\ell)\, f_\ell(x)}{g^0_{\ell-1}(\eta^0_s(x_\ell))} > \frac{\lambda_\ell\, f_\ell(x)}{g^1_{\ell-1}(\eta^1_s(x_\ell))} = \frac{d\eta^1_s(x_\ell)}{dx}.$$

It follows that the two assignment functions can cross at most once, at a point $\hat{x}_\ell$, and that $\eta^0_s(x_\ell) - \eta^1_s(x_\ell) > 0$ for all $x > \hat{x}_\ell$. In the interior case, the cutoff is determined by

$$z^*(\hat{x}_\ell, 0) = \eta^0_s(\hat{x}_\ell) = \eta^1_s(\hat{x}_\ell) = z^*(\hat{x}_\ell, 1),$$

and full separation is strictly optimal for all $x > \hat{x}_\ell$. If no such interior crossing exists, it is optimal to maintain separation until all $n = 0$ teams are exhausted. □

A.5 Proof for Lemma 3 and 4

Proof. In the case that $N = 2$, then $G^0_\ell(y) = (1-\lambda_\ell)F_\ell(y)$ and $G^1_\ell(y) = \lambda_\ell F_\ell(y)$. Hence, Proposition follows directly from Lemma 6. The balanced supply is a special case where Equation A.11 can be written as

$$(1-\lambda_\ell)(F_\ell(\bar{x}_\ell) - F_\ell(\underline{x_\ell})) = (1-\lambda_\ell) = (1-\lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x^0_{\ell',L})),$$

and thus $x^0_{\ell',L} = \underline{x_{\ell'}}$ when $\lambda_\ell = \lambda_{\ell'}$. Hence, full separation between stakeholders must hold. □

A.6 Properties of Matches

Lemma 7 (Properties of the Evolution)**.** *Let* $\hat{i}_\ell \equiv F_\ell(\hat{x}_\ell)$.

1. *For* $i \le \hat{i}_\ell$*, the ranking of the z-index for* $x_\ell[i]$ *remains unchanged after the match. For* $i > \hat{i}_\ell$*, the ranking of the z-index of* $x_\ell[i]$ *increases for profit-driven stakeholders and decreases for purpose-driven stakeholders.*

2. *The separation region* $(1-\hat{i}_\ell)$ *is decreasing in* ℓ.

3. *For all* $n_{\ell-1} < \ell - 1$*, we have* $n_\ell = n_{\ell-1}$*. The measure of teams with index* $n_\ell = \ell$ *after the match is* λ_ℓ.

Proof. Part (1). For $z_{\ell-1} \le z_{\ell-1}[\hat{i}_\ell]$, positive assortative matching between $(z_{\ell-1}, x_\ell)$ in the mixing region implies that $z_\ell[i] = z_{\ell-1}[i] \cdot x_\ell[i]$. Hence,

$$\Pr(z < z_\ell[i]) = \Pr(x < x_\ell[i]) = i.$$

For those in the separation region, we have $z^*(x,0) > z^*(x,1)$, and thus for a profit-driven stakeholder,

$$z_\ell[i] = z^*(x_\ell[i],0) \cdot x_\ell[i] > z^*(x_\ell[i],1) \cdot x_\ell[i].$$

Hence, for a profit-driven stakeholder with ranking i,

$$\Pr(z < z_\ell[i]) > \lambda_\ell\, F_\ell(x_\ell[i]) + (1-\lambda_\ell)\, F_\ell(x_\ell[i]) = i,$$

and vice versa for purpose-driven stakeholders.

Part (2). We show that $\hat{i}_\ell$ is increasing in ℓ. Assuming that the interior cutoff holds, at the cutoff $\hat{i}_\ell$ we must have

$$\frac{G^0_{\ell+1}(\bar{z}_{\ell+1}) - G^0_{\ell+1}\big(z_{\ell+1}(\hat{i}_\ell)\big)}{G^1_{\ell+1}(\bar{z}_{\ell+1}) - G^1_{\ell+1}\big(z_{\ell+1}(\hat{i}_\ell)\big)} = \frac{G^0_\ell(\bar{z}_\ell) - G^0_\ell\big(z(\hat{i}_\ell)\big)}{G^1_\ell(\bar{z}_\ell) - G^1_\ell\big(z(\hat{i}_\ell)\big)} = \frac{1-\lambda_\ell}{\lambda_\ell} < \frac{1-\lambda_{\ell+1}}{\lambda_{\ell+1}},$$

where the first equality uses the fact that $G^0_{\ell+1}\big(z_{\ell+1}(\hat{i}_\ell)\big) = G^0_\ell\big(z(\hat{i}_\ell)\big) = \hat{i}_\ell$. Since the ratio

$$\frac{G^0_{\ell+1}(\bar{z}_{\ell+1}) - G^0_{\ell+1}(z_{\ell+1})}{G^1_{\ell+1}(\bar{z}_{\ell+1}) - G^1_{\ell+1}(z_{\ell+1})}$$

is increasing in $z_{\ell+1}$, it follows that $\hat{i}_{\ell+1} > \hat{i}_\ell$. If instead the cutoff is given by the corner solution, then $\hat{i}_\ell$ is trivially increasing in ℓ.

Part (3). We show that the measure of teams with index $n_\ell = \ell$ after the match is λ_ℓ. It suffices to show that $(x_\ell, 1)$ joins only teams with $n_{\ell-1} = \ell - 1$. This holds by construction in the separation region, so it remains to verify it in the mixing region. Since our construction implies that $(x_\ell, 1)$ always joins the team with the highest $n_{\ell-1}$, we only need to show that there are always relatively more teams with $n_{\ell-1} = \ell - 1$. This is indeed the case, since

$$\frac{G_{\ell-1}(z,\, \ell-1)}{\sum_{\tilde{n}} G_{\ell-1}(z,\, \tilde{n})} \geq \lambda_{\ell-1} \geq \lambda_\ell,$$

where the first inequality uses the fact that there is at least a $\lambda_{\ell-1}$ measure of teams with index $\ell - 1$, as purpose-driven stakeholders ranking after matches is weakly lower than profit-driven stakeholders, according to property (1). □

A.7 Proof for Proposition 3

Proof. We show that Lemma 6 can be applied at any stage N under the sequential ordering. Since the property (3) of Lemma 7 implies that the distribution of n must maximize the second term, we focus on the first term.

Recall that, conditional on $n_{\tau-1} > 0$ (respectively, $n_{\tau-1} = 0$), the team with a higher z-index must be matched with a more productive stakeholder x_τ. Let $x^1_\tau(z_{\tau-1})$ and $x^0_\tau(z_{\tau-1})$ denote the productivity of the stakeholder matched with the team of index z_τ, conditional on $n_{\tau-1} > 0$ and $n_{\tau-1} = 0$, respectively. The first term of the optimization problem can thus be rewritten as

$$\max_{\gamma \in \Gamma_C(\mu,\nu)} \int \Gamma^\theta(z_{\tau-1}\, x_\tau)\, d\gamma, \tag{A.16}$$

where

$$\Gamma^\theta(z_\tau) = z_\tau \prod_{\ell=1}^{N+1} x^\theta_{\tau+\ell}(z_{\tau+\ell-1}),$$

with $z_{\tau+1} = z_\tau\, x^\theta_{\tau+1}(z_\tau)$ and $\theta \in \{0, 1\}$. Importantly, since all purely profit-driven teams must remain purely profit-driven by construction, the recursive relation $z_{\tau+1} = z_\tau\, x^\theta_{\tau+1}(z_\tau)$ holds throughout.

At period N, $x^\theta_N(z_{N-1})$ is given by Lemma 6 and is strictly increasing in z. Hence, $\Gamma^\theta(z_{N-1})$

is a monotone transformation of z_{N-1} for both values of θ. It follows that, at period $N-1$, any solution γ that maximizes (A.15)—where the product is given by $z_{N-1} = z_{N-2}\, x_{N-1}$—must also maximize (A.16), where $x^{\theta}_{N-1}(z_{N-2})$ is characterized by Lemma 6.

We proceed by induction. Suppose that $x^{\theta}_{t+1}(z_t)$ is characterized by Lemma 6 at period $t+1$. Then the matching at period t must maximize $\Gamma^{\theta}(z_{t-1}\, x_t)$, where $\Gamma^{\theta}_z(z_\tau) > 0$. Hence, the solution at period t is again characterized by Lemma 6. This shows that our sequential algorithm is optimal. □

A.8 Proof for Proposition 5

Proof. Let $\hat{x}_{\ell,0}$ and $\hat{x}_{\ell,1}$ denote the cutoff before and after the shock, respectively. Since the separation region must expand, we have $\hat{x}_{\ell,0} > \hat{x}_{\ell,1}$.

For all stakeholders with ability ranking $x_\ell \leq \hat{x}_{\ell,1}$—that is, those who lie in the mixing region both before and after the shock—we have $z^*_t(x,0) = z^*_t(x,1)$. The z-index of the matched team therefore remains unchanged under our sequential construction. Given positive assortative matching between z_N and x_{N+1}, such a stakeholder continues to work at the firm with ability $x_{N+1}[i]$.

For any $x_\ell > \hat{x}_{\ell,1}$, since profit-driven (purpose-driven) stakeholders must be matched with a team of higher (lower) z-index after the shock, we have $z^*_1(x,0) > z^*_0(x,0)$ and $z^*_1(x,1) < z^*_0(x,1)$. This can be seen from our sequential construction. First of all, since matches before period ℓ depend only on the distribution of stakeholders below type $\ell' < \ell$, the team formation among them remains the same.

We now look at the matching at period ℓ, and such an effect can be seen from Figure 3. More purpose-driven stakeholders of type ℓ must result in a downward shift of the green line in Figure 3. This is because that, in the separation region, $\frac{d\eta^1_s(x_\ell)}{dx} = \frac{\lambda_\ell\, f_\ell(x)}{g^0_{\ell-1}\left(\eta^1_s(x_\ell)\right)}$, and thus

$$\eta^1_s(x_\ell) = (1 - c\sigma)\,\bar{z} - \int_x^{\bar{x}} \frac{\lambda_\ell\, f_\ell(x)}{g^0_{\ell-1}\left(\eta^1_s(x_\ell)\right)}\, dx.$$

Hence, a higher λ_ℓ results in lower $\eta^1_s(x_\ell)$. That is, $(x_\ell, 1)$ will now match with a team with lower productivity. The other way around for $(x_\ell, 0)$.

So, an increase in λ_ℓ must increase the z-ranking of purpose-driven (profit-driven) stakeholders after the matches in period ℓ, which is given by $x_\theta\, \eta^\theta_s(x_\ell)$. Thus, the purpose-driven (profit-driven) stakeholders of type ℓ end up in a less (more) productive team after period ℓ. Given that PAM between z and x must hold for any period τ onward, it thus means that $z^*_0(x_\ell, 1) > z^*_1(x_\ell, 1)$ and $z^*_0(x_\ell, 0) < z^*_1(x_\ell, 0)$.

The z-ranking of a treated stakeholder, who moves from $\theta = 0$ to $\theta = 1$, must drop by $x(z^*_0(x,0) - z^*_1(x,1)) > 0$, where the positivity follows from $z^*_0(x,0) \geq z^*_0(x,1) > z^*_1(x,1)$. A

lower ranking thus means that he must therefore work at a smaller firm. Lastly, since a purpose-driven stakeholder of type ℓ must work at a firm with index ℓ, this implies that the treated stakeholder moves to a smaller firm with values-driven index $n = \ell$. □

A.9 Proof for Proposition 4

Proof. Profit-driven stakeholders earn additional rent if and only if they stay in a team $n = 0$, which includes all type 1 $(x_1, 0)$ and for the retaliative productive stakeholders (where $x_\ell > \hat{x}_\ell$) for type $\ell \geq 2$, according to the first property of Lemma. Using the fact that for the firm to be indifferent between ℓ and $\ell - 1$, he must be indifferent between hiring purpose-driven stakeholder of type ℓ,for any $\ell \geq 2$, we thus have

$$U_\ell(x, 0) - U_\ell(x, 1) = \Omega(yx, \ell - 1) - \Omega(yx, \ell) = C(\ell - 1) - C(\ell)$$

and hence $p_\ell(x, 0) - p_\ell(x, 1|\ell) = C(\ell - 1) - C(\ell) - \psi(\xi_\ell^*) = \Delta_\ell(x_\ell)$. For $x > \hat{x}_\ell$,according to the first property of Lemma 7, since $(x_\ell, 0)$ will have higher ranking than $(x_\ell, 1)$,then $z^*(x_\ell, 0) > z^*(x_\ell, 1)$ for $x > \hat{x}_\ell$. This thus means that for any $(x_\ell, 1)$, there exists $x'_\ell < x_\ell$ such that $z^*(x'_\ell, 0) = z^*(x_\ell, 1)$ in equilbiurm. Hence, we have

$$\begin{aligned} U_\ell(x_\ell, 0) - U_\ell(x_\ell, 1) &= p_\ell(x, 0) - (p_\ell(x, 1|\ell) - \psi(\xi_\ell^*)) \\ &= \int \{z^*(x, 0) - z^*(x, 1)\}\, dx + U_\ell(\underline{x}_\ell, 0) - U_\ell(\underline{x}_\ell, 1) > C(\ell - 1) - C(\ell), \end{aligned}$$

and thus $p_\ell(x, 0) - p_\ell(x, 1|\ell) > \Delta_\ell(x_\ell)$. □

A.10 Proof of Proposition 6

Proof. Let $x_\ell^\theta(x_{N+1})$ denote the productivity of firm x_{N+1}'s matched stakeholder of type ℓ with preference $\theta \in \{0, 1\}$. We first establish that the share of firms in the mid-range matched with a values-driven stakeholder of type ℓ is strictly larger than λ_ℓ.

Define $I_\ell(x) \equiv \dfrac{\displaystyle\int_{\underline{x}_\ell}^{x_\ell^1(x)} \lambda_\ell \, dF_\ell(\tilde{x})}{\displaystyle\int_{x_L}^{x} dF_{N+1}(\tilde{x})}$ which represents the proportion of firms below x that are matched with a purpose-driven type-ℓ stakeholder. By the market-clearing condition,

$$\int_{x_L}^{x} dF_{N+1}(\tilde{x}) = \int_{\underline{x}_\ell}^{x_\ell^1(x)} \lambda_\ell \, dF_\ell(\tilde{x}) + \int_{\underline{x}_\ell}^{x_\ell^0(x)} (1 - \lambda_\ell)\, dF_\ell(\tilde{x}),$$

so we may write

$$I_\ell(x) = \frac{\lambda_\ell \int_{\underline{x}_\ell}^{x_\ell^1(x)} dF_\ell(\tilde{x})}{\int_{\underline{x}_\ell}^{x_\ell^1(x)} \lambda_\ell \, dF_\ell(\tilde{x}) + \int_{\underline{x}_\ell}^{x_\ell^0(x)} (1-\lambda_\ell) \, dF_\ell(\tilde{x})} \;\geq\; \frac{\lambda_\ell \int_{\underline{x}_\ell}^{x_\ell^1(x)} dF_\ell(\tilde{x})}{\int_{\underline{x}_\ell}^{x_\ell^1(x)} dF_\ell(\tilde{x})} = \lambda_\ell \, .$$

The inequality uses the fact that when a firm is indifferent between two stakeholders with different preferences, the purpose-driven stakeholder must be weakly more productive, i.e., $x_\ell^1(x_{N+1}) \geq x_\ell^0(x_{N+1})$, which implies $x_\ell^0(x) \leq x_\ell^1(x)$ and hence the denominator on the left is weakly larger. Equality holds if and only if $x_\ell^1(x) = x_\ell^0(x)$, which occurs only when the stakeholder lies in the mixing region (i.e., $x_\ell < \hat{x}_\ell$).

Part (i). Given that $\hat{x}_{\ell,1} < \hat{x}_{\ell,0}$, consider a firm x_{N+1} matched with a stakeholder satisfying $x_\ell < \hat{x}_{\ell,1}$. Such a stakeholder lies in the mixing region both before and after the shock. Hence, $I_{\ell,1}(x) - I_{\ell,0}(x)$, representing the change in the measure of firms affected by stakeholder $(x_\ell, 1)$, satisfies

$$I_{\ell,1}(x) - I_{\ell,0}(x) = \lambda_{\ell,1} - \lambda_{\ell,0} = \delta.$$

Part (ii). For firms that were in the mixing region before the shock but lie in the separation region after the shock (i.e., $x_{N+1} \in [x_{N+1}^*(\hat{x}_{\ell,1}),\, x_{N+1}^*(\hat{x}_{\ell,0})]$), we have

$$I_{\ell,1}(x) - I_{\ell,0}(x) > \lambda_{\ell,1} - \lambda_{\ell,0} = \delta,$$

where we use the fact that $I_{\ell,0}(x) = \lambda_{\ell,0}$, since $x_{N+1} < x_{N+1}^*(\hat{x}_{\ell,0})$ implies that these firms were in the mixing region before the shock.

Part (iii). The total change in share over the top region $[x_{N+1}^*(\hat{x}_{\ell,0}),\, \bar{x}_{N+1}]$ is

$$\begin{aligned}
&\big\{I_{\ell,1}(\bar{x}_{N+1}) - I_{\ell,1}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big)\big\} - \big\{I_{\ell,0}(\bar{x}_{N+1}) - I_{\ell,0}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big)\big\} \\
&\quad = \delta + \big\{I_{\ell,0}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big) - I_{\ell,1}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big)\big\} \\
&\quad < \delta + \lambda_{\ell,0} - \lambda_{\ell,1} < \delta,
\end{aligned}$$

where the first equality uses the fact that $I_{\ell,t}(\bar{x}_{N+1}) = \lambda_{\ell,t}$ for each t. The second inequality uses $I_{\ell,0}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big) = \lambda_{\ell,0}$ together with $I_{\ell,1}\big(x_{N+1}^*(\hat{x}_{\ell,0})\big) > \lambda_{\ell,1}$, the latter holding because $x_{N+1}^*(\hat{x}_{\ell,0}) > x_{N+1}^*(\hat{x}_{\ell,1})$. Hence, the impact on firms at the top is strictly less than δ.

The impact can equivalently be defined from the viewpoint of stakeholders. Let

$$\tilde{I}_\ell(x) \equiv I_{\ell,1}\big(x_{N+1,1}^*(x)\big) - I_{\ell,0}\big(x_{N+1,0}^*(x)\big).$$

Since $x^*_{N+1}(x)$ is monotone in x, the same conclusions apply to the stakeholder-side formulation. □

A.11 Proof of Proposition 8

Proof. According to Proposition 5, the allocation remains the same for stakeholders at the bottom. Hence, their utility remains the same, according to Equation 11.

For stakeholders at the top, consider first type ℓ. As shown in Proposition 5, the z-index $z^*(x_\ell, \theta)$ must decrease (increase) for purpose-driven (profit-driven) stakeholders because of increased competition on the same side of the market. Hence, according to Equation 11, $U_\ell(x_\ell, 1)$ decreases, whereas $U_\ell(x_\ell, 0)$ increases at the top.

The opposite effect arises for stakeholders of any other type $\ell' \neq \ell$. Due to complementarity, purpose-driven stakeholders of type $\ell' \neq \ell$ benefit from the presence of more productive purpose-driven stakeholders of type ℓ, which results in a higher $z^*(x_{\ell'}, 1)$ and thus a higher $U_{\ell'}(x_{\ell'}, 1)$. Conversely, profit-driven stakeholders of type ℓ' are worse off.

To see this, consider stakeholders of type $\ell' < \ell$, whose matches may change from period ℓ onward. Suppose such a stakeholder is in team $z_{\ell-1}(y_{\ell-1}, n_{\ell-1})$. Given that $(x_\ell, 1)$ is now matched with a less productive team, it means that, fixing the productivity, the team can instead match with a more productive stakeholder of type ℓ in period ℓ, which increases post-match team productivity, as $z_{\ell-1}x^*_\ell(z_{\ell-1})$ must increase. Conversely, teams with $n_{\ell-1} = 0$ must be matched with less productive profit-driven stakeholders of type ℓ, since there are now fewer of them. Hence, the distribution of productivity for teams with index $n \geq 1$ ($n = 0$) improves (deteriorates) in the first-order stochastic dominance sense.

This also implies that, for any stakeholder of type $\ell' > \ell$, stakeholder $(x_{\ell'}, 1)$ will be able to match with a better team, since such stakeholders face an improved distribution of teams with index $n \geq 1$. By contrast, matches worsen for stakeholders of type $(x_{\ell'}, 0)$.

□

A.12 Simulation Details

This appendix documents the simulation underlying Figure 8; a replication package accompanies the paper.

Data-generating process. We draw N worker–bank–firm triples from the calibrated distributions of Section 6.3.1 and solve for the $t = 0$ and $t = 1$ equilibrium allocations using the model's own assignment routines, raising the green-bank share from 7% to 15% between periods. All matches, retention indicators, purpose-index counts n, and structural harm levels

are outputs of this equilibrium; no frictions, matching noise, or shocks are added to the assignment. A firm is *treated* if its $t = 0$ bank turns green and *control* if its $t = 0$ bank stays brown; firms already fully green at $t = 0$ are excluded. *Compliance* (take-up) is the indicator that a firm's purpose index rises, $\Delta n \geq 1$; *retention* is the indicator that a firm keeps its $t = 0$ bank. The partial, size-varying take-up used in the text is a property of this equilibrium, not an imposed disturbance: engagers at the bottom have take-up one, exiters in the middle are randomly rematched to a green stakeholder at approximately the green-bank share, and the assortatively matched largest firms have take-up near zero.

Observed harm. Structural harm e is perturbed to form observed harm in each of T pre- and T post-commitment years. For year t, $\tilde{e}_{ft} = e_f \cdot \exp(\varepsilon_{ft}) + \eta_{ft}$, with $\varepsilon_{ft} \sim \mathcal{N}(0, \tau_e^2)$ a multiplicative idiosyncratic shock and $\eta_{ft} \sim \mathcal{N}\big(0, (\tau_m \,\mathrm{sd}(e_0))^2\big)$ additive classical measurement error. Observed harm in each period is the within-period average over its T years, so the transitory components' variance falls at rate $1/T$. The figure uses $\tau_e = 0.10$, $\tau_m = 0.15$, and $T = 5$; the recovery is robust to $\tau_e \leq 0.5$ and $\tau_m \leq 1$. The first stage (the change in a firm's purpose index) is structural and is not perturbed.

Estimators. Firms are grouped into size bins by x_3. Within a bin, the reduced form is the treated-minus-control difference in the pre-to-post change in observed harm, and the first stage is the corresponding difference in take-up. The *differenced* Wald is their ratio; the *treated-only* Wald divides the treated firms' harm change by their take-up rate. Ratios are computed from firm-level pools aggregated across simulation seeds to stabilize the denominator, and bins in which take-up falls below a fixed threshold are treated as unidentified and shaded. Reported point estimates average over independent seeds; the figure uses $N = 40{,}000$ and 14 seeds.

Replication. The estimator, noise layers, and figure are produced by the accompanying `procedure.py` and `make_recovery_figure.py`, which call the model's assignment code without modification.