THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Michele BISCEGLIA

Le 26 juin 2025

Essais en Théorie Microéconomique Appliquée

École doctorale : **TSE Toulouse Sciences Economiques**
Spécialité : **Sciences Economiques - Toulouse**
Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur Jean TIROLE et M. Patrick REY*

Composition du jury
Rapporteur : M. Julian WRIGHT
Rapporteure : Mme Lucy WHITE
Directeur de thèse : M. Jean TIROLE
Co-directeur de thèse : M. Patrick REY

UNIVERSITÉ TOULOUSE CAPITOLE

Essays in Applied Microeconomic Theory

Michele Bisceglia

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
Toulouse School of Economics, Université Toulouse Capitole

Advisors:
Prof. Jean Tirole (Supervisor)
Prof. Patrick Rey (Co-Supervisor)

Toulouse, June 2025

Jury for the Defense of the Thesis

President:	Prof. Jean Tirole (TSE)
Examiner:	Prof. Patrick Rey (TSE)
External Examiner:	Prof. Lucy White (BU Questrom)
External Examiner:	Prof. Julian Wright (NUS)

"Economics is a science of thinking in terms of models joined to the art of choosing models which are relevant to the contemporary world."

— John Maynard Keynes

Abstract

This thesis contains three essays in applied microeconomic theory. It develops frameworks to advance the understanding of collusion in labor markets, digital platforms' strategies, and socially responsible investments, and derives policy implications.

The first chapter develops a theory of collusion in the presence of labor market power. In an oligopoly-oligopsony setting, a firm needs to increase its wage offers to recruit more workers and expand production, which dampens incentives to deviate from a collusive outcome. No-poaching and non-compete agreements, preventing a firm from hiring its rivals' workers, act as facilitating practices. As a result, labor market power increases firms' ability to collude, and collusion harms consumers and workers, underlining the need for antitrust authorities to monitor collusive behavior also in labor markets. However, if only wage collusion is monitored, or is prevented by enforcing a minimum wage, firms fiercely collude on prices, leaving consumers worse off than under unconstrained collusion.

The second chapter examines whether users receive their fair contribution to a digital ecosystem. The frequent accusations of self-preferencing and excessive platform fees leveled at dominant gatekeepers raise the issue of the standard these platforms should be held to. The important role played by two zero lower bounds on the pricing of core and complementary services in the setting of privately and socially optimal platform fees warrants the concerns about equity for business users. A simple "Pigouvian rule" for regulating access conditions ensures that business users appropriate their contribution to the ecosystem, promoting the right level of innovation; it does so by pricing the unpriced positive externality (ancillary benefit) enjoyed by a third-party seller that receives access to the consumer.

The third chapter analyzes the efficiency of socially responsible investing as a market-based mechanism to control firms' externalities. When responsible and profit-motivated investors interact, the former tend to concentrate on a subset of firms in the economy, while excluding others. This concentration of responsible capital can mitigate free-riding and coordination issues in the adoption of green technologies, but it can also create product market power and crowd out the green investments of excluded firms. If these unintended consequences dominate, aggregate green investments and welfare are larger in the absence of responsible investing. In equilibrium, responsible capital concentrates most when such concentration is least desirable.

Acknowledgements

Isaac Newton famously wrote, "If I have seen further, it is by standing on the shoulders of giants." When I began studying Industrial Organization some years ago, I could have never imagined that true giants of the field, Jean Tirole and Patrick Rey, would be willing to supervise the work of someone *coming from nowhere*. They have always been so generous with their time and ideas that I often felt not only lifted by giants but also gently taken by the hand; so, if I have not seen further, it is entirely my fault. Their example of integrity and humility is even more invaluable.

Speaking of giants, I am deeply grateful to Mike Whinston for hosting me during a fantastic and productive semester at MIT.

I am indebted to my coauthors, especially Ale Bonatti, Fiona Scott Morton, Jorge Padilla, and the great Piccolo *bros* Salvatore and Alessio, from whom I have learned, and continue to learn, so much.

The PhD is a long journey, and it is a joyful one with the right companions. I feel fortunate to have shared this path with Enrico, Luca, and Alessio, and to have crossed paths with other amazing people along the way, especially Luisa.

I thank my mum, dad, sister, and grandpa for their unwavering support and presence, and my lifelong hometown friends, who still make me feel like a carefree high-schooler whenever we are together.

Finally, I thank God for a life filled with health, opportunity, and the wonderful people mentioned above – and many more.

Contents

Chapter I

Labor Market Power and Collusive Behavior

1 Introduction

Across many industries, large firms enjoy significant and increasing degrees of oligopsony power in labor markets.[1] This raises concerns about anticompetitive labor market practices, particularly collusive behavior.[2] In July 2021, a White House executive order encouraged "the FTC and DOJ to strengthen antitrust guidance to prevent employers from collaborating to suppress wages".[3] Accordingly, the DOJ has recently brought several criminal cases against naked wage-fixing agreements, the sharing of relevant wage information, and no-poaching and non-compete agreements.[4] In 2023, the FTC has signed a new agreement with the DOL to bolster efforts to protect workers by promoting competitive labor markets, identifying collusive behavior as the first area of mutual interest for the two agencies.[5]

Concerns over collusive behavior in labor markets add to those over collusion in product markets, and the two may often coexist. Indeed, in many industries, due to skill specialization or the local nature of labor and product markets, large firms that recruit from the same labor markets also operate in the same product markets, and enjoy market power in both.[6] Therefore, they can collude both to suppress wages and increase prices. For instance, the chicken-processing companies involved in one of the major wage collusion cases investigated to date have also been sued for price-fixing, with the two cases having factual overlap.[7] Similar concerns have recently arisen in the health care industry.[8]

This paper aims to develop *a theory of collusion in the presence of labor market power*, guiding antitrust enforcement and regulatory interventions. How does oligopsony power affect collusive strategies and their sustainability? Are antitrust authorities' inter-

[1]Berger et al. (2022) develop a general equilibrium oligopsony model of the labor market and find that it is quantitatively consistent with documented empirical regularities suggestive of oligopsony. Several other studies find low estimates for firm-level labor supply elasticities across many different sectors (see, e.g., Manning, 2021).

[2]For empirical evidence of employer collusion, see Sharma (2024).

[3]See the Fact Sheet at `https://www.presidency.ucsb.edu/documents`.

[4]See `https://www.antitrustalert.com/tag/wage-fixing/` for an overview of some recent cases.

[5]See the Press Release at `https://www.ftc.gov/news-events/news/press-releases/2023/09/`. Similarly, in the EU, a recent policy statement from the European Commission treats wage-fixing and no-poaching agreements as *restriction of competition by object*: see `https://competition-policy.ec.europa.eu/document/`.

[6]Using plant-level data, Tortarolo and Zarate (2018) find significant price mark-ups and wage markdowns.

[7]Because plants cluster in the areas in which chickens are raised by farmers, there are multiple plants in relatively small areas in which they compete for workers. In both wage- and price-fixing cases, plaintiffs argue that the defendants exchanged information through various intermediaries, including a company called Agri Stats. The top three chicken industry firms named in the DOJ's Agri Stats lawsuit (Tyson, Pilgrim's Pride, and the recently-merged Sanderson-Wayne Farms) have already been subject to at least $698 million in settlements: see `https://accountable.us/report`.

[8]Revelations about a data analytics firm's role in determining medical payments have raised concerns about potential price fixing in health care, prompting calls for a federal investigation. See `https://www.nytimes.com/2024/05/01/us/multiplan-health-insurance-price-fixing.html`.

ventions to monitor collusion in labor markets aligned with consumer protection, or do they necessarily rely on a *broad mandate* that includes worker protection? How do labor market regulations, such as minimum wage and pay-equity provisions, and the antitrust treatment of no-poaching and non-compete agreements impact collusion and, ultimately, consumer and worker welfare?

To address these questions, Section 2 introduces labor market power, deriving from workers' idiosyncratic preferences for different employers,[9] in a barebone model of collusion in oligopoly: the Bertrand supergame – i.e., the infinitely repeated Bertrand oligopoly game with homogeneous products. At each period, firms simultaneously make wage offers to workers and set product prices. These choices determine the labor force and the consumers' demand for each firm. Firms also employ other variable factors (e.g., flexible capital inputs, such as materials), traded in competitive markets, to produce the demanded output;[10] the production function exhibits constant returns to scale. Within this framework, Section 3 characterizes the cartel outcome – i.e., the industrywide-profit-maximizing wage and price levels, which would arise if cartels were legal and contractually enforceable – and the collusion outcome – i.e., the most profitable (stationary and symmetric) subgame-perfect equilibrium of the supergame. It presents two main sets of results.

First, *labor market power facilitates collusion*: The critical discount factor above which the cartel outcome is sustainable under collusion is lower than in perfectly competitive labor markets. This is because a deviating firm cannot capture the entire cartel profit in the presence of oligopsony power. While it can still attract all its rivals' consumers by slightly undercutting the monopoly price, expanding production to satisfy their demand requires hiring additional workers. Since the (residual) labor supply is upward-sloping, this entails paying higher wages. It is, therefore, impossible for a deviating firm to capture the whole cartel revenue without increasing its average production costs, which weakens deviation incentives, ultimately fostering collusion.

Second, the analysis reveals the *interplay between collusion in labor and product markets*: The best collusive scheme involves both sub-competitive wages and supra-competitive prices. This outcome equalizes the ratio of the marginal profit from collusion to the marginal profit from deviation across labor and product markets: In simple terms, firms exploit *multimarket contact* to allocate their collusive power across both markets, so they simultaneously raise prices and lower wage offers.

These results have relevant policy implications, which are examined in Section 4. In

[9]See, e.g., Azar et al. (2022), who conclude that job differentiation gives employers market power, allowing them to pay workers less than their marginal productivity.

[10]As in standard Bertrand models, a firm is always committed to satisfying all consumers' demand at its posted price. In this model, it can do so for any labor force at its disposal by adjusting its endowment of flexible capital. The *no rationing* assumption ensures the existence of a zero-profit static Nash Equilibrium, which implies that restricting attention to subgame-perfect equilibria in grim-trigger strategies is without loss of generality, but is otherwise inessential to the results.

the presence of labor market power, collusion is easier to sustain and harms consumers and workers alike, so competition authorities need to intensify their enforcement activities. Moreover, they must actively monitor collusion also in labor markets, even if they follow a pure consumer surplus standard. The reason is that, if only price collusion is monitored, firms can still collude in the labor market:[11] By coordinating to reduce their wage offers, they end up hiring fewer workers, making it individually optimal to produce less than under competitive wage-setting behavior. As a result, even if firms' price-setting behavior is non-cooperative due to the monitoring of price collusion, equilibrium output is reduced, to the detriment of consumers. However, shifting antitrust authorities' (limited) budget devoted to monitoring collusive behavior from product to labor markets would lead to the worst outcome from consumers' standpoint: Pure price collusion yields even higher prices than a multimarket collusive arrangement. The reason is that when wage collusion is monitored, competitive behavior in the labor market leads to higher wages, which in turn raises production costs, making higher prices incentive-compatible.[12]

When collusive behavior cannot be monitored, the introduction of a binding minimum wage, or an increase thereof, may similarly lead to an increase in equilibrium employment and prices. In the presence of oligopsony power, such a pass-through of an increased regulatory minimum wage to consumers would not occur under competitive behavior: A binding minimum wage deprives oligopsonists of incentives to hire fewer workers in order to pay lower wages; under competitive behavior, this would give them incentives to expand production, i.e., to set lower prices. However, this effect may be outweighed by the strengthening of price collusion, which is triggered by the inability of colluding in wages. Thus, consumer harm from an increase in the minimum wage, often found in the empirical literature since Card and Krueger (1994), is not inconsistent with the presence of labor market power; instead, it is suggestive of collusive behavior.

The result that oligopsony power fosters collusion also implies that firms capture most of the benefits from their labor market power: The pass-through to consumer prices of wage mark-downs due to firms' market power *vis-à-vis* workers is more limited than under competitive behavior. As a result, policies aimed at limiting the extent of firms' labor market power by protecting workers' bargaining power – e.g., strengthening trade unions

[11]Collusive behavior often takes place through price- and wage-fixing agreements, requiring communication (though, being illegal, hence not enforceable in court, these agreements must be incentive-compatible) – see, e.g., Harrington (2006). If antitrust authorities prevent firms from agreeing on prices (e.g., by monitoring communication among pricing managers), firms can still engage in wage-fixing agreements (e.g., through communication among HR directors); in these cases, provided they offer the agreed wage, firms can set any price level without triggering a punishment, which results in competitive (i.e., static-Nash) price-setting behavior taking wages set at the collusive level. The same *semi-collusive* outcome arises if antitrust authorities can only infer whether the prevailing price is supra-competitive given the observed firms' labor force.

[12]Monitoring of wage collusion is thus complementary to, rather than a substitute for, monitoring of price collusion: The welfare gain to consumers from the monitoring of price collusion is higher when wage collusion is monitored.

– benefit workers mostly at the expense of firms rather than consumers.

Labor market power, deriving from the local nature of many (especially low-skilled) labor markets (e.g., Marinescu and Rathelot, 2018), also has implications for standard policy measures aimed at enhancing the competitiveness of product markets. First, promoting product market globalization (e.g., through free trade agreements), whereby firms in different geographic markets compete for consumers in a global product market, may fail to lead to competitive prices. The opportunity to serve consumers in other markets does not necessarily strengthen firms' incentives to deviate from a collusive outcome, due to the difficulty to recruit workers from those corresponding labor markets in order to expand production. This finding is consistent with the evidence on mark-ups in De Loecker and Eeckhout (2018) and underscores the irreplaceable role of antitrust authorities' monitoring of collusion.

Second, in local labor markets it is often the case that the same workers can be hired by firms selling different products. In these circumstances, oligopsony power introduces cross-market externalities through the labor market: Higher wage offers by any firm also harm firms selling in independent product markets by increasing the wage they need to offer to recruit a given number of workers. As a result, firms selling in independent product markets can collude together (on wages and price levels) to internalize these externalities, and *conglomerate mergers* – i.e., mergers across firms in independent product markets – facilitate such *cross-market collusion*, thereby leading to higher prices in all markets, consistent with empirical evidence (e.g., Ciliberto and Williams, 2014). These findings advocate for a strict conglomerate merger policy when the merging parties recruit workers in the same labor markets.

Finally, extending the model to long-term employment contracts allows for the identification of labor-market-specific facilitating devices. Considering an overlapping generations model with identical cohorts of workers, Section 5 shows that no-poaching agreements (NPA), prohibiting firms from making offers to each others' current employees, can be used as facilitating practices. The impossibility of making poaching offers makes the (residual) labor supply steeper for a deviating firm that wants to expand its labor force, discouraging deviations from a collusive equilibrium. The reason is that, for any wage offer above the (candidate) equilibrium one, a deviating firm can expand its labor force to a greater extent when it can make the same offer also to its rivals' workers, as some of them would accept it. Thus, signing binding NPA enables firms to sustain more collusive outcomes. This result provides an anticompetitive rationale for the widespread use of NPA also in low-skilled labor markets (e.g., Krueger and Ashenfelter, 2022) and justifies their *per se illegality* antitrust status in these markets. Non-compete agreements (NCA), whereby a worker commits with the current employer to not work for a competitor in the future, and which are also widely used in low-skilled labor markets (Starr et al., 2021), can similarly dampen deviation incentives.[13]

Through a different mechanism, also pay-equity regulations, which require firms not to wage-discriminate among workers ("equal pay for equal work"), facilitate collusion: For a firm contemplating an increase in its wage offers to recruit more newcomers and (absent NPA or NCA) rivals' incumbent workers, pay-equity regulations imply the obligation to correspondingly increase the remuneration of its current employees, which reduces the profitability of such deviation. This result may explain why the introduction of pay-transparency rules, which help enforce pay-equity regulations by revealing eventual pay disparities among coworkers performing similar work within a firm, often leads to a reduction in average wages (Cullen, 2024).

Section 6 concludes. All proofs are in Appendix A. Appendix B contains additional material.

Related literature. Starting from Friedman (1971), an extensive literature has analyzed collusion in oligopoly supergames. Previous studies have examined how the sustainability of collusion depends on firms' asymmetries or product differentiation, capacity constraints, market transparency, business cycles or demand fluctuations, *inter alia*, and how firms can employ facilitating practices such as joint venture agreements and resale price maintenance: see Ivaldi et al. (2003) for an excellent overview. All these models, however, assume that firms are price takers in the input markets and produce at a constant marginal cost. Departing from these assumptions, this paper shows that labor market power facilitates collusion.

The mechanism underlying this facilitating effect is that a deviating firm faces increasing marginal costs when expanding production because of the need to raise its wage offers to attract more workers. This result holds whenever each firm faces an upward-sloping labor supply, no matter whether the oligopolists also interact in the same labor market. Diseconomies of scale indeed make it possible to sustain supra-competitive equilibrium prices even absent repeated interactions, as shown by Dastidar (1995) in his analysis of the Bertrand game with convex costs. Unlike in Dastidar (1995) and follow-up work (see Vives, 1999, for an overview), by microfounding the cost function through the modeling of the labor market, this paper also examines how firms that simultaneously interact in the same labor and product markets can collude in both markets.

In the latter respect, this work relates to the literature on multimarket contact. Bernheim and Whinston (1990) were the first to formalize the insight that multimarket contact can facilitate collusion. In their framework with multiple independent product markets, multimarket contact pools the incentive-compatibility constraints for collusion across markets, which can relax binding constraints if markets are asymmetric. This insight

[13] Workers do not expect to receive attractive poaching offers from other employers along any stationary equilibrium path, so they are willing to sign a non-compete clause without asking for compensation.

has been extended to interdependent product markets by Spagnolo (1999) and Choi and Gerlach (2013). Considering vertically related markets,[14] this paper unveils novel mechanisms through which multimarket contact facilitates collusion, driven by the profitability of a joint deviation (in output and input markets). The paper also extends the analysis to multiple product markets, showing that, unlike in Bernheim and Whinston (1990), conglomerate mergers can have anticompetitive multimarket-contact effects even when (product) markets are completely symmetric. This is because such mergers reduce the profitability of a joint deviation across product markets, as expanding production in one market entails higher recruitment costs also in the other market.[15]

Moreover, the present paper is the first to model wage collusion. The long-standing literature on monopsony or oligopsony power in labor markets has abstracted from product market interactions and collusion. Borrowing the terminology from Manning (2021), two main approaches have been taken in this literature to model employers' market power: The "new classical" approach posits employer differentiation deriving from heterogeneity in tastes among workers, and borrows standard industrial organization models of imperfect competition (e.g., Bhaskar and To, 1999); the "modern" approach is instead based on frictions in the labor market deriving from a search-and-matching process (e.g., Burdett and Mortensen, 1998). This model follows the first approach, but its insights are robust to the case where labor market power derives from search frictions (see Appendix B.1). The few papers examining non-cooperative (Tong and Ornaghi, 2021) or cooperative (Gonzaga et al., 2014) oligopoly-oligopsony models have followed the same approach, but have relied on static analyses, thereby neglecting collusive behavior.

Finally, some recent papers have examined the trade-off between reduced worker mobility and increased employee training entailed by no-poaching and non-compete agreements (e.g., Martins and Thomas, 2023, and Shi, 2023, respectively). Mukherjee and Vasconcelos (2012) have similarly considered these agreements as alternative means for employers to avoid engaging in wage wars for hiring each others' "star workers". In all these models, the rationale for these clauses and their competitive and welfare effects only apply to high-skilled labor markets, where workers' training or individual performance are relevant. By showing how these agreements can help firms to sustain more collusive outcomes, this paper provides a novel anticompetitive rationale for their use also in low-skilled labor markets, which is robust to, but does not rely on, firms also interacting in the same product market. The mechanism is similar in spirit to Aghion and Bolton (1987) and Rasmusen et al. (1991), where contractual restrictions to mobility discourage

[14] Previous works on collusion in vertically related markets (e.g., Nocke and White, 2007, Piccolo and Mikl'os-Thal, 2012, and Normann et al., 2015) have focused on models of vertical supply chains.

[15] The industrial organization literature on conglomerate mergers – i.e., mergers of firms selling in independent product markets – is rather sparse. In a recent paper, Chen and Rey (2023) have examined the welfare effects of these mergers in the presence of heterogeneous "consumption synergies" deriving from bundling of independent products. In this paper, independent product markets are instead linked through the labor market.

potential competition (i.e., entry) rather than the competitive behavior (i.e., deviations) of incumbents.

2 Model

This section describes the baseline model and examines the benchmark scenarios where firms have no (individual) labor market power.

2.1 Set-up

Consider a product market with n symmetric firms. To produce, they need to employ workers, whom they hire from the same labor market, and other (variable) production factors.

Product market. Firms produce perfect substitutes and compete *à la* Bertrand. That is, each firm $i \in \mathcal{N} \equiv \{1, ..., n\}$ posts a price p_i at which it is committed to serving all the resulting demand (*no rationing of consumers*). Perfect substitutability implies that consumers only buy from the lowest-priced firm(s): Consumers' demand is a downward-sloping function $Q(p)$ of the lowest available price $p \equiv \min_{i \in \mathcal{N}} p_i$; if several firms charge p, they equally split $Q(p)$. Formally, firm i sells

$$q_i^d \equiv \mathbb{1}[p_i = p] \frac{Q(p)}{\#\{i : p_i = p\}}, \tag{1}$$

where $\#\{i : p_i = p\}$ is the number of firms that charge the lowest price in the market.

Production function. Each firm i produces with a constant returns to scale (hereafter, CRS) production function: Its output is given by $q_i^s \equiv F(\ell_i, k_i)$, where $F(\cdot)$ is (positively) homogeneous of degree one, ℓ_i is the labor force at its disposal, and k_i represents the amount of another variable factor, which will be referred to as *flexible capital*, it employs for production. Flexible capital is traded in a competitive market at rate r.

Labor market. In the labor market, there is a measure J of ex-ante identical workers, indexed by $j \in [0, J]$, where J is (finite but) sufficiently large, so that workers are always in *excess supply* for firms $i \in \mathcal{N}$. Each worker j inelastically supplies one unit of labor when hired by a firm i in exchange for a wage $w_{i,j}$. All workers have the same outside option w_0, whose value is common knowledge, say the *competitive wage* they would earn if hired outside of the considered industry. Worker j's utility from accepting firm i's offer is given by $w_{i,j} + \xi_{i,j}$, where $\xi_{i,j}$ are i.i.d. draws from a continuous c.d.f. $\Xi(\cdot)$ with bounded support $[\underline{\xi}, \overline{\xi}]$, and represent j's non-monetary extra-utility from working in i

rather than in the outside competitive sector. The realizations of $\xi_{i,j}$ are worker j's private information, and workers are *anonymous* from firms' standpoint.

Firms can (costlessly) make personalized wage offers to as many workers as they want; each worker observes the available offers and decides which one to accept (if any). Each firm i is committed to paying $w_{i,j}$ to workers j who have accepted its offer no matter how much it produces – i.e., no new hires or dismissals are possible after consumers' demand realizes (*short-term rigid labor force*).[16] Therefore, firm i's labor force is[17]

$$\ell_i \equiv \int_{j\in[0,J]} \mathbb{1}\left[w_{i,j} + \xi_{i,j} \geq \max\left\{w_0, \max_{i'\in\mathcal{N}\setminus\{i\}} (w_{i',j} + \xi_{i',j})\right\}\right] \mathrm{d}j, \tag{2}$$

with the convention that $w_{i,j} = -\infty$ for any worker j to whom firm i does not make an offer.

Timing and solution concept. Each firm i sets its product price p_i and makes offers $\{w_{i,j}\}_{j\in[0,J]}$ to workers. These choices are simultaneous and determine the amount of output each firm must produce, $q_i = q_i^d$ given in Eqn. (1), and each firm's labor force, ℓ_i given in Eqn. (2), respectively. Under the no-rationing and short-term rigid labor force assumptions, (q_i, ℓ_i) pin down the amount of flexible capital $K(\ell_i, q_i) \equiv F^{-1,k}(\ell_i, q_i)$ (with $F^{-1,k}(\cdot)$ denoting the inverse of the production function with respect to k) firm i needs to employ.

This paper considers the infinitely repeated version of this stage game, with perfect monitoring, long-lived firms and short-lived workers.[18] That is, time is discrete and indexed by $t = 0, 1, ...$, identical cohorts of one-period-lived workers enter the labor market at each t, and firms discount profits at a common rate $\delta \in (0, 1)$. The solution concept is *stationary symmetric Subgame Perfect Nash Equilibrium in pure strategies*, hereafter referred to as SPNE, with *no wage-discrimination* – i.e., along the equilibrium path, all firms charge the same price ($p_i \equiv p$) and offer the same wage to all workers over time ($w_{i,j} \equiv w$); no restrictions are imposed off-equilibrium path.

Discussion and assumption. This paper aims to understand the impact of labor market power on collusive behavior. As price- and wage-fixing cartels are *per se illegal*, cartel provisions are not enforceable in court, and colluding firms are constrained to choose self-sustainable price and wage levels. Throughout the paper, *cartel outcomes* refer to price and wage levels that would prevail if cartels were enforceable and *collusive*

[16] No-rationing and short-term rigid labor force are assumed even if a firm makes negative profits (*no exit*).

[17] Ξ being continuous, workers' tie-breaking rule is immaterial to the analysis.

[18] As flexible capital is not a strategic choice, the stage game is the simultaneous-choice wage- and price-setting game. As firms can only post uniform prices, provided they sell a non-durable good, whether consumers are short- or long-lived is immaterial to the analysis.

outcomes to those arising in the most profitable SPNE of the supergame.[19]

The first building block of the framework is a barebone model of collusion in oligopoly, namely, the Bertrand supergame. This model is particularly tractable because Nash reversion yields a discounted profit of zero, which, by the results in Abreu (1988), implies that restricting attention to grim-trigger strategies (Friedman, 1971) is without loss of generality. This conclusion will still be valid in my model, which significantly simplifies the analysis. Formally, a stationary action profile $(p_i, \{w_{i,j}\}_{j\in[0,J]})_{i\in\mathcal{N}}$ is a SPNE in grim-trigger strategies of the supergame defined above if, at any given t, firms play the specified actions if and only if no firm has played differently at any $t' < t$; else, a Nash Equilibrium of the stage game (hereafter, *static NE*) is played.

The Bertrand game assumes that firms are committed to satisfying all consumers' demand at their posted prices, which is plausible in cases where there are high costs of turning consumers away (see Vives, 1999, for a discussion). For this to be feasible also off-path, firms must be able to adjust their production capacity after demand is realized. In this model, firms produce using labor and flexible capital and,[20] in line with the literature on multimarket contact (Bernheim and Whinston, 1990), I consider simultaneous firms' choices in labor and product markets to rule out commitment effects. Then, flexible capital must be eventually adjusted after consumers express their demand to ensure no-rationing. The assumption that capital inputs, such as materials, lack adjustment costs and monopsony power aligns with empirical evidence (Yeh et al., 2022).

To avoid the results of the paper be driven by the no-rationing assumption, the following will be assumed throughout:

(A) *Serving all consumers' demand would be optimal for a firm deviating from the cartel outcome, even if it could not adjust its wage offers and had the option to target fewer consumers.*

Assumption (A) is formalized in Appendix A, together with other technical assumptions guaranteeing that all problems considered in the analysis are well-behaved.

Following the "new-classical approach" to labor market power (Manning, 2021), firms enjoy labor market power because workers are in excess supply and have idiosyncratic preferences for different employers. The considered model, under a Type I Extreme Value specification for $\Xi(\cdot)$, is equivalent to a logit model, which has often been estimated in

[19]The distinction between *explicit* and *tacit* collusion is immaterial to the analysis, except in Section 4.1, where I examine antitrust authorities' monitoring of collusion, which is only possible for explicit collusion.

[20]If labor were the only (variable) production factor, wage choices alone would determine firms' capacity (at least in the short-run), and output would always be sold at the market-clearing price (Kreps and Scheinkman, 1983). Then, collusion in the labor or product market would be equivalent. Constant returns to scale (with respect to $\{\ell_i, k_i\}$), while being in line with empirical evidence in many manufacturing industries (e.g., Berger et al., 2022), are mainly assumed for consistency with the standard Bertrand supergame; same as for firm symmetry.

empirical works (e.g., Card et al., 2018, Tortarolo and Zarate, 2018).[21] The assumptions that workers are short-lived and can only be employed by firms selling in the same product market will be later relaxed. The robustness of the main results with respect to the other modeling choices is discussed in Appendix B.2.

2.2 Benchmarks: No individual-firm labor market power

Before proceeding with the analysis, this section shows that the considered framework is a straightforward extension of the analysis of collusion in the Bertrand oligopoly supergame whenever firms $i \in \mathcal{N}$ are identical from workers' viewpoint and so, at least individually, have no labor market power.

Perfectly competitive labor market. Suppose that the labor market is perfectly competitive: $\xi_{i,j} \equiv 0$ for all i and j, so that *(i)* firms need to offer at least w_0 to hire any worker, and *(ii)* workers being in *excess supply*, firms can hire as many workers as they want at w_0. Therefore, each firm i only chooses how many offers to make, determining its labor force ℓ_i,[22] and firms' cost-minimization problems are independent of each other.

For any anticipated quantity q_i to sell, the optimal *labor demand* $\ell^*(q_i)$ of firm i is obtained by minimizing the production cost $w_0\ell_i + rK(\ell_i, q_i)$. As the production function is CRS and all factors' prices are constant, standard arguments imply that *(i)* the optimal labor demand is such that $\ell^*(q_i) = q_i\ell^*(1)$, and *(ii)* firms can produce any amount of output at a constant marginal, or average, cost: Formally, the minimized average cost function is

$$C_0(q) \equiv \frac{1}{q}[w_0\ell^*(q) + rK(\ell^*(q), q)] = w_0\ell^*(1) + r\,K(\ell^*(1), 1) \equiv c_0 \qquad \forall q > 0. \tag{3}$$

As a result, the analysis unravels as in a standard Bertrand supergame with constant marginal costs (Friedman, 1971): Firms always choose the individually optimal labor demand in equilibrium, and a firm deviating from a candidate equilibrium price p can appropriate the whole industry profit by slightly undercutting p (so to capture all consumers' demand $Q(p)$ instead of its share $1/n$) and optimally expanding its labor demand (from $\ell^*(Q(p)/n)$ to $\ell^*(Q(p))$) in order to produce $Q(p)$ at the same marginal cost c_0.

No individual-firm labor market power. Suppose now that all firms $i \in \mathcal{N}$ are still identical from workers' viewpoint, but workers value differently working in the considered

[21] The considered firms may be able to mark-down the competitive salary w_0 which, as standard in partial equilibrium analysis, is taken as exogenous (this is always the case if, e.g., $\underline{\xi} \geq 0$). However, the results of the paper only rely on each firm $i \in \mathcal{N}$ facing an upward-sloping (residual) labor supply function.

[22] Formally, each firm i offers w_0 to all workers with probability ρ_i such that it ends up hiring ℓ_i workers in equilibrium – e.g., with two firms and workers breaking ties with equal probability across them, $J\rho_i\left(\rho_{i'}\frac{1}{2} + (1-\rho_{i'})\right) = \ell_i$.

industry compared to the outside competitive sector: Formally, for any worker j, $\xi_{i,j} \equiv \xi_j \sim \Xi$ for all i.

Then, the cost-efficient way for a cartel to produce a quantity q is defined as

$$C_0(q) \equiv \frac{1}{q} \min_w \{wL_0(w) + rK(L_0(w), q)\}, \tag{4}$$

where $L_0(w) \equiv J[1 - \Xi(w_0 - w)]$ is the industrywide labor supply when all firms $i \in \mathcal{N}$ offer $w_{i,j} = w$ to all workers $j \in [0, J]$. Slightly abusing notation, denote by $p = c_0$ the solution to $p = C_0(Q(p))$. For any $p > c_0$, a firm deviating from a candidate equilibrium (p, w) can appropriate the whole industry profit. In fact, it can capture all consumers' demand $Q(p)$ by slightly undercutting the candidate equilibrium price p; and it can hire all workers $L_0(w)$ by slightly overcutting the candidate equilibrium wage w.[23]

Results and implications. The previous arguments imply that, in both the considered scenarios, the following results hold:

Proposition 0 (no individual-firm labor market power). *If the labor market is perfectly competitive ($\xi_{i,j} \equiv 0$) or firms have no individual labor market power ($\xi_{i,j} \equiv \xi_j \sim \Xi$), then:*

- *For all $\delta < \delta_0^M \equiv (n-1)/n$, the unique SPNE is the repetition of the static NE, in which firms price at the minimized average cost and make zero profits;*
- *For all $\delta \geq \delta_0^M$, the cartel outcome – i.e., the price and labor demand or wage offers that maximize industry profits – is a SPNE.*

The results of Proposition 0 show that, provided firms have no individual labor market power, the analysis yields the standard results of collusion in the Bertrand oligopoly supergame, whose main features can be summarized as follows:

1. *Competitive static equilibrium*: The static game admits a unique (symmetric) NE, where firms price at the optimized average cost ($p = c_0$) and make zero profits. Starting from any candidate NE with $p > c_0$, where firms make positive profits $\pi = (p - c_0)Q(p)/n$, each firm can profitably deviate by slightly undercutting this price to serve all consumers' demand $Q(p)$; this output is produced at the same average cost c_0 because the deviating firm recruits the extra workers at the same wage.

2. *"Bang-bang" collusion*: For any price $p \in (c_0, p_0^M]$, where $p_0^M \equiv \arg\max_p (p - C_0(Q(p)))Q(p)$ denotes the cartel price, the deviation examined above allows the

[23] This argument also implies that, in any candidate static NE, firms produce at the optimized average cost.

deviating firm to reap the whole industry profit in the deviation period. However, it triggers Nash reversion and so a profit of zero forever after. As a result, no collusion at all is sustainable (i.e., $p = c_0$ for all t) if $\delta < \delta_0^M$, whereas the cartel outcome is sustainable for all $\delta \geq \delta_0^M$.

3. *Pure price collusion*: The cartel outcome can be implemented by colluding firms by only coordinating on setting price p_0^M. Firms do not need to also coordinate their behavior in the labor market: Each firm always acts in its individual best interest in the labor market, to produce its share of the monopoly output in the most cost-efficient way (i.e., at the optimized average cost defined above). Put it differently, once firms have agreed on *how much* to produce (i.e., on the product price p), there is no need for colluding also on *how* to produce (i.e., on how many workers to hire or on wage offers to make).[24]

To sum up, in the absence of (individual-firm) labor market power, antitrust authorities only need to monitor firms' price-setting behavior in industries featuring relatively few competitors to ensure that consumers have access to competitively-priced products, and efficient employment and wage levels. The analysis of Section 3 will show that none of these results holds in the presence of individual-firm labor market power.

3 Equilibrium analysis

When workers derive firm-specific non-monetary utility from working in the considered industry ($\xi_{i,j} \sim^{\text{i.i.d.}} \Xi$ for all i, j), each firm has labor market power: It can offer a lower wage relative to its competitors, and still hire any worker with positive probability. This section contains the equilibrium analysis in this scenario. Section 3.1 characterizes the Nash Equilibria of the static game. Section 3.2 analyzes the supergame, first characterizing the conditions under which the cartel outcome can be sustained as a SPNE, and then deriving the collusive outcome when it cannot.

3.1 Static game

This section characterizes firm behavior in the static game to grasp the first insights on the impact of labor market power and to show that grim-trigger strategies are without loss of generality in the supergame. To simplify the exposition and without loss of insights, I restrict attention to symmetric NEs.

[24] Conversely, coordination only on labor market behavior (i.e., agreeing to restrict the number of offers or to offer the wage prevailing in the cartel outcome) would not suffice to implement the cartel outcome, because firms would have incentives to undercut p_0^M even holding fixed the number of workers they hire in the cartel outcome.

Wage-offer game. The next lemma describes firm behavior in the personalized-offers game *vis-à-vis* workers:

Lemma 1 (wage offers). *For each firm i, making non-discriminatory wage offers – i.e., offering $w_{i,j} \equiv w_i \ \forall j \in [0, J]$ – is a dominant strategy in the static game. Then, if $w_{i'} \equiv w \ \forall i' \in \mathcal{N} \setminus \{i\}$, firm i hires a measure $\ell_i \equiv L(w_i, w)$ of workers, where the function $L(\cdot)$ is such that*

$$\frac{\partial L(w_i, w)}{\partial w_i} > 0 > \frac{\partial L(w_i, w)}{\partial w}, \quad \text{and} \quad \left(\frac{\partial L(w_i, w)}{\partial w_i} + \frac{\partial L(w_i, w)}{\partial w} \right) \bigg|_{w_i = w} > 0.$$

As making offers is costless and workers are anonymous from firms' viewpoint, the minimal-cost way for a firm i to hire any labor force ℓ_i is, irrespective of its rivals' behavior, to offer the same wage to all workers $j \in [0, J]$. Because, in the case of a non-degenerate distribution of offers, the higher ones are accepted with larger probability, differentiating offers across workers (and so, *a fortiori*, not approaching some of the workers) would increase the average wage a firm ends up paying to recruit any given measure ℓ_i of workers, as i would recruit mostly workers to whom it has offered high wages.[25]

As a result, in a static setting, this game of personalized offers is equivalent to a standard wage-competition model with differentiated employers, where each firm posts a wage and is committed to hiring all workers who accept this wage. Each firm then faces a labor supply function obtained from Eqn. (2) for $w_{i,j} \equiv w_i$ for all $j \in [0, J]$, which is upward-sloping in its offer and downward-sloping in the competitors' offers: Labor market power thus introduces oligopsonistic competition for workers.

Diseconomies of scale. How does oligopsonistic competition affect firms' cost structure? Defining firm i's optimized average cost function, for any given rivals' symmetric offers $w_{i'} \equiv w \ \forall i' \in \mathcal{N} \setminus \{i\}$, as

$$C(q; w) \equiv \frac{1}{q} \min_{w_i} \left[w_i L(w_i, w) + r\, K(L(w_i, w), q) \right], \tag{5}$$

the following result holds:

Lemma 2 (diseconomies of scale). *For any symmetric rivals' offer w, firm i's optimized average production cost is increasing in output: $\partial C(q; w)/\partial q > 0$.*

Lemma 2 states the well-known result that a firm operating under a CRS production function and enjoying monopsony power in some input markets faces an increasing average

[25] The logic of this result is the same behind the optimality of uniform pricing for a monopolist facing consumers with unit demand (Riley and Zeckhauser, 1983). As workers are anonymous, from each firm i's viewpoint any set of rivals' wage offers $\{w_{i',j}\}_{i' \in \mathcal{N} \setminus \{i\}, j \in [0,J]}$ just translates into a different distribution of workers' best alternative option, which is immaterial to the optimality of a uniform wage offer.

cost function (see, e.g., Gelles and Mitchell, 1996). In the present setting, where firms recruit from the same labor market, this result applies, holding fixed rivals' wage offers, if firm i behaves as a monopsonist given its upward-sloping residual labor supply function, which must be the case in any static NE.

Equilibrium characterization. For any candidate equilibrium price p, each firm i expects to sell $q_i = Q(p)/n$, and accordingly chooses what wage w_i to offer workers in order to minimize its production cost – i.e., for given rivals' offers w, it produces at the optimized average cost $C(Q(p)/n; w)$. The equilibrium wage offers are thus a fixed point of firms' cost-minimization problems:

Lemma 3 (competitive wage offers). *For any candidate NE price p, there is a unique symmetric equilibrium wage offer $W^*(p)$, and it is decreasing in p.*

As in the benchmark models examined in Section 2.2, in a candidate equilibrium where firms charge a higher price, each anticipates a lower consumer demand and thereby finds it optimal to hire fewer workers; hence, firms optimally reduce their wage offers.

Each firm's profit in the candidate equilibrium with price p can be written as

$$\pi(p) \equiv \left[p - C\left(\frac{Q(p)}{n}; W^*(p)\right)\right]\frac{Q(p)}{n}. \tag{6}$$

A firm's best deviation consists in slightly undercutting the candidate equilibrium price p, attracting all consumers' demand $Q(p)$,[26] and optimally increasing its wage offers to hire more workers, in order to minimize the corresponding production cost, thus obtaining a profit

$$\pi^D(p) \equiv [p - C(Q(p); W^*(p))]\, Q(p). \tag{7}$$

Any price p such that $\pi(p) \geq \max\{\pi^D(p), 0\}$, together with the corresponding wage offer $w = W^*(p)$, is a static NE.

Proposition 1 (static NE). *The static game admits a continuum of NEs: There exist prices $\underline{p}^N < \overline{p}^N$ such that $(p, W^*(p))$ set by all firms is a NE for all $p \in [\underline{p}^N, \overline{p}^N]$; firms' equilibrium profit is zero for $p = \underline{p}^N$ and is strictly increasing in p.*

The diseconomies of scale effect (Lemma 2) implies that if a firm undercuts the candidate equilibrium price p and so serves all the demand $Q(p)$ alone, its (optimized) average production cost increases: $C(Q(p); W^*(p)) > C(Q(p)/n; W^*(p))$. The price $\underline{p}^N = C(Q(\underline{p}^N)/n; W^*(\underline{p}^N))$, at which firms make zero profits, is an equilibrium price because a deviating firm undercutting $\underline{p}^N$ would end up selling $Q(\underline{p}^N)$ below the corresponding average production cost – i.e., $\pi(\underline{p}^N) = 0 > \pi^D(\underline{p}^N)$. Higher prices, such that

[26] This is the case for any candidate NE price below the monopoly price given wage offers $W^*(p)$ (considering static NEs with higher prices would be uninteresting).

$p > C(Q(p)/n; W^*(p))$ and so firms make positive profits, can also be sustained as static NE.[27]

Implications for the analysis of collusion. The results of Proposition 1 have the following implications:

Corollary 1 (static NE). *The static-game equilibrium characterization has the following implications for the supergame:*

- *Perpetual reversion to the zero-profit static NE* $(\underline{p}^N, W^*(\underline{p}^N)) \equiv (\underline{p}^N, \overline{w}^N)$ *following any deviation constitutes an optimal punishment. Therefore, restricting attention to these SPNE in grim-trigger strategies is without loss of generality;*[28]
- *For* $\delta = 0$*, the most profitable SPNE is such that firms play the static NE* $(\overline{p}^N, W^*(\overline{p}^N)) \equiv (\overline{p}^N, \underline{w}^N)$ *for all* t *and so make positive profits.*

3.2 Cartel and collusive outcomes

This section characterizes first the cartel outcome and then the most profitable SPNE of the supergame for any value of δ.

Cartel outcome. The following proposition characterizes the cartel outcome – i.e., the prices and wage offers that maximize firms' joint profits, without imposing stationarity or symmetry assumptions – and provides conditions under which it is sustainable as a SPNE of the supergame:

Proposition 2 (cartel outcome). *The cartel outcome is stationary and symmetric: Maximizing industry profits requires all firms to charge the same price* p^M *to consumers and offer the same wage* w^M *to all workers over time, with*

$$p^M > \overline{p}^N \quad \text{and} \quad w^M < \underline{w}^N.$$

This outcome can be sustained as a SPNE of the supergame if and only if $\delta \geq \delta^M$*, with*

$$\delta^M < \delta_0^M.$$

[27] The equilibrium characterization is as in a Bertrand game with increasing marginal costs (Dastidar, 1995).

[28] That is, for any discounted profit that some SPNE can achieve, there always exists a SPNE using these grim-trigger strategies that yields the same discounted profit: The most profitable SPNE is in grim-trigger strategies. The no-rationing assumption is crucial for the existence of this zero-profit static NE, which greatly simplifies the analysis. Assumption (A) ensures that the possibility of rationing does not change qualitatively any of the results of the paper provided that there exists a punishment scheme yielding a discounted profit of zero to a deviating firm in the continuation game following any deviation.

Maximization of industry profits requires equally splitting production among all firms. The reason is that, because of employer differentiation from workers' viewpoint, the wage level to attract any overall labor force in the industry is minimized when all firms are active and offer the same wage to all workers: First, by the arguments in Lemma 1, wage discrimination by any firm would not be cost-efficient; second, any asymmetry in wage offers across firms would imply some misallocation of workers – i.e., that some workers do not work for their preferred firm – which necessarily increases the average wages to recruit any overall labor force.

A cartel internalizes the negative cross-firm consumer-demand externalities from setting low prices (for a given labor force) and labor-supply externalities from offering high wages (for a given output). These two anticompetitive effects reinforce each other: The internalization of price-externalities calls for a larger price, and the correspondingly lower production reduces the needed labor force, which depresses the wage offers; through the same mechanism, the lower wage offers because of the internalization of wage-externalities, reducing the hired labor force, make it optimal to reduce production, hence to raise prices. As a result, the cartel price is higher, and the cartel wage is lower than in the most profitable static NE.

The main result of Proposition 2 is that labor market power facilitates collusion – i.e., it lowers the critical discount factor beyond which the cartel outcome is sustainable as a SPNE of the supergame relative to the benchmark scenarios analyzed in Section 2.2. This is because, while punishment profits are still equal to zero (recall Corollary 1), in the presence of employer differentiation from workers' standpoint a firm deviating from the cartel outcome cannot reap the whole industry profit.

This is for two reasons. First, a firm i's deviation profit is strictly below the profit it would make if it were the only firm active in the labor market. Indeed, even though i can eliminate rivals' product market competition by slightly undercutting p^M, it still faces their competition in the labor market, as they do not expect the deviation and so offer w^M to all workers. This competition strictly reduces i's deviation profit: As, for all w_i, $L(w_i, w^M) < L(w_i, -\infty)$, it is more expensive for i to hire workers relative to the case where its rivals are out of the market. Second, even this hypothetical single-firm profit is strictly below the industry profit in the cartel outcome.[29] As argued above, employer differentiation implies that, for all w, $L(w, -\infty) < nL(w, w)$: Even in the absence of its rivals, a firm would need to raise its offers to get the same labor force of a n-firm industry, which implies facing larger costs to produce the cartel output $Q(p^M)$.

[29]This single-firm profit would be obtained by a deviating firm if the stage game was a sequential, production-to-order, game. The main results are robust with respect to the timing of the stage game: see Appendix B.2.

Collusion outcomes. Can some collusion manifest itself also when the cartel outcome is not sustainable (i.e., for $\delta < \delta^M$)? The most profitable SPNE, for any given $\delta < \delta^M$, is the pair (p, w) obtained by solving the following problem:

$$\max_{p,w} \pi(p, w) \quad \text{s.t. } \delta \geq 1 - \frac{\pi(p, w)}{\pi^D(p, w)}, \tag{P}$$

where the per-firm profit in the candidate SPNE outcome (p, w) is

$$\pi(p, w) \equiv p\frac{Q(p)}{n} - \left[wL(w, w) + r\, K\left(L(w, w), \frac{Q(p)}{n}\right)\right], \tag{8}$$

whereas the highest profit that a firm can make when deviating is given by

$$\pi^D(p, w) \equiv pQ(p) - C(Q(p); w), \tag{9}$$

in the period of deviation, and zero afterward, given that (by Corollary 1) firms revert to the zero-profit static NE after any deviation.[30] Then, the incentive-compatibility constraint in Problem (P) follows from the definition of grim-trigger strategies. The following proposition describes the solution to this problem:

Proposition 3 (multimarket collusion). *For all $\delta \in [0, \delta^M]$, the most profitable SPNE $(P^M(\delta), W^M(\delta))$ is obtained from the binding incentive-compatibility constraint and the optimality condition*

$$\frac{\partial\pi(\cdot)/\partial p}{\partial\pi^D(\cdot)/\partial p} = \frac{\partial\pi(\cdot)/\partial w}{\partial\pi^D(\cdot)/\partial w}, \tag{10}$$

and is such that, as δ increases from 0 to δ^M, $P^M(\delta)$ continuously increases from $\overline{p}^N$ to p^M and $W^M(\delta)$ continuously decreases from $\underline{w}^N$ to w^M.

For $\delta = 0$, firms are myopic and so cannot sustain any outcome more collusive than the most profitable static NE (Corollary 1); for all $\delta \geq \delta^M$, firms can sustain the cartel outcome, which features both a higher price and a lower wage (Proposition 2). Proposition 3 shows that firms optimally exploit any increase in the discount factor in the range $(0, \delta^M)$ to set both a more collusive (higher) price and (lower) wage. In particular, for any such level of the discount factor, the incentive-compatibility constraint binds, and the optimal collusive scheme equalizes the ratio of the marginal profit from collusion to the marginal profit from deviation across labor and product markets (i.e., with respect to w and p).

[30] As slightly undercutting p to capture all consumers' demand (and accordingly increasing the wage offer above w to minimize the production cost of $Q(p)$) is the most profitable deviation if $w = W^*(p)$, it is *a fortiori* so starting from a candidate collusive SPNE where $w < W^*(p)$ (colluding firms would never choose $w > W^*(p)$). The reason is that a deviating firm maximizes the gains from its rivals' more accommodating behavior in the labor market by maximally expanding its production, i.e. by serving the whole demand $Q(p)$.

The pattern of collusive prices in the presence of labor market power resembles the one arising in oligopoly (with constant marginal costs) when firms sell differentiated products (e.g., Ross, 1992). Indeed, similar to labor market power, product differentiation implies that a deviating firm, undercutting a candidate SPNE price p, can never appropriate the whole industry profit at p,[31] and its incentives to deviate continuously increase with p. The same collusive price patterns thus arise when firms exploit differentiation *vis-à-vis* workers in the labor market, rather than *vis-à-vis* consumers in the product market.[32]

Summing up. The presence of (individual-firm) labor market power implies that collusive behavior has completely different features:

1. *"Collusive" static equilibria*: Besides a zero-profit equilibrium, firms can sustain positive-profits equilibria, with inefficiently low production and employment levels, even absent repeated interactions (i.e., if $\delta = 0$).

2. *"Smooth" collusion*: The ability to collude – i.e., to raise prices and reduce wages – increases continuously with the discount factor (before firms reach the cartel outcome). Moreover, the critical discount factor to sustain the cartel outcome is lower than in the absence of individual-firm labor market power.

3. *Multimarket collusion*: The most profitable SPNE cannot be implemented by only colluding on a price level p, leaving each firm i free to choose its static-profit-maximizing wage offer to produce $q_i = Q(p)/n$: Firms find it optimal to also coordinate their behavior in the labor market – i.e., to suppress their wage offers below $W^*(p)$ in order to reduce the industry production costs (formally, $W^M(\delta) < W^*(P^M(\delta))$ for all $\delta > 0$: see Section 4.1).

4 Policy implications

Section 3 has shown that colluding firms are able to set supra-competitive prices and sub-competitive wage levels, harming consumers and workers alike. To prevent this possibility, antitrust authorities must monitor collusive behavior. Their monitoring activities may target preventing collusion in the labor market, the product market, or both, as examined in Section 4.1. When collusive behavior is hard to monitor – e.g., firms are able to

[31] Even if firms sell differentiated products, in the presence of labor market power a deviating firm's average cost would increase when it lowers its price and expands its production, which *ceteris paribus* facilitates collusion.

[32] As shown in Section 4.1 below, this holds irrespective of whether firms collude only on prices or also on wages. Yet, product differentiation also implies that the static NE features positive profits, which limits the severity of punishments, at least in a SPNE in grim-trigger strategies (as considered in Ross, 1992), thereby destabilizing collusion. This effect is absent in this model: When firm differentiation, which facilitates collusion by dampening defection profits, comes from the labor market, severe punishments are ensured in the presence of product homogeneity.

tacitly collude, without leaving hard evidence to the competition watchdog – competition authorities or regulators can nonetheless employ several policy measures, in labor and/or product markets, to make collusion harder to sustain, as examined in Section 4.2.

4.1 Monitoring of collusion

In practice, monitoring of collusive behavior takes place through the detection of price- and wage-fixing agreements. Suppose that collusion in any market leaves hard evidence, and the antitrust authority can commit to a monitoring policy and levy hefty fines on firms caught colluding (as in Motta and Polo, 2003, and Choi and Gerlach, 2013, among many others). Then, if the antitrust authority monitors collusion in both labor and product markets, the best firms can do is play the most profitable static NE $(\overline{p}^N, \underline{w}^N)$ over time (the results of this section are unchanged if firms play any other static NE). If, on the contrary, collusive behavior is left unmonitored, firms reach the *multimarket collusion* outcome characterized in Proposition 3. However, the antitrust authority, when constrained by limited budget, can monitor collusion only in either the labor market or the product market. This section first characterizes the most profitable SPNE in these scenarios and then derives implications of the antitrust authority's monitoring policies.

Preliminaries. Consider a candidate SPNE outcome (p, w). In the analysis of Section 3.2, any deviation from (p, w) triggers the reversion to the zero-profit static NE. This section characterizes the most profitable SPNE if only deviations from p (resp., from w) trigger Nash-reversion – i.e., (p, w) is played at any t if and only if all firms have set p (resp., w) for all $t' < t$, no matter their choices of w (resp., of p); else, $(\underline{p}^N, \overline{w}^N)$ is played. In these equilibria, the variable whose choice does not trigger the punishment is set by each firm to maximize its static profit, resulting in static Nash behavior taking as given the value of the other variable.

These scenarios arise if collusion in each market requires communication among specialized middle managers – e.g., firms' HR directors need to communicate in order to coordinate on collusive wage levels, given the price set by firms' pricing managers (i.e., to set $w < W^*(p)$) – and such communication can be prevented, in either market, by the competition watchdog. Alternatively, the antitrust authority can only infer whether the prevailing price (resp., wage) is collusive given the observed labor force (resp., consumer demand) – e.g., because it cannot estimate the labor supply function, it is not able to ascertain whether $w < W^*(p)$; in this case, firms can be sued for collusion only if the observed price is supra-competitive given the observed firms' wage and labor force ($p > P^*(w)$ in the notation below). Both these microfoundations are detailed in Appendix B.1.

Monitoring of wage collusion. If wage collusion is monitored, firms collude only on the product price: As described above, given a candidate SPNE price p, along the equilibrium path each firm chooses the wage offers in order to minimize its production cost of $Q(p)/n$ anticipating that rivals do the same. This competitive behavior in the labor market introduces the *competitive wage constraint* $w = W^*(p)$ in the collusion Problem (P).

Denoting by $p^P \equiv \arg\max_p \pi(p, W^*(p))$ the *price cartel outcome*, which would emerge if firms could write down p in a legally binding contract, but could not collude on w, the following results hold:

Proposition 4 (price collusion). *Under price collusion, there exists a threshold $\delta^P \in (0, \delta_0^M)$ such that the most profitable SPNE is $(P^P(\delta), W^P(\delta))$, with $W^P(\delta) = W^*(P^P(\delta))$, and $P^P(\delta)$ being increasing in δ and such that $P^P(0) = \overline{p}^N$ and $P^P(\delta) = p^P$ for all $\delta \geq \delta^P$. Moreover,*

$$P^P(\delta) > P^M(\delta) \quad \textit{and} \quad W^P(\delta) > W^M(\delta),$$

for all $\delta > 0$.

This proposition shows two main results. First, even though wage collusion is not in place (i.e., firms compete in wages, given the price chosen in a collusive fashion), price collusion also harms workers. Price coordination allows firms to raise their prices, which implies that they face a lower demand, and so find it individually optimal to reduce their wage offers and hire fewer workers. Thus, as the discount factor grows, firms can sustain SPNE featuring both higher prices and lower wages, even if they can coordinate only on their pricing behavior.

Second, the monitoring of wage collusion only is beneficial to workers ($W^P(\delta) > W^M(\delta)$) but harmful to consumers ($P^P(\delta) > P^M(\delta)$). Intuitively, the optimality condition (10) entails that, under multimarket collusion, firms exploit their ability to collude (for a given value of δ) to induce a collusive allocation in both labor and product markets: In particular, $W^M(\delta) < W^*(P^M(\delta))$ for all $\delta > 0$. The higher wage levels prevailing because of competitive behavior in the labor market make it efficient to produce less and more expensive for a firm contemplating a deviation to recruit more workers to expand its production, rendering higher prices incentive-compatible.

Monitoring of price collusion. If price collusion is monitored, firms collude only on their wage offers: As explained above, given a candidate SPNE wage w, along the equilibrium path all firms have labor force $L(w, w)$ and compete in prices. The most profitable SPNE in this class solves Problem (P) subject to the *competitive price constraint* $p \leq P^*(w)$, where $P^*(w)$ is the highest price that prevents a firm from undercutting holding fixed the wage level w, obtained by equating the revenue from capturing rivals' demand and the cost of the extra endowment of capital needed to satisfy the additional demand given that the labor force is fixed at $L(w, w)$:

$$\frac{n-1}{n}pQ(p) = r\left[K(L(w,w), Q(p)) - K\left(L(w,w), \frac{Q(p)}{n}\right)\right]. \tag{11}$$

Denoting by $w^W \equiv \arg\max_w \pi(P^*(w), w)$ the *wage cartel outcome*, which would emerge if firms could write down w in a legally binding contract but could not collude on p, the following results hold:

Proposition 5 (wage collusion). *Under wage collusion, there exists a threshold $\delta^W \in (0, \delta_0^M)$ such that the most profitable SPNE is $(P^W(\delta), W^W(\delta))$, with W^W and $P^W \leq P^*(W^W)$ being decreasing and increasing in δ, respectively, and such that $(P^W(0) = \overline{p}^N < P^*(\underline{w}^N), W^W(0) = \underline{w}^N)$ and $(P^W(\delta) = P^*(w^W), W^W(\delta) = w^W)$ for all $\delta \geq \delta^W$. Moreover, for all $\delta > 0$,*

$$P^W(\delta) \leq P^M(\delta) \quad \text{and} \quad W^W(\delta) \leq W^M(\delta),$$

with strict inequalities whenever the competitive price constraint binds ($p = P^*(w)$).

By colluding in the labor market, firms can sustain higher prices in the supergame, relative to the highest static NE price $\overline{p}^N$, even absent price collusion. By coordinating to lower their wage offers, they end up hiring fewer workers; a lower labor force, in turn, makes individually rational to produce less – i.e., to set higher prices (indeed, $P^*(w)$ is decreasing in w). Wage collusion is thus harmful not only to workers, but to consumers as well.

However, as the competitive price constraint is violated at the multimarket cartel outcome – i.e., $p^M > P^*(w^M)$ – for sufficiently large values of δ, the monitoring of price collusion prevents firms from achieving the profits they make under multimarket collusion.[33] For these values of δ, when they cannot collude on prices, firms inefficiently exploit all their ability to collude to suppress wages, implying that the monitoring of price collusion only is harmful to workers ($W^W(\delta) < W^M(\delta)$) but beneficial to consumers ($P^W(\delta) < P^M(\delta)$). In particular, competitive price behavior depresses the marginal revenue product of labor and entails a more significant increase in demand for a deviating firm undercutting the candidate equilibrium price exacerbating the diseconomies of scale effect, thereby making lower wages both more efficient and incentive-compatible.

[33] Wage collusion instead suffices to implement the multimarket collusion outcome for relatively small values of δ. As in the static game firms would be indifferent between undercutting or not the price $P^*(w)$ when holding fixed their wage offers at w, they have strict *static* incentives to undercut $P^*(w)$ given that they can indeed also increase their offers – i.e., $\pi^D(P^*(w), w) > \pi(P^*(w), w)$ for all w. By a continuity argument, so long as δ is sufficiently small, for any given w the price $P^*(w)$ is too large to be sustainable as a SPNE outcome under multimarket collusion. Then, the competitive price constraint does not bind, so (unlike price collusion) wage collusion can achieve the multimarket collusion outcome. This result is, however, specific to perfect Bertrand competition in the product market.

Remark (Multimarket-contact effect). To conclude the analysis, it is interesting to notice that, unlike in the multimarket contact model with independent markets (Bernheim and Whinston, 1990), the multimarket cartel outcome (p^M, w^M) might be sustainable at a lower critical discount factor relative to *both* cartel outcomes in each single market under competitive behavior in the other market – i.e., it might be that[34]

$$\delta^M < \min\{\delta^P, \delta^W\}.$$

The reason is that competitive behavior in one market does not eliminate the gains from a deviation in that market once a firm deviates in the other one: $w = W^*(p)$ is a competitive wage level for fixed p (i.e., to produce $q_i = Q(p)/n$), but a deviating firm who undercuts p has also incentives to raise its wage; similarly as for the price level $p = P^*(w)$. For the reasons discussed above, a price (resp., wage) cartel inefficiently pushes the price up (resp., the wage down) relative to a multimarket cartel, which may enhance the profitability of such a joint deviation.[35]

Implications. Suppose that the antitrust authority adopts a consumer-surplus standard, in line with its current *narrow mandate* in many countries. Then, the foregoing analysis has the following immediate policy implications on the monitoring of collusion:

Corollary 2 (monitoring of collusion). *For all $\delta > 0$, $P^P(\delta) > P^M(\delta) \geq P^W(\delta) > \overline{p}^N$, implying that consumers:*

- *benefit from the monitoring of both price and wage collusion;*
- *benefit from the monitoring of wage collusion if and only if price collusion is also monitored;*
- *always benefit from the monitoring of price collusion, and more so when also wage collusion is monitored.*

Labor market power makes collusion easier to sustain, which increases the need to monitor collusive behavior. Moreover, if only price collusion is monitored, firms have the incentives and ability to collude to depress wages, which also causes consumer harm by making higher prices individually rational. Consumer protection thus requires antitrust

[34]This is a theoretical possibility in light of the results of Proposition 4 and 5. The possibility result can be established by simulating the model using a Cobb-Douglas production function, an isoelastic product demand, and a logit labor supply function (Matlab code is available upon request). In the case of independent product markets considered by Bernheim and Whinston (1990), instead, in order for the cartel outcome to be sustainable in both markets under multimarket contact, it needs to be sustainable in at least one of the two markets (no matter the behavior in the other market) absent multimarket contact.

[35]Similar to the previous literature on multimarket contact (Matsushima, 2001), labor market power can have further procollusive effects by allowing firms to detect better deviations in settings with demand shocks and imperfect monitoring *à la* Green and Porter (1984). See Appendix B.2 for a discussion.

authorities to monitor also wage collusion actively.[36] Yet, shifting all their attention and budget towards monitoring wage collusion would not align with their statutory objective: Under pure price collusion, consumers are even worse off than under no monitoring at all of collusive behavior. For these reasons, the welfare stakes for consumer-surplus-oriented authorities deciding whether to monitor price collusion are higher when wage collusion is monitored.

Overall, these results point to a complementarity in the monitoring of collusive behavior in labor and product markets. Granting antitrust authorities a *broad mandate* – specifically, including worker protection in their objectives – is unnecessary to drive their efforts against wage collusion,[37] as long as they are sufficiently resourced to address collusion in both labor and product markets.

4.2 Restraining collusion

Suppose now that antitrust authorities are unable to monitor collusive behavior. Firms' ability to set supra-competitive prices and sub-competitive wages can be constrained by competition or regulatory measures in labor and product markets.

Minimum wage regulation. One of the most important policies, especially in low-skilled labor markets, is minimum wage regulation, which constrains colluding firms' possibility of suppressing wages.

Proposition 6 (minimum wage regulation). *For every $\delta \in (0, \delta^M)$ there exists $\varepsilon > 0$ such that a minimum wage $\underline{w} \in (W^M(\delta), W^M(\delta) + \varepsilon)$ imposed by regulation raises both employment and price levels.*

The effects of minimum wage regulation crucially depend not only on whether labor markets are perfectly competitive or oligopsonistic, as recognized by a long-standing literature (see, e.g., Belman and Wolfson, 2014, for a comprehensive survey) but also on firms' competitive conduct. If labor markets are perfectly competitive, any minimum wage regulation raising the wage above the competitive level w_0 implies that firms find it optimal to hire fewer workers, which, no matter whether they compete or collude in the product market, leads to higher prices. Conversely, in the presence of labor market power, firms would ration workers only if the minimum wage is much larger than the wage level prevailing in equilibrium. A *locally binding* minimum wage instead raises employment levels: Firms no longer have incentives to hire fewer workers in order to be able to pay lower wages when these are set by regulation. This employment-enhancing effect, in turn,

[36] Similar arguments justify the *per se illegality* antitrust status of wage cartels even based on a pure consumer surplus standard. Indeed, in the presence of a wage cartel, firms colluding on prices would be able to sustain equilibria with higher prices and lower wages relative to the multimarket collusion outcomes characterized in Proposition 3 for all $\delta < \delta^M$ (see Appendix B.2).

[37] A broad mandate may indeed have unintended consequences, as discussed in Tirole (2023).

makes it optimal to expand production, resulting in lower consumer prices. Once again, this is true in equilibrium under competitive behavior (e.g., Tong and Ornaghi, 2021) and in the cartel outcome (or, under collusion, for $\delta \geq \delta^M$).

The novel implication of this model is that, if firms engage in collusive behavior, but their ability to collude is constrained by incentive-compatibility (i.e., $\delta \in (0, \delta^M)$), the introduction of a (locally) binding minimum wage, or a (local) increase thereof, implies that firms, being unable to depress wages as much as they would like to, exploit their ability to collude in the product market. That is, firms offer to all workers the minimum wage to comply with regulation (as above, rationing of workers is not optimal if the minimum wage is only slightly higher than the equilibrium wage in the absence of regulation), but they are able to sustain higher prices.

Minimum wage regulation can thus have the same effects as monitoring wage collusion only: it can backfire on consumer surplus if antitrust authorities are not able to monitor collusive behavior in product markets. Labor market power and collusive behavior thus jointly rationalize the evidence in several studies since Card and Krueger (1994) that increases in the minimum wage can raise both employment and prices. Such firms' reactions are thus suggestive of collusive conduct.

Regulation of unions and collective bargaining. In the considered model, labor market power derives from workers' heterogeneous preferences for working at different firms. In reality, however, the extent of firms' labor market power – to be intended as their ability to mark down the competitive wage w_0 even without engaging in collusion – also depends on workers' bargaining power: In many countries, the prevailing wage levels are determined through collective bargaining between employers and unions.

Appendix B.2 develops a model where firms $i \in \mathcal{N}$ choose employment levels ℓ_i and wages are determined through collective bargaining. Suppose that the distribution Ξ is such that the market clearing wage, which would prevail if unions have no bargaining power, would always be below the competitive wage w_0, which would prevail if unions have full bargaining power. The wage prevailing under collective bargaining is the weighted average of the market clearing wage and the competitive wage, where weights reflect unions' bargaining power.[38]

Weaker unions imply a lower cost to hire any labor force for the firms, resulting into lower consumer prices. However, as labor market power facilitates collusion, weakening unions' bargaining power increases firms' ability to set collusive prices, which results into relatively small pass-throughs of the wage mark-downs to consumer prices. As a result, when collusion is a concern, the trade-off between consumer and worker surplus calls for

[38]Formally, abstracting away from employer differentiation, the market clearing wage as function of industrywide labor demand $L \equiv \sum_i \ell_i$, $w^*(L) < w_0$, is obtained from $J[1 - \Xi(w_0 - w^*(L))] = L$; the prevailing wage is then $W(L) \equiv \alpha w^*(L) + (1 - \alpha) w_0$, where $\alpha \in [0, 1]$ is an inverse measure of trade unions' bargaining power.

the protection of unions' bargaining power: Increasing firms' labor market power acts as a facilitating device, implying that firms retain much of the corresponding gains, and so workers lose much more than consumers gain.

Open product markets. Policy measures that open up product markets (e.g., free trade agreements), by exposing firms within each local market to competition by firms producing in different geographic markets, can make it impossible for them to sustain collusion, ensuring consumer access to competitively priced products. This section argues that this result does not necessarily hold if, in line with empirical evidence (e.g., Marinescu and Rathelot, 2018), labor markets remain local.

Consider H distinct geographic markets $h \in \{1, ..., H\}$, each composed of a labor and a product market as described so far. Such geographic markets are independent and, for simplicity but without loss of insights, identical. Within each market, therefore, the cartel outcome (p^M, w^M) is as in Proposition 2 for all H. Suppose that firms in each local market h can sell their products in any of the H markets: the larger H, the more globalized the product market;[39] still, firms can only recruit workers from their local labor market h. The no-rationing assumption still holds within each market, but a firm charging a price below the one prevailing in some other markets is free to choose how many of these markets it wants to serve.

The following proposition describes the impact of the available number of markets H on the critical discount factor, denoted by $\delta^M(H)$, to sustain the cartel outcome (p^M, w^M) within any geographic market.[40]

Proposition 7 (product market globalization). *There exists a (finite) number of markets H^* such that $\delta^M(H)$ is increasing in H if and only if $H < H^*$ and, for all $H \geq H^*$, $\delta^M(H) = \delta^M(H^*) < (nH^* - 1)/(nH^*)$.*

For $\delta = 0$, at $p = \overline{p}^N$ firms are indifferent between undercutting or not when, if they do so, they have to serve only consumers in their local market. By the diseconomies of scale effect, the opportunity for a deviating firm to serve also consumers in other markets is then valueless: As a result, for all $H \geq 1$, the most profitable SPNE features $(\overline{p}^N, \underline{w}^N)$ in each market. When δ grows larger, however, an increase in H can limit the extent of collusion: In the SPNE characterized in Proposition 3 (played in all markets), outputs and wages are relatively low, implying that a deviating firm may gain strictly more by also serving consumers in some other markets; accordingly, colluding firms in

[39] Indeed, one can equivalently consider an exogenous number $\overline{H}$ of geographic markets divided into $\overline{H}/H$ disjoint *free trade areas*, so that each firm can sell its products in up to H markets: $H = 1$ corresponds to the case of *local product markets* analyzed so far, $H = \overline{H}$ corresponds instead to a *single market*.

[40] There is no need of *cross-market collusion* to sustain (p^M, w^M) within each geographic market – i.e., firms within a market h only need to collude among each other to sustain (p^M, w^M) (see Appendix A).

each market have to set lower prices and higher wages to preserve incentive-compatibility. As this is especially true at the cartel outcome, it follows that the critical discount factor to implement (p^M, w^M) may well be larger when $H > 1$. Nevertheless, serving more and more markets requires a deviating firm to recruit more and more workers, which entails paying increasingly high wages (this is true even when the dimension J of the local labor market is sufficiently large – i.e., there are more available workers than a deviating firm would want to hire at w_0 to serve consumers in all H markets). This implies that a deviating firm would not serve consumers in more than H^* markets.

Therefore, for product market globalization to guarantee a competitive outcome, it needs to be accompanied by the globalization of labor markets and the absence of labor market power.As soon as one of these conditions fails, the monitoring of collusion is needed to protect consumers' and workers' interests. Indeed, the procollusive effects of increased labor market power (Yeh et al., 2022) may outweigh the procompetitive impact of the increasing globalization of product markets, thereby leading to higher mark-ups, consistent with the empirical evidence in De Loecker and Eeckhout (2018).

Merger policy. Another traditional policy tool to preserve competition and avoid facilitating collusion consists in preventing concentration, by adopting a strict merger policy that prohibits horizontal mergers (absent substantial efficiencies). A merged entity, in fact, has weaker incentives to deviate from a collusive arrangement because of the internalization of business-stealing externalities across the merged units. This section argues that, in the presence of labor market power, mergers that increase labor market concentration can produce similar anticompetitive effects even if the merging parties sell demand-independent products. These mergers can therefore be labeled as *conglomerate mergers*.

Indeed, while the foregoing analysis has considered (strategic) firms in only one industry, often workers (especially in low-skilled labor markets) can be employed by firms selling different products to consumers within a geographic market. Suppose that the n firms considered throughout (with n being even for the sake of the exercise) still recruit from the same labor market, but now firms $i \in \{1, ..., n/2\}$ sell product A and the other firms $i \in \{n/2 + 1, ..., n\}$ sell product B; product markets $z \in \{A, B\}$ are identical and independent, with consumer demand $q_z \equiv Q(p_z)/2$ in each market, where p_z denotes the (minimum available) price for product z (this normalization ensures that the cartel outcome remains as in Proposition 2). This section will contrast the scenario in which the n firms are independent (*single-product firms*) with the one in which the same *multiproduct firms* are active in both product markets – i.e., firm i, selling product A, merges with firm $i+n/2$, selling product B, for all $i \in \{1, ..., n/2\}$. In either case, workers' preferences depend on the features (e.g., the location) of each production plant, irrespective of its ownership: in both scenarios, workers' labor supply is as above. I again restrict attention to the most profitable stationary symmetric SPNE, where all firms set price $p_z \equiv p$ and offer wage $w_z \equiv w$ for $z = A, B$.[41]

As higher wage offers by firms selling product z entail a negative externality on the labor supply of firms selling the other product, firms selling in independent product markets have incentives to collude together to internalize these cross-market wage-externalities.[42] Conglomerate mergers allow to relax incentive-compatibility constraints in this *cross-market collusion* problem. This is because, while along any equilibrium path a multiproduct firm earns the sum of its units' profits, deviation profits are *subadditive*.

The intuition is as follows. A multiproduct firm anticipates that any deviation triggers punishments yielding zero profits in both $z = A, B$, so it optimally deviates by undercutting p in both product markets. Yet, increasing the wage offer to recruit more workers for its production unit in market A imposes a negative externality on its subsidiary in market B, as it makes it increasingly costly to expand the labor force hired for production in market B. As a result, multiproduct firms cannot obtain the same per-market deviation profit of a single-product firm, and so have weaker incentives to deviate from any candidate SPNE (p, w), and in particular from the cartel outcome:

Proposition 8 (conglomerate mergers). *Compared to the scenario with distinct single-product firms in each product market $z = A, B$, the critical discount factor to sustain the cartel outcome is strictly lower when the same firms operate in both markets.*

The result that conglomerate mergers, by inducing multimarket contact, can lead to more collusive outcomes dates back to Bernheim and Whinston (1990). However, in their setting with perfectly competitive input markets, pooling incentive constraints across independent product markets *(i)* has an effect only if markets are asymmetric along some specific dimensions, and *(ii)* cannot yield a strictly higher price in all markets. In contrast, the anticompetitive multimarket-contact effects of conglomerate mergers are much more robust in the presence of labor market power, given that, even absent asymmetries, they simultaneously lead to higher prices in all markets and, on top of this, also entail lower wages.[43] Both effects have been found in the data: Ciliberto and Williams (2014) have shown that, in the airline industry, carriers with a significant amount of multimarket contact can sustain near-perfect cooperation in setting fares; Arnold (2019) has found that

[41] Once again, such equilibrium is in grim-trigger strategies, as the results of Corollary 1 hold in each product market z, for any w_{-z} offered by firms in the other market – i.e., perfect within-market Bertrand competition ensures the existence of a zero-profit continuation equilibrium following a deviation in either market.

[42] Under such cross-market collusion, firms may be bound, for incentive-compatibility reasons, to reduce consumer prices, relative to the case of within-market collusion, where firms selling each product z only collude among each other correctly anticipating the wage offer by firms in the other product market (see Appendix B.3 for the details). By contrast, the internalization of wage-externalities implies that cross-market collusion unambiguously results in lower wages.

[43] The subadditivity of deviation profits, which drives the results, generalizes to an arbitrary number of product markets and applies also to asymmetric SPNE outcomes $(p_z, w_z)_{z=A,B}$, which shall be considered if product markets are asymmetric.

mergers that result in significant increases in local labor market concentration produce a decline in wages and these effects are not driven by changes in product market power (see also Prager and Schmitt, 2021, and Berger et al., 2023).

5 Employment contracts and facilitating devices

The foregoing analysis has considered, for simplicity, a spot labor market. In reality, however, workers typically sign long-term employment contracts, which can be subject to horizontal and vertical restraints, such as no-poaching and non-compete agreements, and a body of regulations constraining firms' wage-setting behavior, including pay-transparency and pay-equity provisions. This section explores how these labor-market-specific features affect the sustainability of collusion. The results on labor market collusion do not rely on the assumption that oligopsonists also interact in the same product market (see Appendix B.4). However, the analysis is conducted within the oligopoly-oligopsony setting to also provide insights into price collusion applicable in this scenario.

Set-up. Consider an overlapping generation model with cohorts of *myopic* T-period lived workers.[44] Each generation is composed of a measure J of workers j as defined in Section 2 with $\xi_{i,j} \sim^{\text{i.i.d.}} \Xi(\cdot)$ being time-invariant (*persistent type*). Consistent with real-world practices, firms offer long-term stationary contracts, specifying a per-period salary with the possibility of Pareto-improving renegotiation and workers' option to quit at any future period (*one-sided commitment*).[45] To simplify the analysis, suppose that if a worker is not employed in the considered industry at *age* $\tau = 1, ..., T$, it cannot be hired in future periods, i.e., at any age $\tau' > \tau$ – e.g., it leaves the considered industry or local labor market and enjoys a per-period outside option w_0.[46]

Therefore, at each period t, workers who were aged $\tau = T$ in $t-1$ *retire*, and firms, simultaneously, *(i)* make offers to newcomers ($\tau = 1$); *(ii)* possibly renegotiate wages with their incumbent employees (i.e., previously hired workers of age $\tau = 2, ..., T$) and make offers to rivals' incumbent employees (if allowed: see below); and *(iii)* set their products' price. Then, all workers in the market observe their available offers and choose which firm to work in (if any), and consumers make their purchase decisions. A firm i then needs to employ variable capital $k_i = K(\sum_\tau \ell_{\tau,i}, q_i)$ to satisfy consumers' realized

[44] The results are qualitatively unchanged if workers are farsighted and discount future payoffs at the common rate δ (see Appendix A for the details). To keep the model stationary, I assume that the first T generations of workers are simultaneously available at the initial time $t = 0$.

[45] Permanent employment contracts specify the current salary, which normally can only be increased in the future, as *downward nominal wage rigidity* is prevalent (Lebow et al., 2003). Nevertheless, employees can decide to leave the firm at any time. Whether firms can fire workers hired in period t in future periods $(t+1, ...)$ is immaterial to the results.

[46] Together with long-term contracts, this assumption rules out *ratchet effects* deriving from employers learning workers' types $\xi_{i,j}$ over time through their contract acceptance decisions.

demand, q_i in Eqn. (1), given the measure $\ell_{\tau,i}$ of workers of each age at its disposal, which depends on contract offers.

In particular, newcomers' labor supply, if i's rivals all offer them the same wage w_1, is still given by $\ell_{1,i} = L(w_{1,i}, w_1)$, the function $L(\cdot)$ being defined in Lemma 1. The allocation of the available workers $\sum_i \ell_{1,i}$ of age $\tau = 2, ..., T$ across firms again depends on currently available wage offers in a static fashion – i.e., worker j chooses at each age τ to work for the firm i providing the highest overall utility $w_{i,j} + \xi_{i,j}$ in that period.

5.1 No-poaching agreements

A no-poaching agreement (hereafter, NPA) among any subset of firms prevents the signatories from hiring each others' incumbent workers: If firm i has signed a NPA in period t, it cannot propose any offer to workers who have worked for the other signatories in period $t-1$.[47] If allowed by competition authorities, these agreements are legally binding: Even if a firm deviates from a collusive outcome path, it cannot violate the NPA (e.g., the agreement would be enforced in court, or hefty fines are levied on the violating firm). At each period of the game, each firm decides whether to join the NPA (that rules for one period) before labor- and product-market decisions are made as described above, with the no-poaching constraint for the signatories.

Analysis. Along any stationary equilibrium path, where all employed workers obtain the same wage w and firms set the same price p over time, even in the absence of NPAs, firms do not make poaching offers and only replace *retirees* with newcomers hired at the same wage w. Indeed, so long as a firm wants to hold the same labor force over time, poaching rivals' incumbent workers would not be optimal, as it would require paying them a higher wage than the equilibrium level w at which it can recruit newcomers to replace all its retirees (given that, by revealed preference, rivals' incumbent workers prefer working for their current employer at the equilibrium wage). Thus, the fact that firms do not poach each others' workers *per se* tells little about whether or not their behavior is collusive. However, this does not mean that a ban on NPAs is inconsequential, because the presence of binding NPAs affects the defection profit that can be captured by a firm deviating from a candidate stationary equilibrium path, as argued below.

Suppose, first, that NPAs are not in place (e.g., they are banned by competition authorities). Then, starting from any candidate most profitable SPNE (p, w), the highest one-shot deviation profit for a firm i is obtained by slightly undercutting the candidate equilibrium price p and making wage offers that minimize the production cost of $Q(p)$:

[47] In order for NPAs to have a bite, there must be some incumbent workers in the market, so that the following analysis, formally speaking, applies for $t \geq 1$.

formally,

$$\min_{\{w_{\tau,i}\}} \sum_{\tau=2,...,T} \left[wL(w,w) + w_{\tau,i}\tilde{\ell}_{\tau,i}(w_{\tau,i},w) \right] + w_{1,i}L(w_{1,i},w) + r\,K(\cdot), \tag{12}$$

as the labor force of a deviating firm i consists of *(i)* its incumbent workers $L(w,w)$, whom it keeps at the candidate equilibrium wage w, *(ii)* poached rivals' workers $\tilde{\ell}_{\tau,i}$, i.e. their incumbent workers that i can hire by offering $w_{\tau,i} > w$,[48] and *(iii)* newcomers, whose labor supply is $L(w_{1,i},w)$. Then, for any collusive level of wages (i.e., $w < W^*(p)$, this function being defined following the same steps as in Lemma 3), the deviating firm finds it optimal to set $w_{\tau,i} > w$ for all $\tau = 1,...,T$. The option to make poaching offers to rivals' incumbent workers is thus valuable from a deviating firm's standpoint. The reason is that the wage offer needed to expand its labor force to any given level is lower when it can be offered to a larger pool of workers, which is the case when a deviating firm has the chance of poaching rivals' workers.

Following any deviation, firms revert to a continuation equilibrium where they set prices competitively and the deviating firm makes zero profits; this constitutes an optimal punishment and implies that the best deviation is indeed the one, characterized above, that maximizes profits in the period of defection.

If, instead, NPAs are allowed, a stationary equilibrium where on-path all firms sign these agreements at any period always exists. If one or more firms deviate and do not sign the NPA, this deviation is detected before firms make wage offers and set their prices, so they immediately revert to a zero-profit continuation equilibrium.[49] Therefore, a deviation can be profitable only in the wage- and price-setting stage after signing the NPA. Then, by the arguments above, the impossibility of poaching rivals' workers strictly reduces the profits that a deviating firm can obtain in the deviation period; following any deviation, a continuation equilibrium where no NPAs are signed anymore and the deviating firm makes zero profits is played.[50]

[48] Formally, $\tilde{\ell}_{\tau,i}(w_{\tau,i},w) \equiv (n-1)L(w,w)\Pr\left[w_{\tau,i} + \xi_i \geq w + \max_{i'\in\mathcal{N}\setminus\{i\}} \xi_{i'} \middle| \xi_i < \max_{i'\in\mathcal{N}\setminus\{i\}} \xi_{i'}\right]$. As these workers are identical from i's viewpoint, by the same arguments as in Lemma 1, offering them a uniform wage is optimal. In this model, as a deviating firm deprives its rivals of all consumers and workers are contestable in future periods, these firms would have no incentives to make their workers a counteroffer to the poaching offer.

[49] A deviating firm would derive no advantage from the possibility of making poaching offers if rivals can renegotiate their incumbent employees' wages, given that these workers always prefer remaining where they are if their current employer matches the poaching offer; moreover, the signatories could try to poach the deviating firm's workers, inducing it to increase the remuneration to its incumbent workers. Firms thus remain symmetric in the continuation game, so a zero-profit continuation equilibrium always exists.

[50] Starting from a candidate equilibrium where each firm expects others not to sign the NPA, unilaterally signing it is inconsequential; then, all workers hired by the deviating firm can be poached away by its rivals in the future, exactly as in the game without NPAs, so that a zero-profit continuation equilibrium still exists.

Results and implications. By the above analysis, the possibility of signing a binding NPA weakens defection incentives, allowing firms to sustain more collusive arrangements – i.e., the most profitable SPNE features firms signing binding NPAs, and setting lower wages and higher prices relative to the scenario where these agreements are banned by competition authorities:

Proposition 9 (no-poaching agreements). *Banning NPAs increases the critical discount factor to sustain the cartel outcome and leads to strictly higher wages and lower prices for all lower values of* δ*.*

This model provides an anticompetitive rationale for the use of NPAs in low-skilled labor markets, where relation-specific investments are not a concern and workers are easily replaceable: These agreements can be employed as facilitating practices – i.e., are instrumental to (wage and, eventually, price) collusion. A ban on NPAs, by tightening incentive-compatibility constraints, thus leads to higher wages and, if employers also interact in the same product market, lower prices. Therefore, the *per se illegality* status of naked NPAs in the US legislation, and their similar consideration as *by object restrictions of competition* in the EU, is consistent with both worker and consumer protection.

The above analysis also rationalizes the recent empirical evidence on franchise NPAs and a ban thereof. Commentators have observed that a series of vertical NPAs between a franchisor and multiple franchisees eliminates competition among the latter as effectively as horizontal agreements among themselves orchestrated by the franchisor.[51] Even if competing firms in the industry are franchisees of different franchisors, all of them imposing franchise NPAs would facilitate intra- and inter-brand collusion (though to a lesser extent than an industrywide NPA) by reducing the pool of available workers that any deviating franchisee can attract. Krueger and Ashenfelter (2022) have documented that, until recently, over half of all franchise agreements in the US, at companies including fast-food restaurants and consumer staples (actually, more than in industries with higher average wages and education levels), included provisions barring franchisees from hiring one another's workers. Then, because of the legal cases and proposed legislation, many chains have removed such clauses from their contracts, which, as shown by Lafontaine et al. (2023) using data from the chain restaurant industry, has led to higher wages.

5.2 Related facilitating devices

Non-compete agreements. A non-compete agreement (hereafter, NCA) between a firm and any of its employees prohibits the latter from working for rival firms in the following period. Formally, when negotiating with a newcomer ($\tau = 1$), a firm i can include a one-period NCA in its offer at an additional remuneration $\omega_{1,i}$, on top of the

[51] See, e.g., `https://www.concurrences.com/en/bulletin/special-issues/no-poach-agreements/`.

per-period wage $w_{1,i}$; in this case, if the worker accepts the offer, it commits not to work for a rival firm $i' \in \mathcal{N} \setminus i$ at age $\tau = 2$. In the following period, a firm can renegotiate the wage (i.e., eventually offer $w_{2,i} > w_{1,i}$) and eventually propose a remuneration $\omega_{2,i}$ to sign another one-period NCA, and so on.

As, along any stationary equilibrium path, workers never expect to receive any attractive poaching offer from other firms, NCAs are signed for free (i.e., $\omega_{\tau,i} = 0$ for all τ and i) even by farsighted workers. In an equilibrium where all workers sign NCAs, a firm contemplating a deviation is *de facto* unable to make offers to all rivals' incumbent workers, precisely as in the presence of an industrywide NPA. However, NCAs may not be a perfect substitute for NPAs. The reason is that a deviating firm offering NCAs obtains a larger pool of non-contestable workers, which may make it harder for rivals to punish its deviation. Unless this effect dominates, banning NCAs results in less collusive outcomes.

Corollary 3 (non-compete agreements). *Banning NCAs may increase the critical discount factor to sustain the cartel outcome and lead to higher wages and lower prices for all lower values of δ.*

This result is a possible theoretical explanation for empirical evidence that, similar to NPAs, NCAs cover a significant fraction of less-educated and low-wage workers (Starr et al., 2021), and their ban for these workers resulted in an increase in wages (Lipsitz and Starr, 2022).

Pay-transparency and pay-equity regulations. Pay-equity rules, such as the US Equal Pay Act of 1963, are based on the principle "equal pay for equal work"; pay-transparency regulations, making workers aware of eventual discrimination, are instrumental to their enforcement.

In this model, all employed workers receive equal compensation along any stationary equilibrium path; yet, when a firm deviates, it optimally offers a higher wage to newcomers (and, absent NPAs or NCAs, to rivals' incumbent workers) without correspondingly increasing the wage of its incumbent workers (as these receive no poaching offers in the deviation period). Pay-equity regulations would instead force the deviating firm to do so, reducing its deviation incentives:[52]

Corollary 4 (pay-equity regulations). *Pay-equity regulations forbidding firms from wage-discriminate among their employees reduce the critical discount factor to sustain*

[52] Unlike the possibility of signing NCAs, pay-equity regulations cannot help a deviating firm to obtain a positive profit in the punishment phase: Together with downward nominal wage rigidity, pay-equity regulations imply that the deviating firm needs to pay an even (weakly) higher wage in the subsequent periods, which disadvantages it *vis-à-vis* its rivals. Thus, similar to NPAs, pay-equity regulations only add a binding constraint to the (one-shot) deviation profit-maximization problem.

the cartel outcome and lead to strictly lower wages and higher prices for all lower values of δ.

The facilitation of employer collusion, coming as an unintended consequence of pay-transparency and pay-equity regulations, may be an explanation for the empirical evidence showing that, on top of narrowing coworker wage gaps, these regulations have often also led to lower average wages (see Cullen, 2024, and references therein).

6 Conclusion

Employer collusion is not a recent phenomenon. The idea that employers have the incentives and the ability to collude to depress wages dates back to Adam Smith; recently, Delabastita and Rubens (2022) have provided historical evidence of wage cartels among Belgian coal firms in the 19th century. The fact that "until recently economists assumed that labor markets are fairly competitive" (Krueger and Posner, 2018) explains why collusive behavior has only been investigated, both theoretically and empirically, in product markets. However, the recent evidence that firms enjoy significant labor market power and a rising number of antitrust cases involving employers' coordinated behavior make it paramount to understand the economics of collusion in labor markets, guiding antitrust and regulatory interventions.

Prima facie, one might think that *turning upside-down* models of collusion in oligopoly would suffice for this purpose: The behavior of firms with labor market power colluding to depress wages may simply mirror the well-studied one of firms with product market power colluding to raise prices. Adopting this simplistic view would be misleading, as it overlooks two key aspects. First, firms often interact in the same labor and product markets, and have market power in both. Therefore, it would be erroneous to view labor market and product market collusion as separate phenomena that do not interact: Understanding labor market collusion requires building a theory of multimarket contact in vertically related markets. Second, the labor market has distinctive features absent in product markets. In particular, complex long-term employment contracts, whose terms are partly subject to public regulation, are prevalent, as opposed to simple one-shot transactions at firms' posted prices occurring in most product markets. A theory of employer collusion must therefore unveil what contractual clauses can constitute labor-market-specific facilitating practices and how regulation of employment contracts can affect the scope for collusion.

This paper has developed a theory of collusion in the presence of labor market power that encompasses both these aspects. It has shown that the impact of oligopsony power on collusive behavior is twofold. First, labor market power increases the scope for collusion, in several respects: *(i)* Firms can sustain the cartel outcome at a lower critical discount factor than in the absence of individual-firm labor market power; *(ii)* when firms operating in independent product markets hire from the same labor market, oligopsony power fosters broader cross-market collusion, and conglomerate mergers can produce anticompetitive multimarket-contact effects; *(iii)* firms can employ labor-market-specific horizontal or vertical restraints, such as no-poaching and non-compete agreements, as facilitating practices. Second, labor market power makes collusion a multimarket phenomenon: Firms have incentives to cooperate in both labor and product markets; when colluding in one market is unfeasible (e.g., because of monitoring of collusive behavior by antitrust authorities), cooperation in the other market strengthens.

Since collusion harms workers and consumers alike, antitrust authorities' efforts to monitor also labor market collusion align with consumer protection – i.e., with a *narrow* statutory objective, as opposed to a *broad mandate* including worker protection. Effective monitoring of collusive behavior in labor and product markets is indispensable in safeguarding consumers' and workers' interests because, in the presence of labor market power, standard policy measures to promote product market competitiveness (e.g., free trade agreements) have limited effects on firms' ability to collude, and labor market regulations (e.g., pay-equity provisions) can have unintended procollusive implications.

Finally, some implications of this theory can help infer the presence of collusive behavior from observable wage and price data, which can guide antitrust authorities' monitoring actions. Specifically, *(i)* a simultaneous increase in employment and price levels following a rise in the minimum wage, *(ii)* higher average wages and prices following a ban on no-poaching or non-compete agreements, and *(iii)* a low pass-through to consumer prices of wage mark-downs due to an increase in firms' labor market power (determined by, e.g., the weakening of trade unions' bargaining power), are all indicative of collusive behavior. Existing evidence often aligns with these predictions; this paper will hopefully encourage further investigations.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica, 56*(2), 383–396.

Aghion, P., & Bolton, P. (1987). Contracts as a barrier to entry. *The American Economic Review, 77*(3), 388–401.

Arnold, D. (2019). Mergers and acquisitions, local labor market concentration, and worker outcomes. Available at SSRN 3476369.

Azar, J. A., Berry, S. T., & Marinescu, I. (2022). Estimating labor market power. NBER working paper No. w30365.

Belman, D., & Wolfson, P. J. (2014). *What does the minimum wage do?*. WE Upjohn Institute.

Berger, D., Herkenhoff, K., & Mongey, S. (2022). Labor market power. *The American Economic Review, 112*(4), 1147–1193.

Berger, D. W., Hasenzagl, T., Herkenhoff, K., Mongey, S., & Posner, E. A. (2023). Merger guidelines for the labor market. NBER Working Paper No. w31147.

Bernheim, B. D., & Whinston, M. D. (1990). Multimarket contact and collusive behavior. *The RAND Journal of Economics, 21*(1), 1–26.

Bhaskar, V., & To, T. (1999). Minimum wages for Ronald McDonald monopsonies: A theory of monopsonistic competition. *The Economic Journal, 109*(455), 190–203.

Burdett, K., & Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review, 39*(2), 257–273.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics, 36*(S1), S13-S70.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review, 84*(4), 772–793.

Chen, Z., & Rey, P. (2023). A theory of conglomerate mergers. TSE Working Paper No. 1447.

Choi, J. P., & Gerlach, H. (2013). Multi-market collusion with demand linkages and Antitrust enforcement. *The Journal of Industrial Economics, 61*(4), 987–1022.

Ciliberto, F., & Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry. *The RAND Journal of Economics, 45*(4), 764–791.

Cullen, Z. (2024). Is pay transparency good?. *Journal of Economic Perspectives, 38*(1), 153–180.

Dastidar, K.G. (1995). On the existence of pure strategy Bertrand equilibrium. *Economic Theory, 5*, 19–32.

Delabastita, V., & Rubens, M. (2022). Colluding against workers. Available at SSRN 4208173.

De Loecker, J., & Eeckhout, J. (2018). Global market power. NBER Working Paper No. w24768.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies, 38*(1), 1–12.

Gelles, G. M., & Mitchell, D. W. (1996). Returns to scale and economies of scale: Further observations. *The Journal of Economic Education, 27*(3), 259–261.

Gonzaga, P., Brandão, A., & Vasconcelos, H. (2014). Theory of semi-collusion in the labor market. Universidade do Porto, Faculdade de Economia do Porto, Working paper No. 522.

Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica, 52*(1), 87–100.

Harrington Jr, J. E. (2006). How do cartels operate?. *Foundations and Trends in Microeconomics, 2*(1), 1–105.

Ivaldi, M., Jullien, B., Rey, P., Seabright, P., & Tirole, J. (2003). The economics of tacit collusion. Final Report for DG Competition, European Commission.

Kreps, D. M., & Scheinkman, J. A. (1983). Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics, 14*(2), 326–337.

Krueger, A. B., & Ashenfelter, O. (2022). Theory and evidence on employer collusion in the franchise sector. *Journal of Human Resources, 57*(S), S324–S348.

Krueger, A. B., & Posner, E. A. (2018). A proposal for protecting low-income workers from monopsony and collusion. The Hamilton Project. Policy Proposal, 2018–05.

Lafontaine, F., Saattvic, S., & Slade, M. (2023). No-poaching clauses in franchise contracts, anticompetitive or efficiency enhancing?. Available at SSRN 4404155.

Lebow, D. E., Saks, R. E., & Wilson, B. A. (2003). Downward nominal wage rigidity: Evidence from the employment cost index. *Advances in Macroeconomics, 3*(1), Article 2.

Lipsitz, M., & Starr, E. (2022). Low-wage workers and the enforceability of noncompete agreements. *Management Science, 68*(1), 143–170.

Manning, A. (2021). Monopsony in labor markets: A review. *ILR Review, 74*(1), 3–26.

Marinescu, O. & Rathelot, R. (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics, 10*(3), 42–70.

Martins, P. S., & Thomas, J. P. (2023). Employer collusion and employee training. Working paper, University of Edinburgh.

Matsushima, H. (2001). Multimarket contact, imperfect monitoring, and implicit collusion. *Journal of Economic Theory, 98*(1), 158–178.

Motta, M., & Polo, M. (2003). Leniency programs and cartel prosecution. *International Journal of Industrial Organization, 21*(3), 347–379.

Mukherjee, A., & Vasconcelos, L. (2012). Star wars: Exclusive talent and collusive outcomes in labor markets. *The Journal of Law, Economics, & Organization, 28*(4), 754–782.

Nocke, V., & White, L. (2007). Do vertical mergers facilitate upstream collusion?. *The American Economic Review, 97*(4), 1321–1339.

Normann, H. T., Rösch, J., & Schultz, L. M. (2015). Do buyer groups facilitate collusion?. *Journal of Economic Behavior & Organization, 109*, 72–84.

Piccolo, S., & Miklós-Thal, J. (2012). Colluding through suppliers. *The RAND Journal of Economics, 43*(3), 492–513.

Posner, E. A. (2021). *How antitrust failed workers.* Oxford University Press.

Prager, E., & Schmitt, M. (2021). Employer consolidation and wages: Evidence from hospitals. *The American Economic Review, 111*(2), 397–427.

Rasmusen, E. B., Ramseyer, J. M., & Wiley Jr, J. S. (1991). Naked exclusion. *The American Economic Review, 81*(5), 1137–1145.

Ross, T. W. (1992). Cartel stability and product differentiation. *International Journal of Industrial Organization, 10*(1), 1–13.

Sharma, G. (2024). Collusion among employers in India. Working Paper, available at: `https://garimasharma.com/files/collusion_gs_july2024.pdf`

Shi, L. (2023). Optimal regulation of noncompete contracts. *Econometrica, 91*(2), 425–463.

Spagnolo, G. (1999). On interdependent supergames: multimarket contact, concavity, and collusion. *Journal of Economic Theory, 89*(1), 127–139.

Starr, E. P., Prescott, J. J., & Bishara, N. D. (2021). Noncompete agreements in the US labor force. *The Journal of Law and Economics, 64*(1), 53–84.

Tirole, J. (2023). Socially responsible agencies. *Competition Law & Policy Debate, 7*(4), 171–177.

Tong, J., & Ornaghi, C. (2021). Joint oligopoly-oligopsony model with wage markdown power. University of Southampton, Discussion Papers in Economics and Econometrics No. 2101.

Tortarolo, D., & Zarate, R. D. (2018). Measuring imperfect competition in product and labor markets. An empirical analysis using firm-level production data. Development Bank of Latin America Working Paper No. 2018/03.

Vives, X. (1999). *Oligopoly pricing: Old ideas and new tools.* MIT Press.

Yeh, C., Macaluso, C., & Hershbein, B. (2022). Monopsony in the US labor market. *The American Economic Review, 112*(7), 2099–2138.

Appendix A: Proofs

Technical assumptions. To guarantee firm viability, assume firms sell a positive quantity when pricing at the optimized average cost. This assumption always holds if $Q(c_0) > 0$, with c_0 defined in Eqn. (3), and $\underline{\xi} \geq 0$.

Moreover, I assume that all profit functions considered in the maximization problems are globally concave and admit an interior maximum point, which is therefore obtained from the first-order condition(s). Sufficient conditions for these properties are given by $Q(p) \to 0$ as $p \to \infty$, $Q''(p) \leq 0$, and $\partial^2 L(\cdot)/\partial w_i^2$, $\partial^2 L(\cdot)/(\partial w_i \partial w)$, $\partial^2 L(\cdot)/\partial w^2 \leq 0$.

Finally, using the notation in Section 3, Assumption (A) can be written as

$$\frac{Q(p^M)}{|Q'(p^M)|} \geq r \int_{L(w^M,w^M)}^{nL(w^M,w^M)} \left| \frac{\partial^2 K(\tilde{\ell}, Q(p^M))}{\partial \ell \partial q} \right| \mathrm{d}\tilde{\ell}.$$

Proof of Proposition 0. The results easily follow from the arguments given in the text. □

Proof of Lemma 1. For any rivals' offers $\{w_{i',j}\}_{i' \in \mathcal{N}\setminus\{i\}, j \in [0,J]}$, fix any ℓ_i that firm i may want to hire (on- or off-path). Suppose firm i offers the same wage w_i to all workers $j \in [0, J]$. Then, from Eqn. (2) it follows that w_i must be set such that

$$\ell_i = J\,\mathbb{E}[\Pr[w_i + \xi_i \geq \max\{w_0, \max_{i' \in \mathcal{N}\setminus\{i\}} w_{i'} + \xi_{i'}\}]],$$

the expectation being taken with respect to the vector $\{\xi_i\}_{i \in \mathcal{N}}$ and, eventually, as workers are anonymous, the distribution of rivals' wage offers $w_{i',j}$.

To show that offering a uniform wage is the minimal-cost way to hire ℓ_i workers, suppose by contradiction that firm i makes two distinct offers, $\underline{w}_i$ to a mass J' of workers and $\overline{w}_i > \underline{w}_i$ to the other workers. Then, $\underline{w}_i$ and $\overline{w}_i$ must be such that

$$\ell_i = J'\,\mathbb{E}[\Pr[\underline{w}_i + \xi_i \geq \max\{w_0, \max_{i' \in \mathcal{N}\setminus\{i\}} w_{i'} + \xi_{i'}\}]] + (J - J')\mathbb{E}[\Pr[\overline{w}_i + \xi_i \geq \max\{w_0, \max_{i' \in \mathcal{N}\setminus\{i\}} w_{i'} + \xi_{i'}\}]],$$

where, workers being anonymous, the distribution of rivals' offers is the same across the two groups of workers. As, for all ξ_i and $w_{i'}$,

$$\Pr[\underline{w}_i + \xi_i \geq \max\{w_0, \max_{i' \in \mathcal{N}\setminus\{i\}} w_{i'} + \xi_{i'}\}] < \Pr[\overline{w}_i + \xi_i \geq \max\{w_0, \max_{i' \in \mathcal{N}\setminus\{i\}} w_{i'} + \xi_{i'}\}],$$

the deviation increases the average wage paid to the recruited workers. This argument generalizes to an arbitrary number of different offers. As choosing $w_{i,j} = -\infty$ is equivalent not to approach worker j, this suffices to show that offering uniform wages is a dominant strategy in the static game.

Then, each firm's labor supply function is obtained from Eqn. (2) for $w_{i,j} \equiv w_i$ for all i. As all $\xi_{i,j}$ are drawn from the same distribution Ξ, which is assumed smooth, firms' labor supply is symmetric and almost everywhere differentiable. Then, inspection of Eqn. (2) immediately implies that, in a situation where all i's rivals offer the same wage w to all workers, i's labor supply, denoted by $L(w_i, w)$, is increasing in w_i and decreasing in w. Finally, for $w_i = w$, aggregate labor supply writes as

$$\sum_{i \in \mathcal{N}} L(w, w) = J \Pr[w + \max_{i \in \mathcal{N}} \xi_i \geq w_0],$$

which is clearly increasing in w. □

Proof of Lemma 2. Given the wage w offered (to all workers) by all its rivals, firm i's cost-minimization problem for any given quantity q to produce can be written as

$$\begin{aligned} &\min_{\ell_i, k_i} W(\ell_i, w)\ell_i + r\, k_i \\ &\text{s.t. } q = F(\ell_i, k_i), \end{aligned}$$

where

$$W(\ell_i, w) \equiv L^{-1, w_i}(w_i, w),$$

is firm i's residual (inverse) labor supply, with $L^{-1,w_i}(\cdot)$ being the inverse of the labor supply function $L(\cdot)$ with respect to w_i. As $L(\cdot)$ is strictly increasing in w_i, this inverse function is well defined and is increasing in ℓ_i. Then, by the steps in Gelles and Mitchell (1996), given that $F(\cdot)$ is CRS, at the optimum, the elasticity of the average cost $C(q; w)$, defined in Eqn. (5), with respect to q equals

$$\frac{1}{qC(\cdot)} \ell_i^2 \frac{\partial W_i(\cdot)}{\partial \ell_i} > 0,$$

which implies that the optimized average cost is an increasing function of q. □

Proof of Lemma 3. For any candidate symmetric equilibrium price p and wage w offered by rivals, each firm i expects to sell $q_i = Q(p)/n$ and chooses w_i solving

$$\min_{w_i} w_i L(w_i, w) + r\, K\left(L(w_i, w), \frac{Q(p)}{n}\right).$$

Taking the first-order condition yields

$$L(w_i, w) + \left[w_i + r \frac{\partial K(L(w_i, w), Q(p)/n)}{\partial \ell}\right] \frac{\partial L(w_i, w)}{\partial w_i} = 0. \tag{13}$$

Any symmetric equilibrium is then obtained from Eqn. (13) imposing symmetry – i.e., $w_i = w$. At $w \to -\infty$, its left-hand side is negative, as $L(-\infty, -\infty) = 0$ and the term in square brackets is strictly negative and is multiplied by $\partial L(\cdot)/\partial w_i > 0$. Conversely, for $w \to +\infty$, this left-hand side is clearly positive.

Therefore, the wage-offer game (given p) admits a unique symmetric equilibrium $w_i = W^*(p)$ for all $i \in \mathcal{N}$ if the derivative of the left-hand side of Eqn. (13) for $w_i = w$ with respect to w is positive. This boils down to the following condition:

$$\frac{\partial L(\cdot)}{\partial w_i} + \left[1 + r\frac{\partial L(\cdot)}{\partial w_i}\frac{\partial^2 K(\cdot)}{\partial \ell^2}\right]\left(\frac{\partial L(\cdot)}{\partial w_i} + \frac{\partial L(\cdot)}{\partial w}\right) - \frac{L(\cdot)}{\frac{\partial L(\cdot)}{\partial w_i}}\left(\frac{\partial L^2(\cdot)}{\partial w_i^2} + \frac{\partial L^2(\cdot)}{\partial w_i \partial w}\right) > 0,$$

which, as $[\partial L(\cdot)/\partial w_i + \partial L(\cdot)/\partial w]|_{w_i=w} > 0$, always holds for $\partial L^2(\cdot)/\partial w_i^2 + \partial L^2(\cdot)/(\partial w_i \partial w) \leq 0$.

Finally, the derivative of the left-hand side of Eqn. (13) with respect to p equals

$$r \underbrace{\frac{\partial L(\cdot)}{\partial w_i}}_{+} \underbrace{\frac{\partial^2 K(\cdot)}{\partial \ell \partial q}}_{-} \underbrace{\frac{Q'(\cdot)}{n}}_{-} > 0,$$

which, by the Implicit Function Theorem and the second-order condition of firms' cost-minimization problem, shows that $W^*(p)$ is decreasing in p. □

Proof of Proposition 1. As $\pi(0) < 0 < \pi(p^P)$,[53] and, by concavity, $\pi(p)$ is strictly increasing in p for all $p < p^P$, there exists a price level $\underline{p}^N \in (0, p^P)$ such that $\pi(\underline{p}^N) = 0$. As firms can always overcut any candidate equilibrium price and make zero profits, any candidate equilibrium price is such that $p \geq \underline{p}^N$. Then, for all $p \leq p^P$, the best deviation for a firm always consists in slightly undercutting the price, yielding the deviation profit $\pi^D(p)$ in Eqn. (7).

The results of Lemma 2, which apply here as firms are best-responding to rivals' wages both on- and off-path, imply that holding fixed rivals' wage offers at $W^*(p)$, each firm faces a convex cost function. Then, by the same steps as in Dastidar (1995), it follows that the difference $\pi^D(p) - \pi(p)$ is negative at $p = \underline{p}^N$ and is increasing in p, and so that there is a price $\overline{p}^N \in (\underline{p}^N, p^P]$ such that $\pi(p) \geq \pi^D(p)$ for all $p \leq \overline{p}^N$. Hence, all prices $p \in [\underline{p}^N, \overline{p}^N]$, with corresponding wage levels $w = W^*(p)$, constitute NEs of the static game. As $\pi(p)$ is strictly increasing in p for all $p < p^P$, equilibria with higher prices are associated with a larger profit. □

[53]Note that *(i)* $\pi(p) \to 0$ as $p \to \infty$ (as $Q(p) \to 0$ and so firms are *de facto* inactive, i.e. also $W^*(p) \to 0$), and *(ii)* as firms sell a positive quantity when pricing at the optimized average cost, they can obtain positive profits by raising their price. Hence, p^P is finite and is such that $\pi(p^P) > 0$.

Proof of Corollary 1. Since optimal punishments profits cannot be negative because firms always have the option to shut down, and the simple perfect equilibrium consisting of the repeated static NE profile $(\underline{p}^N, W^*(\underline{p}^N))$ yields discounted profits of zero to all firms, by the results in Abreu (1988), restricting attention to SPNE in grim-trigger strategies specifying reversion to $(\underline{p}^N, W^*(\underline{p}^N))$ is without loss of generality.

For $\delta = 0$, in any SPNE firms must play a static NE at each period. Hence, the most profitable SPNE is the repetition of $(\overline{p}^N, W^*(\overline{p}^N))$. □

Proof of Proposition 2. As seen in Lemma 1, uniform wage offers ($w_{i,j} \equiv w_i$ for all $j \in [0, J]$) minimize production costs. Then, industry cost minimization requires each worker to be hired by the firm providing the highest non-wage benefit, which is achieved by all firms offering the same wage level w. As a result, industry profits' maximization requires that all firms are active and equally split the production of the monopoly output – i.e., charge the same price p to consumers and offer the same wage w to workers.

Maximizing industry profits is then equivalent to maximize $\pi(p, w)$ in Eqn. (8) with respect to p and w. This yields the first-order conditions

$$Q(p) + \left[p - r\frac{\partial K(L(w,w), Q(p)/n)}{\partial q}\right] Q'(p) = 0, \tag{14}$$

and

$$L(w,w) + \left[w + r\frac{\partial K(L(w,w), Q(p)/n)}{\partial \ell}\right] \left(\frac{\partial L(w_i, w)}{\partial w_i} + \frac{\partial L(w_i, w)}{\partial w}\right)\bigg|_{w_i = w} = 0, \tag{15}$$

respectively.

Defining $\widehat{P}(w)$ as the solution to Eqn. (14) for fixed w, by the Implicit Function Theorem,

$$\frac{\partial \widehat{P}(w)}{\partial w} = -\frac{-rQ'(\cdot)\frac{\partial^2 K(\cdot)}{\partial \ell \partial q}\left(\frac{\partial L(\cdot)}{\partial w_i} + \frac{\partial L(\cdot)}{\partial w}\right)}{\frac{\partial^2 \pi(\cdot)}{\partial p^2}} < 0,$$

as $\partial^2 \pi(\cdot)/\partial p^2 < 0$ by the second-order condition of the cartel's problem. Similarly, defining $\widehat{W}(p)$ as the solution to Eqn. (15) for fixed p, the Implicit Function Theorem implies that

$$\frac{\partial \widehat{W}(p)}{\partial p} = -\frac{\frac{r}{n}\frac{\partial^2 K(\cdot)}{\partial \ell \partial q}\left(\frac{\partial L(\cdot)}{\partial w_i} + \frac{\partial L(\cdot)}{\partial w}\right)}{\frac{\partial^2 \pi(\cdot)}{\partial w^2}} < 0,$$

as $\partial^2 \pi(\cdot)/\partial w^2 < 0$ by the second-order condition of the cartel's problem.

Comparing Eqn. (15) with Eqn. (13) at a symmetric solution implies that holding fixed the price p, $\widehat{W}(p) < W^*(p)$: This result follows from the second-order conditions because the left-hand side of Eqn. (15) contains a positive extra-term, as $w + r\partial K(\cdot)/\partial \ell < 0$ for the two equations to hold, and $\partial L(\cdot)/\partial w < 0$.

If $p^M > \overline{p}^N$, as $\widehat{W}(p)$ is decreasing in p, this would imply that $w^M < W^*(\overline{p}^N) \equiv \underline{w}^N$. Therefore, given that $\widehat{P}(w)$ is decreasing in w, and it is an upper bound on the cartel price given w (even if $\widehat{P}(w) < P^*(w)$, with $P^*(\cdot)$ defined by Eqn. (11), the cartel, coordinating on the most profitable equilibrium, would still set $\widehat{P}(w)$), one can conclude that indeed $p^M > \overline{p}^N$ and $w^M < \underline{w}^N$.

In order for (p^M, w^M) to be sustainable as a SPNE, it must be that

$$\delta \geq \delta^M \equiv 1 - \frac{\pi(p^M, w^M)}{\pi^D(p^M, w^M)},$$

where $\pi(\cdot)$ and $\pi^D(\cdot)$ are defined in Eqn. (8) and (9), respectively (the result that the profit in Eqn. (9) is the highest one attainable in the deviation period is established in the Proof of Proposition 3 below, and firms' profits are zero afterwards by Corollary 1).

Therefore, $\delta^M < \delta_0^M$ if and only if $\pi^D(p^M, w^M) < n\pi(p^M, w^M)$. Let $\widehat{\pi}^M$ denote the highest attainable profit in the hypothetical scenario where only one firm is active. As firm i's rivals being inactive is equivalent to $w_{i'} = -\infty$ for all $i' \in \mathcal{N} \setminus \{i\}$,

$$\widehat{\pi}^M = \max_{p,w} pQ(p) - [wL(w, -\infty) + r\,K\,(L(w, -\infty), Q(p))]\,.$$

Then, as $L(w, -\infty) > L(w, w^M)$ (because $\partial L(\cdot)/\partial w < 0$), it immediately follows that

$$\widehat{\pi}^M > \max_{p,w} pQ(p) - \left[wL(w, w^M) + r\,K\left(L(w, w^M), Q(p)\right)\right] = \pi^D(p^M, w^M).$$

Next, as, by definition, $\pi(p^M, w^M) = \max_{p,w} \pi(p, w)$, it holds that

$$n\pi(p^M, w^M) = \max_{p,w} pQ(p) - [nwL(w, w) + r\,K\,(nL(w, w), Q(p))] > \widehat{\pi}^M,$$

because employers' differentiation implies that

$$nL(w, w) = J\,\Pr[w + \max_{i \in \mathcal{N}} \xi_i \geq w_0] > J\,\Pr[w + \xi_i \geq w_0] = L(w, -\infty).$$

Summing up, $\pi^D(p^M, w^M) < \widehat{\pi}^M < n\pi(p^M, w^M)$, and so $\delta^M < \delta_0^M$. □

Proof of Proposition 3. First, I prove that, given a candidate most profitable SPNE (p, w) and the corresponding per-period profit given in Eqn. (8), the highest deviation profit is the one given in Eqn. (9). To this end, I first establish that, supposing that a deviating firm wants to deviate in price, its best deviation always consists in slightly undercutting p. As $\pi(p, w) > 0$ (this holds at the most profitable static NE, hence *a fortiori* it shall be true at the most profitable SPNE for all $\delta > 0$), and a firm i would make no sales for any $p_i > p$, one can restrict attention to $p_i \leq p$, with the convention that all consumers' demand $Q(p)$ is served by the deviating firm i even if $p_i = p$ (i.e.,

$p_i = p$ corresponds to a slight undercutting of p). The deviating firm solves

$$\max_{p_i \leq p, w_i} p_i Q(p_i) - [w_i L(w_i, w) + r\, K\,(L(w_i, w), Q(p_i))]\,,$$

and, hence, $p_i = p$ is the best deviation price if and only if

$$Q(p) + \left[p - r\frac{\partial K(L(w_i, w), Q(p))}{\partial q}\right] Q'(p) \geq 0.$$

As the collusive price p is weakly below the monopoly price given $w_i = w$, denoted by $\widehat{P}(w)$ (any higher price would be suboptimal), i's profit is increasing in p_i and so this inequality always holds at $(L(w, w), Q(p)/n)$, or equivalently (given that $\partial K(\cdot)/\partial q$ is homogeneous of degree zero) at $(nL(w, w), Q(p))$. As $\partial^2 K(\cdot)/\partial \ell \partial q < 0$, it holds *a fortiori* in the deviation if $nL(w, w) > L(w_i, w)$. This inequality always holds: As colluding firms at most fully internalize the negative cross-firm wage externalities, $nL(w, w)$ is weakly larger than the overall labor force needed to produce $Q(p)$ at the minimum cost for the industry; hence, a deviating firm who wants to produce $Q(p)$ alone will optimally hire at most $nL(w, w)$ workers; however, given that it needs to raise the wage to attract workers hired by rivals in equilibrium, it follows that $nL(w, w) > L(w_i, w)$ at the optimal deviation.

Then, the highest deviation profit is given by

$$\tilde{\pi}^D(p, w) \equiv \max_{q_i \in \{Q(p)/n, Q(p)\}, w_i} pq_i - [w_i L(w_i, w) + r\, K\,(L(w_i, w), q_i)]\,.$$

Fix a price level p. If $w = W^*(p)$, firm i is already offering its static profit maximizing wage level to produce $q_i = Q(p)/n$, and so, for all $p > \overline{p}^N$, only choosing $q_i = Q(p)$ strictly increases the deviating firm's profit above $\pi(p, w)$. Colluding firms can, however, set $w < W^*(p)$, which increases the deviation profit from choosing $q_i = Q(p)/n$ less than it increases the deviation profit from choosing $q_i = Q(p)$. Indeed, by the Envelope Theorem,

$$\frac{\partial \tilde{\pi}^D(\cdot)}{\partial w} = L(w_i, w)\frac{\frac{\partial L(w_i, w)}{\partial w}}{\frac{\partial L(w_i, w)}{\partial w_i}},$$

is higher, in absolute value, the higher the deviating firm's w_i, which in turn is higher when it produces more. As a result, any reduction in w below $W^*(p)$ further increases the relative profitability for a deviating firm of choosing $q_i = Q(p)$ rather than $q_i = Q(p)/n$. Therefore, the highest deviation profit is obtained for $q_i = Q(p)$ and is given by Eqn. (9).

Next, to obtain the most profitable SPNE for any $\delta \in [0, \delta^M]$, I first show that the ratio $\pi(p, w)/\pi^D(p, w)$ is decreasing in p and increasing in w at any candidate most profitable

SPNE. As, for all $p \leq \widehat{P}(w)$, $\partial\pi(\cdot)/\partial p \geq 0$ and, by the arguments above,

$$\frac{\partial\pi^D(\cdot)}{\partial p} = Q(p) + \left[p - r\frac{\partial K(L(w_i, w), Q(p))}{\partial q}\right] Q'(p) > 0,$$

it holds that

$$\frac{\partial}{\partial p}\left(\frac{\pi(\cdot)}{\pi^D(\cdot)}\right) < 0 \iff \frac{\pi(\cdot)}{\pi^D(\cdot)} > \frac{\partial\pi(\cdot)/\partial p}{\partial\pi^D(\cdot)/\partial p},$$

which, using the fact that $K(\cdot)$ is homogeneous of degree one, and so $K(\ell, q) = \ell\,\partial K(\cdot)/\partial\ell + q\,\partial K(\cdot)/\partial q$, can be written (omitting arguments) as

$$\frac{\frac{1}{n}\left[\left(p - r\frac{\partial K}{\partial q}\right)Q - \left(w + r\frac{\partial K}{\partial \ell}\right)L\right]}{\left(p - r\frac{\partial K^D}{\partial q}\right)Q - \left(w^D + r\frac{\partial K^D}{\partial \ell}\right)L^D} > \frac{\frac{1}{n}\left[Q + \left(p - r\frac{\partial K}{\partial q}\right)Q'\right]}{Q + \left(p - r\frac{\partial K^D}{\partial q}\right)Q'},$$

where w^D is the best deviation wage, $L^D \equiv L(w^D, w)$, and $K^D \equiv K(L^D, Q(p))$, whilst L and K denote the corresponding values in the candidate SPNE outcome. The two sides of this inequality are equal if $w^D = w$ and $L^D = nL$. Yet, as argued above, because $w^D > w$, the deviating firm sets $L^D < nL$. By a revealed preference argument, $\pi^D(\cdot)$ is then smaller, hence the left-hand side is larger; conversely, the denominator of the right-hand side is larger (as its derivative with respect to L^D equals $-rQ' \cdot \partial^2 K^D/(\partial\ell\partial q) < 0$), so that the right-hand side is smaller. As a result, the inequality is always satisfied.

Similarly, as, for all $w \geq \widehat{W}(p)$, $\partial\pi(\cdot)/\partial w \leq 0$ and, as seen above, $\partial\pi^D(\cdot)/\partial w = L^D(\partial L^D/\partial w)/(\partial L^D/\partial w_i) < 0$, it holds that

$$\frac{\partial}{\partial w}\left(\frac{\pi(\cdot)}{\pi^D(\cdot)}\right) > 0 \iff \frac{\pi(\cdot)}{\pi^D(\cdot)} > \frac{\partial\pi(\cdot)/\partial w}{\partial\pi^D(\cdot)/\partial w}.$$

As colluding firm always set $w \leq W^*(p)$,

$$\frac{\partial\pi(\cdot)}{\partial w} > L\frac{\partial L/\partial w}{\partial L/\partial w_i},$$

and so a sufficient condition for the above inequality to hold is $\pi(\cdot)/\pi^D(\cdot) > L/L^D$, or, equivalently,

$$\frac{\pi^D(\cdot)}{L^D} < \frac{\pi(\cdot)}{L}.$$

This inequality always holds because a deviation, entailing an increase in the wage to be paid to attract more workers, unambiguously reduces the profit per worker (i.e., the deviating firm is less efficient than the whole industry in producing the same output $Q(p)$).

Therefore, the colluding firms would always like to raise p (holding fixed w) or reduce w (holding fixed p), but cannot do so when $\delta < 1 - \pi(\cdot)/\pi^D(\cdot)$ and so the incentive-compatibility constraint is violated. It follows that the incentive-compatibility constraint is binding for all $\delta < \delta^M$, and colluding firms optimally exploit any increase in δ to both raise the price and reduce the wage offers. By the Implicit Function Theorem, the results that $\pi(\cdot)/\pi^D(\cdot)$ is decreasing in p and increasing in w implies that the incentive-compatibility constraint defines an upward-sloping function $P^{\mathrm{IC}}(w)$ (resp., $W^{\mathrm{IC}}(p)$) in the (w,p)-plane (resp., (p,w)-plane), and the colluding firms choose the point on this function that maximizes $\pi(p,w)$. In particular, the Lagrangian of Problem (P) writes as

$$\mathcal{L}(\cdot) = \pi(p,w) - \lambda \left[\delta - 1 + \frac{\pi(p,w)}{\pi^D(p,w)}\right],$$

with λ denoting the Lagrange multiplier associated to the incentive-compatibility constraint. Taking the first-order condition with respect to p and w yields, respectively,[54]

$$\frac{\partial \pi(\cdot)}{\partial p} = \frac{\lambda}{[\pi^D(\cdot)]^2}\left[\frac{\partial \pi(\cdot)}{\partial p}\pi^D(\cdot) - \frac{\partial \pi^D(\cdot)}{\partial p}\pi(\cdot)\right],$$

and

$$\frac{\partial \pi(\cdot)}{\partial w} = \frac{\lambda}{[\pi^D(\cdot)]^2}\left[\frac{\partial \pi(\cdot)}{\partial w}\pi^D(\cdot) - \frac{\partial \pi^D(\cdot)}{\partial p}\pi(\cdot)\right].$$

Diving these equations and rearranging yields the optimality condition (10). □

Proof of Proposition 4. Maximizing $\pi(p, W^*(p))$ w.r.t. p yields the first-order condition

$$Q(\cdot) + Q'(\cdot)\left[p - r\frac{\partial K(\cdot)}{\partial q}\right] + n\frac{\partial L(\cdot)/\partial w}{\partial L(\cdot)/\partial w_i}L(\cdot)\frac{\partial W^*(\cdot)}{\partial p} = 0,$$

which, as $\partial W^*(p)/\partial p < 0$, contains a positive extra-term relative to the first-order condition (14) of the multimarket cartel: hence, the price cartel has incentives to set a higher price holding fixed w. Together with $W^*(p) > \widehat{W}(p)$, this implies that $p^P > p^M$ and $W^*(p^P) > w^M$.

Under Assumption (A), $p^P > \overline{p}^N$ is such that $\pi(p^P, W^*(p^P)) < \pi^D(p^P, W^*(p^P))$, implying that the critical discount factor to implement the price cartel outcome is

$$\delta^P = 1 - \frac{\pi(p^P, W^*(p^P))}{\pi^D(p^P, W^*(p^P))} > 0.$$

Indeed, as $w^M < W^*(p^M)$, also $L(w^M, w^M) < L(W^*(p^M), W^*(p^M))$, and so

$$p^M \geq r\frac{\partial K(L(w^M, w^M), Q(p^M))}{\partial q} > r\frac{\partial K(L(W^*(p^M), W^*(p^M)), Q(p^M))}{\partial q}.$$

Therefore, Assumption (A), which is equivalent to the first inequality above, implies that

[54] The functions $\pi(p,w)$ and $\pi^D(p,w)$ defined in Eqn. (8) and (9) are continuously differentiable. Hence, the following expressions are always well-defined.

a firm is willing to undercut p^M even though it must stick to $w = W^*(p^M)$, implying that *a fortiori* $\pi(p^M, W^*(p^M)) < \pi^D(p^M, W^*(p^M))$. Then, as the difference $\pi^D(p, W^*(p)) - \pi(p, W^*(p))$ is increasing in p (see the Proof of Proposition 1), and, as seen above, $p^P > p^M$, one can conclude that $\pi(p^P, W^*(p^P)) < \pi^D(p^P, W^*(p^P))$.[55]

Next, the result $\delta^P < \delta_0^M$ follows from Lemma 2. Indeed, offering $w_i = W^*(p)$ when rivals do the same implies that firm i produces at the minimized average cost function $C(Q(p)/n; W^*(p))$. Then, as rivals still offer $w = W^*(p)$ when i deviates, the minimized average cost to produce $Q(p)$ is larger: $C(Q(p); W^*(p)) > C(Q(p)/n, W^*(p))$, from which one can conclude that, at $p = p^P$,

$$\pi^D(p^P, W^*(p^P)) = [p^P - C(Q(p^P); W^*(p^P))]Q(p^P) <$$

$$< [p^P - C(Q(p^P)/n; W^*(p^P))]Q(p^P) = n\pi(p^P, W^*(p^P)),$$

which implies that $\delta^P < (n-1)/n = \delta_0^M$.

Next, consider values of $\delta > 0$ such that both $P^P(\delta) < p^P$ and $P^M(\delta) < p^M$ hold – i.e., the incentive-compatibility constraint binds under both price collusion and multimarket collusion. For all such values of δ, the multimarket collusion outcome $(P^M(\delta), W^M(\delta))$ violates the wage constraint $w = W^*(p)$. Indeed, at $w = W^*(p)$, the inequality

$$\frac{\partial\pi(p, W^*(p))/\partial p}{\partial\pi^D(p, W^*(p))/\partial p} < \frac{\partial\pi(p, W^*(p))/\partial w}{\partial\pi^D(p, W^*(p))/\partial w}$$

simplifies, using the notation introduced in the proof of Proposition 3 above, as

$$\frac{\frac{1}{n}\left[Q + \left(p - r\frac{\partial K}{\partial q}\right)Q'\right]}{Q + \left(p - r\frac{\partial K^D}{\partial q}\right)Q'} < \frac{-L\frac{\partial L(\cdot)}{\partial w}\frac{\partial L(\cdot)}{\partial w_i}}{-L^D\frac{\partial L^D}{\partial w}\frac{\partial L^D}{\partial w_i}}.$$

The two sides of this inequality are equal when $w^D = w$ and $L^D = nL$; then, $w^D > w_i$ implies that $L^D < nL$ (see the Proof of Proposition 2), which reduces the denominator of the left-hand side and increases the denominator of the right-hand side, implying that this inequality always holds.

Therefore, as the optimality condition (10) is violated under the wage constraint $w = W^*(p)$, it follows that, under price collusion, this constraint must bind for all $\delta > 0$. Then, as under multimarket collusion firms would never set $w > W^*(p)$ and the binding incentive-compatibility constraint defines an upward-sloping function in the (p, w)-plane, it follows that, for any given δ, the price collusion outcome is such that $P^P(\delta) > P^M(\delta)$ and $W^P(\delta) = W^*(P^P(\delta)) > W^M(\delta)$. Still, as any increase in δ entails an outward shift of the locus defined by the incentive-compatibility constraint and firms' profits increase

[55] This implies that $\overline{p}^N < p^P$ is such that $\pi(\overline{p}^N, W^*(\overline{p}^N)) = \pi^D(\overline{p}^N, W^*(\overline{p}^N))$, and so $(\overline{p}^N, W^*(\overline{p}^N))$ is the most profitable SPNE under price collusion for $\delta = 0$.

in p for all $p \leq \widehat{P}(w)$, $P^P(\delta)$ is increasing in δ, and so $W^P(\delta) = W^*(P^P(\delta))$ is decreasing in δ.

Therefore, if the value $P^P(\delta^M)$, obtained by the intersection between the upward-sloping function $W^{\text{IC}}(p)$ defined by the incentive-compatibility constraint at δ^M and the downward-sloping function $W^*(p)$, is smaller than p^P, then $\delta^P > \delta^M$ (in this case, $P^P(\delta)$ keeps increasing, and $W^P(\delta) = W^{\text{IC}}(P^P(\delta))$ keeps decreasing for all $\delta \in [\delta^M, \delta^P]$, where the multimarket cartel outcome can be sustained); else, $\delta^P < \delta^M$ (in this case, $P^M(\delta)$ keeps increasing, and $W^M(\delta)$ keeps decreasing for all $\delta \in [\delta^P, \delta^M]$, where the price cartel outcome can be sustained). □

Proof of Proposition 5. The constraint $p \leq P^*(w)$ is obtained by imposing that a firm does not gain by undercutting p holding fixed its wage offer at w – i.e.,

$$\frac{n-1}{n} pQ(p) - r\left[K(L(w,w), Q(p)) - K(L(w,w), Q(p)/n)\right] \leq 0. \tag{16}$$

To see that this inequality defines an upper bound on p, note that its left-hand side is negative at $p = 0$ and is increasing in p: its derivative with respect to p equals

$$\frac{n-1}{n}\left[Q(p) + pQ'(p)\right] - r\left[\frac{\partial K(L(w,w), Q(p))}{\partial q} - \frac{1}{n}\frac{\partial K(L(w,w), Q(p)/n)}{\partial q}\right],$$

this value being positive as $\partial K(L(w,w), Q(p))/\partial q > \partial K(L(w,w), Q(p)/n)/\partial q$ and $Q(p) + pQ'(p) - r\,\partial K(L(w,w), Q(p)/n)/\partial q \geq 0$ for all $p \leq \widehat{P}(w)$. Therefore, $P^*(w)$ is the value that satisfies (16) with equality. Moreover, $P^*(w)$ is a decreasing function of w. This result follows from the Implicit Function Theorem and the fact that the left-hand side of inequality (16) is increasing in p, as seen above, and in w: Indeed, its derivative with respect to w equals

$$-r\left[\frac{\partial K(L(w,w), Q(p))}{\partial \ell} - \frac{\partial K(L(w,w), Q(p)/n)}{\partial \ell}\right]\left(\frac{\partial L(w,w)}{\partial w_i} + \frac{\partial L(w,w)}{\partial w}\right) > 0.$$

Next, the most profitable static NE $(\overline{p}^N, \underline{w}^N)$ is the most collusive SPNE outcome under wage collusion only for $\delta = 0$: Under Assumption (A), it is such that $\pi(\cdot) = \pi^D(\cdot)$, so any different wage level would imply that firms have incentives to deviate in wage offers holding fixed their price. As argued in the text, $\overline{p}^N < P^*(\underline{w}^N)$, because a firm i is indifferent between undercutting or not $p = \overline{p}^N$ when it can optimally expand its labor force, which implies that it is strictly better off by not undercutting when it cannot choose $w_i > \underline{w}^N$. By a continuity argument, the price constraint $p \leq P^*(w)$ does not bind for sufficiently low values of δ, so that wage collusion can replicate the multimarket collusion outcome.

Conversely, Assumption (A) implies that, at the multimarket cartel outcome, any firm would have strict incentives to undercut $p = p^M$ even though it cannot raise its wage, or, equivalently, that $p^M = \widehat{P}(w^M) > P^*(w^M)$. Therefore, for all $\delta \geq \delta^M$, the price constraint binds, implying that under wage collusion, firms have to move down along the upward-sloping function in the (p, w)-plane defined by the incentive-compatibility constraint for $\delta = \delta^M$ by choosing lower price and wage levels such that the binding price constraint is satisfied. Therefore, by a continuity argument, $W^W(\delta) < W^M(\delta)$ and $P^W(\delta) = P^*(W^W(\delta)) < P^M(\delta)$ for all δ sufficiently large.

This argument also implies that the price constraint must bind at the wage cartel optimum. Maximizing $\pi(P^*(w), w)$ yields the first-order condition

$$\left[L(\cdot) + \left(w + r\frac{\partial K(\cdot)}{\partial \ell}\right)\left(\frac{\partial L(\cdot)}{\partial w_i} + \frac{\partial L(\cdot)}{\partial w}\right)\right] - \frac{1}{n}\left[Q(\cdot) + \left(p - r\frac{\partial K(\cdot)}{\partial q}\right)Q'(\cdot)\right]\frac{\partial P^*(\cdot)}{\partial w} = 0,$$

which, as $\partial P^*(w)/\partial w < 0$, contains a positive extra-term relative to the first-order condition (15) of the multimarket cartel. Together with the fact that, under Assumption (A), $\widehat{P}(w) > P^*(w)$ at the optimum, by the corresponding second-order condition this implies $w^W < w^M$ and $p^W < w^M$.

Therefore, $\delta^W > \delta^M$ if the wage level $W^W(\delta)$ pinned down, as explained above, by the intersection (in the (p, w)-plane) between the upward-sloping locus of incentive-compatible points given $\delta = \delta^M$ and the price constraint $p = P^*(w)$, is such that $W^W(\delta) > w^W$; else, $\delta^W < \delta^M$.

In either case, $\delta^M < \delta_0^M$. This result follows from the same argument given in the Proof of Proposition 2: At $w = w^W$, the industry is producing $Q(p^W)$ at the minimal cost; a deviating firm that wants to produce the same quantity alone will always face higher costs, because *(i)* of employer differentiation ($nL(w^W, w^W) > L(w^W, -\infty)$), and *(ii)* of labor market competition by rivals ($L(w_i, -\infty) > L(w_i, w^W)\, \forall w_i$); as a result, $\pi^D(P^*(w^W), w^W) < n\pi(P^*(w^W), w^W)$, and so $\delta^W < (n-1)/n = \delta_0^M$. □

Proof of Corollary 2. The results in Proposition 3, 4, and 5 yield the inequalities in the statement. Then, the results immediately follow from consumer surplus being a decreasing function of the equilibrium price. □

Proof of Proposition 6. Take $\hat{\delta} \in (0, \delta^M)$ and consider the multimarket collusion outcome $(P^M(\hat{\delta}), W^M(\hat{\delta}))$, which is the point on the upward-sloping function $P^{\text{IC}}(w)$ defined by the binding incentive-compatibility constraint that maximizes the equilibrium profit, i.e. satisfies condition (10). Suppose that the regulator introduces a minimum wage $\underline{w} = W^M(\hat{\delta}) + \varepsilon$, with $\varepsilon > 0$ small enough, and denote by $(\underline{p}^M, \underline{w}^M)$ the most profitable SPNE under the wage constraint $w \geq \underline{w}$. Then, this constraint must bind: $\underline{w}^M(\delta) = \underline{w}$; moreover, as the regulatory minimum wage is still lower than the wage at which firms would like to ration workers, $L(\underline{w}, \underline{w}) > L(W^M(\hat{\delta}), W^M(\hat{\delta}))$. As the incentive-compatibility constraint is still binding, $\underline{p}^M = P^{\mathrm{IC}}(\underline{w}) > P^{\mathrm{IC}}(W^M(\hat{\delta})) = P^M(\hat{\delta})$. $\square$

Proof of Proposition 7. Irrespective of H, the cartel outcome in each product market is $(p^M, w^M) \equiv \arg\max_{p,w} \pi(p,w)$, with $\pi(p,w)$ given in Eqn. (8) as in the baseline model because each firm only sells to consumers in its geographic market. The critical discount factor to sustain (p^M, w^M) as a SPNE of the supergame in each product market is

$$\delta^M(H) = 1 - \frac{\pi(p^M, w^M)}{\pi^D(p^M, w^M)},$$

where

$$\pi^D(p^M, w^M) \equiv \max_{h \in \{1,\ldots,H\}} \left\{ hp^M Q(p^M) - \min_{w_i} \left[w_i L(w_i, w^M) + r\, K(L(w_i, w^M), hQ(p^M)) \right] \right\}.$$

This is because, as seen in the baseline model, a deviating firm always finds it optimal to undercut the cartel price and serve all consumers within its local market. Yet, it may be profitable to serve consumers in $h > 1$ markets, if possible (i.e., if $H > 1$). However, if $H \to \infty$, the maximand H^* in the above problem is always finite. The reason is that a deviating firm faces a globally increasing average cost function (this is true by the arguments of the baseline analysis whenever the dimension J of the local labor market, whose workers can be all attracted by paying w_0, is not binding, and even more so for higher production levels, where the deviating firm can expand production only by increasing its endowment of capital) and hence has a finite optimal level of production.

Therefore, $\delta^M(H)$ is strictly increasing in H for $H < H^*$, and constant at $\delta^M(H^*)$ for all $H \geq H^*$. Finally, $\delta^M(H^*) < (nH^* - 1)/(nH^*)$ because a deviating firm serving consumers in h markets always obtains $\pi^D(p^M, w^M) < hn\pi(p^M, w^M)$: This holds for $h = 1$ by Proposition 2, and *a fortiori* holds for $h > 1$ as the deviating firm serves consumers also in geographic markets from which it cannot attract any worker.

Finally, note that there is no need for *cross-market collusion* to sustain (p^M, w^M) within each geographic market. The reason is as follows. Suppose firms in market h expect this outcome to prevail in all other markets. Then, a single firm's best deviation is always more profitable than a joint deviation of all firms in market h from (p^M, w^M), given that a joint deviation, implying higher wage offers by all firms in local labor market h, makes it costlier to expand production for any individual firm. Hence, for all $\delta \geq \delta^M(H)$, this critical discount factor being pinned down by the individual-firm incentive-compatibility constraint defined above, all firms in market h optimally choose (p^M, w^M) even if they only collude among each other. $\square$

Proof of Proposition 8. In what follows, $\ell_{i,z} = L(w_{i,z}, w_z; w_{i',-z}, w_{-z})$ denotes the labor force hired by firm i in market z as function of the own wage offer, the wage offer of firm i' in the other product market $-z$ with which it can merge (as detailed above), and the symmetric offers made by any other firm in each of the two markets.

If all firms selling product z set the same price p_z and offer the same wage w_z to all workers, each firm in product market $z \in \{A, B\}$ makes a per-period profit

$$\pi(p_z, w_z; w_{-z}, w_{-z}) \equiv p_z \frac{Q(p_z)}{n} - \left[w_z L(w_z, w_z; w_{-z}, w_{-z}) + r\, K \left(L(w_z, w_z; w_{-z}, w_{-z}), \frac{Q(p_z)}{n} \right) \right].$$

With single-product firms, the highest deviation profit equals

$$\pi^D(p_z, w_z; w_{-z}, w_{-z}) \equiv p_z Q(p_z)/2 - \min_{w_i} \left[w_i L(w_i, w_z; w_{-z}, w_{-z}) + r\, K \left(L(w_i, w_z; w_{-z}, w_{-z}), Q(p_z)/2 \right) \right].$$

Comparing this deviation profit of a single-product firm with that of a multiproduct firm, denoted here by $\pi_m^D(p_z, w_z; w_{-z})$, at a symmetric SPNE candidate ($p_z \equiv p$, $w_z = w_{-z} \equiv w$, and also $w_{i',-z} = w$ with single-product firms), one has that $\pi_m^D(p, w; w) < 2\pi^D(p, w; w, w)$. To see this, note that, at the optimal deviation wage w_z^D for a single-product firm operating in market z,

$$\frac{\partial \pi^D(p, w; w_{i',-z}, w)}{\partial w_{i',-z}} = -\left(w_z^D + r \frac{\partial K(L(w_z^D, w; w_{i',-z}, w), Q(p)/n)}{\partial \ell} \right) \frac{\partial L(w_z^D, w; w_{i',-z}, w)}{\partial w_{i',-z}} =$$

$$= -L(w_z^D, w; w_{i',-z}, w) \frac{\frac{\partial L(w_z^D, w; w_{i',-z}, w)}{\partial w_{i',-z}}}{\frac{\partial L(w_z^D, w; w_{i',-z}, w)}{\partial w_i}} < 0.$$

Therefore, the deviation profit in market z is lower when the deviating multiproduct firm raises its wage in the other market $-z$. As a result, a deviating multimarket firm, raising its wage offers to expand production in both markets, cannot obtain twice as much the deviation profit of a single-product firm.

As the cartel outcome (p^M, w^M) – which, by construction, is as in the baseline model – does not depend on whether firms are single- or multiproduct, the results $\pi_m^D(p, w; w) < 2\pi^D(p, w; w, w)$ and, using the same notation for on-path profits, $\pi_m(p, w; w) = 2\pi(p, w; w, w)$ for all (p, w) immediately imply that

$$1 - \frac{\pi_m(p^M, w^M; w^M)}{\pi_m^D(p^M, w^M; w^M)} < 1 - \frac{\pi(p^M, w^M; w^M, w^M)}{\pi^D(p^M, w^M; w^M, w^M)},$$

which concludes the proof. □

Proof of Proposition 9. Take a stationary equilibrium (p, w) and any wages $w_{\tau,i}^D \geq w$ offered by the deviating firm in the period of deviation (as undercutting the equilibrium wage is never profitable). Then, there always exists a continuation equilibrium where the deviating firm makes zero profit, which constitutes an optimal punishment by the results in Abreu (1988).

To establish this claim, as firms' products are perfect substitutes, it is sufficient to prove that the average cost of the deviating firm is weakly larger than the average cost of the other firms. To this end, consider a continuation equilibrium where no NPAs are signed anymore. Such a continuation equilibrium always exists because when each firm expects rivals not to sign the NPA, unilaterally signing it would be inconsequential.

Consider the first period of the punishment phase. The aggregate labor force in the industry of workers of age $\tau \geq 3$ is $nL(w,w)$ (irrespective of whether NPAs were in place along the equilibrium path), as they were first hired when all firms offered the equilibrium wage: This is a symmetric pool of workers, in the sense that no firm can hire more workers from this pool than its rivals unless it offers a higher wage. This is true also for newcomers ($\tau = 1$), given that all workers are available for hire at the initial age. Finally, the industry labor force of workers of age $\tau = 2$ is

$$nL(w,w) + J\Pr[w^D_{1,i} + \xi_i \geq \max\{w_0, w + \max_{i' \in \mathcal{N}\setminus\{i\}} \xi_{i'}\} | w + \max_{\tilde{i} \in \mathcal{N}} \xi_{\tilde{i}} < w_0],$$

with $w^D_{1,i}$ denoting the offer they received from the deviating firm in the deviation period. That is, this pool of workers consists of the on-path industry labor force and the additional workers attracted to the deviating firm i because $w^D_{\tau,i} \geq w$. As $\xi_{i,j}$ are i.i.d. draws from the same distribution Ξ, also this is a symmetric pool of workers in the sense described above.

Therefore, irrespective of its wage offers in the deviation period, the deviating firm has never a cost-advantage *vis-à-vis* its rivals – i.e., it can never recruit a given labor force paying a lower average wage than its rivals.[56] These arguments also hold in subsequent periods, implying that the deviating firm makes zero profits forever after the deviation period.

Given that, for any profile of deviation wages, the deviating firm is guaranteed a payoff of zero in the continuation game, its highest deviation profit is obtained by minimizing the production cost of $Q(p)$ in the period of deviation. This is because the best static deviation, by the same arguments as in the baseline model, consists in slightly undercutting the candidate equilibrium price p – i.e., starting from any candidate SPNE (p,w), one can define the deviation profit as

$$\pi^D(p,w) \equiv pQ(p) - \min_{\{w_{\tau,i}\}} \sum_{\tau=2,\ldots,T} \left[wL(w,w) + w_{\tau,i}\tilde{\ell}_{\tau,i}(w_{\tau,i},w)\right] + w_{1,i}L(w_{1,i},w) + r\,K(\cdot).$$

[56]The latter indeed collectively have a strict cost advantage because *(i)* they know they can keep all their incumbent workers $\tau \geq 2$ even offering a wage lower by an amount $w^D_{\tau-1,i} - w$ relative to the one offered by the deviating firm (this only holds for $\tau = 2$ aged workers in the presence of a NPA on-path), and *(ii)* downward nominal wage rigidity is less stringent for them than for the deviating firm who has offered higher wages in the period of deviation (this gives them an advantage if the wage level in the punishment phase is lower than the deviation wage).

In the presence of a binding NPA in the period of deviation, this minimization problem is subject to the constraint $w_{\tau,i} = w$ for all $\tau = 2, ..., T$, so that $\tilde{\ell}_{\tau,i}(w, w) = 0$, given that offering rivals' workers any $w_i \leq w$ is equivalent not to make a poaching offer.

With a ban on NPAs, the first-order condition of the deviating firm's problem with respect to $w_{\tau,i}$, for all $\tau = 2, ..., T$, is given by

$$\tilde{\ell}_{\tau,i}(\cdot) + \left(w_{\tau,i} + r\frac{\partial K(\cdot)}{\partial \ell}\right)\frac{\partial \tilde{\ell}_{\tau,i}(\cdot)}{\partial w_{\tau,i}} = 0.$$

For $w_{\tau,i} = w \,\forall \tau \geq 2$, from the definition of $\tilde{\ell}_{\tau,i}(\cdot)$ it immediately follows that $\tilde{\ell}_{\tau,i}(\cdot) = 0 < \partial\tilde{\ell}_{\tau,i}(\cdot)/\partial w_{\tau,i}$, and so the left-hand side of the above equation equals

$$\left(w + r\frac{\partial K(\cdot)}{\partial \ell}\right)\frac{\partial \tilde{\ell}_{\tau,i}(\cdot)}{\partial w_{\tau,i}} < -\frac{L(w, w)}{\frac{\partial L(\cdot)}{\partial w_{\tau,i}}}\frac{\partial \tilde{\ell}_{\tau,i}(\cdot)}{\partial w_{\tau,i}} < 0,$$

where the first inequality follows from *(i)* $W^M(\delta) < W^*(P^M(\delta))$ and *(ii)* as $\sum_\tau L(w_{\tau,i}, w) < n\sum_\tau L(w, w)$ by the diseconomies of scale effect (*a fortiori* so if $w_{\tau,i} = w$ for all $\tau = 2, ..., T$), $\partial K(\sum_\tau L(w_{\tau,i}, w), Q(p))/\partial \ell < \partial K(\sum_\tau L(w, w), Q(p)/n)/\partial \ell$. By the second-order condition of the cost-minimization problem, this inequality implies that $w_{\tau,i} > w$ at the optimum for all τ. Therefore, the constraint $w_{\tau,i} = w$ for all $\tau = 2, ..., T$ (so that $\tilde{\ell}_{\tau,i}(\cdot) = 0$) imposed by the presence of NPAs binds, strictly lowers the maximum deviation profit.

By the same arguments of the baseline analysis, wage discrimination, even across workers of different ages, is always inefficient, implying that the cartel outcome is stationary and symmetric and so is defined as

$$(p^M, w^M) \equiv \arg\max_{p,w} \pi(p, w) \equiv p\frac{Q(p)}{n} - \left[wTL(w, w) + r\,K\left(TL(w, w), \frac{Q(p)}{n}\right)\right],$$

where maximization with respect to p and w gives the same first-order conditions (14)-(15) as in the baseline model. As punishment profits are zero, the cartel outcome is sustainable as a SPNE of the dynamic game if and only if

$$\delta \geq \delta^M \equiv 1 - \frac{\pi(p^M, w^M)}{\pi^D(p^M, w^M)}.$$

For lower values of the discount factor, by the same steps as in Proposition 3, the optimal collusive outcome (within the class of stationary and symmetric allocations) is $(P^M(\delta), W^M(\delta))$ obtained from the binding incentive-compatibility constraint and the optimality condition (10), with $P^M(\delta)$ and $W^M(\delta)$ being increasing and decreasing in δ, respectively, and such that, for all $\delta \in (0, \delta^M)$, $W^M(\delta) < W^*(P^M(\delta))$, with the static equilibrium offer $W^*(p)$ defined by Eqn. (13).

Then, as the constraint $w_{\tau,i} = w$ for all $\tau \geq 2$ imposed by NPAs in the deviating firm's problem strictly reduces the deviation profits from any candidate stationary equilibrium (p, w), in the presence of NPAs along the equilibrium path: *(i)* the critical discount factor δ^M is lower, and *(ii)* whenever the incentive-compatibility constraint binds absent NPAs, firms will optimally use the resulting slackness in the incentive constraint in the presence of NPAs to sustain an equilibrium with both higher price and lower wage levels.

The equilibrium where all firms sign a NPA at all periods is always sustained by the threat of reverting to a zero-profit equilibrium immediately if a firm does not sign the NPA. Such a zero-profit continuation equilibrium exists because the deviating firm can poach the signatories' incumbent workers and *vice versa*, and firms compete *à la* Bertrand. This drives the deviating firm's profit to zero, making a deviation not to sign the NPA always unprofitable. Therefore, as the most profitable equilibrium with NPAs features higher profits than the most profitable equilibrium without NPAs, the only way to prevent firms from signing NPAs is through a ban on these agreements, which, by the arguments above, has the effects stated in Proposition 9.

Finally, the same results hold if workers are *farsighted* and discount payoffs at the common rate δ. Along any stationary equilibrium path, newcomers anticipate that they will not get higher wages in future periods, so their acceptance decisions are unchanged. Conversely, when a firm deviates by offering newcomers higher wages, farsighted workers anticipate getting higher offers in the future because firms are reverting to a competitive equilibrium, which changes their labor supply decision.[57] This effect, however, is identical no matter whether NPAs are banned or not as, in either case, firms will not sign NPAs in the competitive continuation equilibrium following the deviation. □

Proof of Corollary 3. If NCAs are banned, the most collusive stationary SPNE is as in the Proof of Proposition 9 above in the scenario of a ban on NPAs. Otherwise, firms always find it optimal to sign NCAs, for two reasons. First, workers are willing to sign these clauses at no extra remuneration ($\omega_{\tau,i} = 0$ for all τ and i), as they do not expect to receive poaching offers along the equilibrium path, so signing these clauses is costless for firms. Second, NCAs signed by other firms with their workers reduce a deviating firm's profit in the best static deviation, as they play the exact same role as binding NPAs, which can only help sustaining more profitable equilibria. Then, if the possibility by the deviating firm of signing NCAs in the deviation period is valuable – i.e., if it increases its profits in the punishment phase – this firm will do so anyway. Otherwise, its punishment profits are zero, implying that its best deviation is indeed the optimal static deviation,

[57] This is true if newcomers can observe all offers or have *symmetric beliefs* – i.e., each believes that the deviating firm is offering the same contract to all other newcomers. In the case of private contracting and *passive beliefs*, as each worker still believes that the deviating firm is offering the equilibrium contract to any other worker and so does not expect a wage war, the labor supply is identical to the case of myopic workers.

and hence, the presence of NCAs hinders its deviation incentives. In this case, a ban on NCAs produces the same effects as those of a ban on NPAs characterized in Proposition 9. □

Proof of Corollary 4. Also in the presence of pay-equity regulations, following any deviation firms can play a continuation equilibrium in which the deviating firm makes zero profits, which constitutes an optimal punishment. This result follows from the same arguments in the Proof of Proposition 9 above and from the fact that the regulatory constraint further penalizes the deviating firm, which, because of downward nominal wage rigidity and the deviation wage w^D being higher than w, faces the tighter constraint $w_i \geq w^D$ in the punishment phase if it wants to hire newcomers. Therefore, the optimal deviation from any stationary symmetric SPNE (p, w) is the best static deviation – i.e., slightly undercutting p and choosing wage offers to minimize the production cost of $Q(p)$:

$$\min_{w_i} \sum_{\tau=2,\ldots,T} \left[w_i L(w, w) + w_i \tilde{\ell}_{\tau,i}(w_i, w) \right] + w_i L(w_i, w) + r\, K(\cdot),$$

whose first-order condition immediately yields that $w_i > w$. Then, the deviation profit is strictly lower than absent such regulations because keeping the wage at $w_i = w$ to own incumbent workers is optimal absent pay-equity rules. Therefore, introducing pay-equity regulations dampens deviation incentives from any candidate SPNE outcome (p, w). As a result, it reduces the critical discount factor to sustain the cartel outcome (which, being symmetric, does not depend on whether these rules are in place), and allows firms to sustain SPNE with strictly higher prices and lower wages for all lower values of δ. □

Appendix B: Supplementary material

B.1 Microfoundations

Price or wage collusion and antitrust monitoring. Suppose that firms need pre-play communication to coordinate on collusive price and wage levels and that such communication takes place among specialized middle managers of each firm, say pricing managers for prices and HR directors for wage offers (each acting in its employing firm's overall best interests), who then report to the top management that takes final decisions.

If such communication is not prevented by antitrust authorities, it leads firms to implement the multimarket collusive outcome, yielding the results of Section 3.2; if, on the contrary, antitrust authorities are able to prevent all middle managers from communicating, the best firms can do is playing the most profitable static NE $(\overline{p}^N, W^*(\overline{p}^N))$ over time. However, antitrust authorities may be able to prevent communication only *in one market*. If only communication among pricing managers is prevented (*monitoring of price* *collusion*), firms will agree upon w, and so only a different wage offer made by any firm will be considered a deviation and trigger Nash reversion; accordingly, each firm is free to choose its price in its individual best interest; anticipating this introduces the constraint $p \leq P^*(w)$ in the collusion Problem (P), yielding the wage collusion outcome described in Proposition 5. Conversely, if only communication among HR directors is prevented (*monitoring of wage collusion*), firms will only agree upon p, each firm will then set its individual profit-maximizing wage, which adds the constraint $w = W^*(p)$ to the collusion Problem (P), yielding the price collusion outcome described in Proposition 4.

The same results hold if antitrust authorities can only look for ex-post evidence of communication when they have a clue of collusive behavior. Assume the authorities know the production function and can observe product prices and wage offers, and firms' sales and labor force. Then, suppose they also know $L(\cdot)$, but not $Q(\cdot)$. In that case, they can observe whether the prevailing wage level is too low, i.e., $w < W^*(p)$ (as this equilibrium value only depends on the equilibrium quantity and not on the demand function), and in this case will look for, and eventually find, evidence of collusion. Anticipating this, colluding firms will optimally set $w = W^*(p)$, which implements the price collusion outcome (as authorities, not knowing $Q(\cdot)$, cannot infer whether a price is collusive). *Vice versa*, if authorities know $Q(\cdot)$, but not $L(\cdot)$, firms will optimally implement the wage collusion outcome. Authorities' knowledge of both consumer demand and labor supply functions or of none leads, respectively, to $(\overline{p}^N, W^*(\overline{p}^N))$ or the multimarket collusive outcome.

Search frictions. While, in the baseline model, firms have oligopsony power because of their differentiation *vis-à-vis* workers, similar conclusions hold in the presence of labor market power deriving from search frictions in the labor market.

To make this point in the simplest way, suppose that there is no employer differentiation: $\xi_{i,j} \equiv \xi_j \sim \Xi$ for all $i \in \mathcal{N}$. Then, worker j will join (one of) the firm(s) i from which it has received the highest offer $w_{i,j}$, provided that $w_{i,j} + \xi_j \geq w_0$. Yet, because of search frictions in the labor market, firms cannot reach all workers. For simplicity, say that a fraction λ/n of workers only receive the offer from one firm i, whereas the other workers actively search and consider offers from all firms.[58] Then, denoting $\underline{w} \equiv \min_{i \in \mathcal{N}} w_i$, one has

$$\ell_i = L(w_i, \underline{w}) \equiv J\left(\frac{\lambda}{n} + \mathbb{1}[w_i = \underline{w}]\frac{1-\lambda}{\#\{i : w_i = \underline{w}\}}\right)[1 - \Xi(w_0 - w_i)].$$

Starting from any candidate SPNE where all firms set price $p_i = p$ and try to reach all workers (which is always optimal to maximize the labor force for any given wage) offering

[58] The following argument applies, more generally, whenever each firm i cannot reach a sufficiently large fraction of workers within the deviation period.

$w_i = w$, a deviating firm, by slightly overcutting w, would hire

$$\ell_i = J\left(\frac{\lambda}{n} + 1 - \lambda\right)[1 - \Xi(w_0 - w)] < J[1 - \Xi(w_0 - w)] = nL(w, w)$$

workers. Under the no-rationing assumption, this translates into a diseconomies of scale effect, yielding similar results to those in the baseline analysis.

To see this, consider the cartel outcome (p^M, w^M), which is independent of λ.[59] As $nL(w^M, w^M)$ is the cost-efficient labor force to produce $Q(p^M)$, a deviating firm, who shall serve all this demand alone, would necessarily face a larger average cost (because it cannot recruit the same labor force when offering w^M to the subset of workers it can reach). This implies $\pi^D(p^M, w^M) < n\pi(p^M, w^M)$, and hence $\delta^M < \delta_0^M \; \forall \lambda > 0$.[60] Moreover, as the measure of workers that a deviating firm can steer away from its rivals by slightly overcutting w^M is decreasing in λ, it immediately follows that

$$\frac{\partial \delta^M}{\partial \lambda} < 0,$$

so that more significant search frictions (captured by a larger value of λ), which translate into labor market power even absent firm differentiation, unambiguously facilitate collusion.

B.2 Discussion and robustness

No-rationing assumption. The analysis has maintained the standard assumption of Bertrand models that firms are always committed to satisfy all consumers' demand at their posted price. How would the results change if firms could choose how much to produce? To answer this question in a standard model, suppose that firms simultaneously choose, on top of p and w, also the amount of the other production factor k to employ. Then, the static game would not admit a pure-strategy NE with zero profits: In the candidate NE where firms are supposed to sell at $p = c(Q(p)/n)$, they would choose (w, k) so that $F(L(w, w), k) = Q(p)/n$ (i.e., no excess capacity in equilibrium); but then, any firm would have incentives to overcut this price, as it would face no competition for the residual demand.

[59] As in the baseline model, it is straightforward to prove that the cartel outcome is indeed stationary and symmetric, and so is defined by:

$$(p^M, w^M) \equiv \arg\max_{p,w}\, pQ(p) - n\left[wL(w, w) + r\,K\left(L(w, w), \frac{Q(p)}{n}\right)\right] =$$
$$= pQ(p) - \left[w\,J[1 - \Xi(w_0 - w)] + r\,K\left(J[1 - \Xi(w_0 - w)], Q(p)\right)\right],$$

where the equality follows from the definition of $L(\cdot)$ and from $K(\cdot)$ being homogeneous of degree one.

[60] Indeed, a zero-profit static NE to which firms can revert in the punishment phase, characterized as in the baseline analysis, always exists under this specification of the model.

Indeed, even if k could be adjusted ex-post, the deviating firm's rivals would have no incentives to serve the consumers it has left because doing so would entail an increase in the average production cost. In other words, the only role played in the analysis by the possibility of adjusting the endowment of k after demand is realized is that jointly with the no-rationing assumption, it implies that overcutting the competitive price $p = c(Q(p)/n)$ is not profitable. In turn, this entails the existence of a zero-profit static NE, which allows to restrict attention to grim-trigger strategies.

Nevertheless, provided that there exists a punishment yielding a discounted profit of zero to a deviating firm in the continuation game, the qualitative results are unchanged under the simultaneous timing considered here.[61] The reason is that the price-overcutting incentives described above are absent in the most profitable SPNE: Colluding firms' ability to charge high prices is instead constrained by their incentives to undercut the candidate equilibrium price. Such undercutting incentives are stronger than in the baseline analysis whenever, starting from a candidate SPNE (p, w), a deviating firm would find it optimal to sell a quantity $q \in (Q(p)/n, Q(p))$, rather than $q = Q(p)$ as is constrained to do in the base model. In these circumstances, this alternative timing of the game strengthens deviation incentives, resulting in lower prices and higher wages in the most profitable SPNE. However, this is the case only for relatively low values of δ: As Assumption (A) implies that serving all the demand is optimal for a firm deviating from the cartel outcome, by a standard continuity argument it follows that the most profitable SPNE is unchanged for sufficiently large values of δ.

Sequential stage game. The main results are robust under sequential-moves specifications of the stage game. In a *production-to-order* game, in which prices are set (and publicly-observed) in advance of purchasing production factors, a firm undercutting a candidate equilibrium price benefits from its rivals optimally avoiding hiring any worker – i.e., $p_i < p_{i'}$ implies no demand for all firms $i' \in \mathcal{N} \setminus \{i\}$, and so $w_{i'} = k_{i'} = 0$. This feature of the production-to-order game strengthens undercutting incentives, resulting in less collusive equilibria; yet, because of employer differentiation, it does not eliminate the diseconomies of scale effect. Except for this, the analysis is unchanged (in particular, there exists a zero-profit static NE, implying that grim-trigger strategies are without loss of generality).[62]

In a *production-in-advance* game, where firms purchase production factors before setting prices – i.e., they first make wage offers and choose capital endowments, and then,

[61] As shown by Abreu (1988), grim-trigger strategies are not optimal punishments unless, as in the baseline model, they yield a discounted profit of zero. Otherwise, stick-and-carrot strategies suffice to yield zero discounted profits in the continuation game following any deviation for values of δ *not too small*.

[62] Note that a deviation only in the wage offers is *a fortiori* suboptimal given that a price-deviation is more profitable than in the baseline model.

upon observing these choices, set their prices – as known since Kreps and Scheinkman (1983), prices will be non-cooperatively set at the market clearing level – i.e., $p = Q^{-1}\left(\sum_{i\in\mathcal{N}} F(\ell_i, k_i)\right)$ – which maximizes individual and industrywide profits for the chosen *capacities*. Therefore, a production-in-advance game boils down to a model of collusion in production capacities, making monitoring of wage collusion even more crucial to ensure low prices. The incentives to deviate from a collusive outcome (w, k) – or, equivalently, (p, w), with each firm buying $k = K(L(w, w), Q(p)/n)$ – are weaker than in the simultaneous-choice stage game as a deviating firm not only has to pay a higher wage to produce more but will also sell at a lower price. In a SPNE in grim-trigger strategies, this effect would be at least in part outweighed by the fact that reversion to the static NE does not lead to zero profits. As above, however, the result that, through the diseconomies of scale effect, labor market power facilitates collusion always holds provided that firms can employ optimal punishment schemes yielding zero discounted continuation profits to a defector.

Production function. Throughout the analysis, firms only employ two variable production factors. As in Yeh et al. (2022), the analysis immediately generalizes to additional variable factors traded in competitive markets. Formally, let $k_1, ..., k_V$ be variable production factors, and r_v be the unit price of each factor k_v, $v = 1, ..., V$. Suppose that the production function is CRS with respect to $\{\ell, k_1, ..., k_V\}$. Then, for any quantity q_i that firm i needs to produce, given its labor force ℓ_i, it chooses $(k_1, ..., k_V)$ solving

$$\min_{k_1,\ldots,k_V} \sum_{v=1,\ldots,V} r_v k_V$$
$$\text{s.t. } q_i = F\Bigg(\ell_i, \sum_{v=1,\ldots,V} k_V\Bigg).$$

Letting $k_v = K_V(\ell_i, q_i)$, for $v = 1, ..., V$, denote the solution of the above cost-minimization problem, the same expression for firms' profit obtains by considering the vector notation $r = (r_1, ..., r_V)$ and $K = (K_1(\ell_i, q_i), ..., K_V(\ell_i, q_i))$; the analysis is unchanged as each $K_v(\cdot)$ is homogeneous of degree one.

Next, suppose that on top of the (at least two) variable production factors considered so far, also fixed factors enter the production function. For conciseness, consider the two variable factors (ℓ, k) as in the base model and only one fixed factor, denoted by z; its level is first set by each firm and becomes common knowledge at the outset of period $t = 0$, then can be publicly revised at the outset of period $t = T$, for some $T \geq 1$, then again at the outset of period $t = 2T$, and so on (except for this, the timing of the game is as in the baseline model). Provided that any deviation in the choice of the fixed factor triggers a zero-profit equilibrium of the continuation game, colluding firms optimally choose the value of z to facilitate price and wage collusion. Formally, they solve the following problem:

$$\max_{p,w,z} \pi(p,w,z) \equiv p\frac{Q(p)}{n} - \left[r_z z + wL(w,w) + r\,K\left(z, L(w,w), \frac{Q(p)}{n}\right)\right]$$
$$\text{s.t. } \delta \geq 1 - \frac{\pi(p,w,z)}{\pi^D(p,w,z)},$$

where r_z is the unit price of z, $K(\cdot) \equiv F^{-1,k}(z,\ell,q)$ is the inverse of the production function $F(z,\ell,k)$ w.r.t. k, and the deviation profit is

$$\pi^D(p,w,z) \equiv \max_{q_i \in \left\{\frac{Q(p)}{n}, Q(p)\right\}, w_i} pq_i - \left[w_i L(w_i,w) + r\,K\left(L(w_i,w), z, q_i\right) + r_z z\right],$$

given that the deviating firm cannot change its choice of z even at the outset of periods $t \in \{0, T, 2T, ...\}$, because doing so would entail triggering a zero-profit continuation equilibrium. The first-order condition w.r.t. z, when the incentive-compatibility constraint binds, gives

$$\frac{\partial \pi(\cdot)}{\partial z} = \lambda \frac{\partial}{\partial z}\left[\frac{\pi(\cdot)}{\pi^D(\cdot)}\right],$$

with λ being the Lagrange multiplier of the above problem, so that firms account for how the choice of z affects deviation incentives. In particular, setting an inefficiently low z from an industrywide profit maximization standpoint (given p and w) can be optimal to weaken price-undercutting incentives.

Importantly, if the production function is CRS with respect to all factors $\{\ell, k, z\}$, the presence of a fixed factor facilitates collusion through a diseconomies of scale effect, given that a deviating firm cannot expand its endowment of z, and so faces decreasing returns w.r.t. $\{\ell, k\}$. This, in turn, further exacerbates the diseconomies of scale effect brought up by the presence of labor market power: The decreasing returns of labor when z is fixed imply that paying higher wages to attract more workers is less appealing than in the baseline analysis.

Imperfect monitoring. Imperfect labor market competition can have further pro-collusive effects by allowing firms to detect better deviations in settings with demand shocks and imperfect monitoring *à la* Green and Porter (1984) – see, e.g., Matsushima (2001). In perfectly competitive labor markets, firms can recruit as many workers as they want at the competitive wage. So, the labor market delivers no information on whether a rival has deviated from the collusive price. Conversely, in the presence of labor market power, even if there is uncertainty on the distribution Ξ at every period, firms ending up with an unlikely low labor force (given their wage offer) can draw inferences on the fact that a rival has increased its wage to recruit more workers.

Hence, a deviating firm undercutting the price either also increases its wage offer, which maximizes its deviation profit but increases the chances of triggering the punishment phase, or it does not deviate in the labor market, which makes price-undercutting less profitable in the first place. In either case, it has weaker deviation incentives.

Wage cartels. Suppose that wage cartels are allowed and firms reach an agreement, enforceable in court, which constrains them to offer a wage w to all workers. As explicit price cartels are still illegal, colluding firms solve Problem (P) with a deviation profit

$$\max_{q \in \{Q(p)/n, Q(p)\}} p\, q - [wL(w,w) + r\, K(L(w,w), q)] < \pi^D(p,w),$$

which softens the incentive-compatibility constraint. That is, by making it impossible to recruit more workers at any period, a legally binding wage cartel reduces the profitability of undercutting any candidate SPNE price. The reason is that a deviating firm can satisfy the increased demand by only increasing its endowment of variable capital, which is not cost-efficient.

Then, for all $\delta \in [0, \delta^M)$ – i.e., whenever the incentive-constraint binds without an explicit wage cartel – firms would find it optimal to write down a legally binding wage-fixing agreement to sustain both lower wages and higher prices than in the most profitable SPNE characterized in Proposition 3. This argument provides a rationale for the *per se illegality* antitrust status of explicit wage-fixing agreements, even based on a pure consumer surplus standard.

Collective bargaining. Consider a model where firms open vacancies, and the wage they end up paying workers in equilibrium depends on industrywide trade unions' bargaining power.

Formally, each firm $i \in \mathcal{N}$ simultaneously chooses how many workers ℓ_i to employ and its product price p_i. The latter choices determine consumers' demand as in the main model. The former choices determine the prevailing wage level. Suppose for simplicity that firms are homogeneous from workers' viewpoint – i.e., as above, $\xi_{i,j} \equiv \xi_j \sim \Xi$. Absent trade unions' bargaining power, the wage level $w^*(L) < w_0$, where $L \equiv \sum_{i \in \mathcal{N}} \ell_i$, is obtained from the market clearing condition $J[1 - \Xi(w_0 - w^*(L))] = L$. If, on the contrary, trade unions have full bargaining power, they are always able to impose the competitive wage: $w_0(L) \equiv w_0\ \forall L$. More generally, suppose that the expected or average wage paid by each firm equals $W(L) \equiv \alpha w^*(L) + (1-\alpha) w_0$, where the parameter $\alpha \in [0,1]$ is an inverse measure of trade unions' bargaining power – e.g., there is a probability α with which trade unions will manage to ex-post impose competitive wages for all workers or a fraction α of workers is unionized and is ex-post able to extract the competitive wage.

For $\alpha = 0$, this model is identical to the competitive input markets benchmark examined in Section 2.2: The cartel outcome is thus sustainable by colluding firms if and only if $\delta \geq \delta_0^M$. For $\alpha = 1$, the analysis is similar to Section 3. In particular, starting from a symmetric candidate SPNE (ℓ, p), a deviating firm undercutting the price finds it optimal to expand its labor demand. Doing so, however, implies a rise in the market clearing wage. Because of this diseconomies of scale effect, the critical discount factor to sustain the cartel outcome, defined here as $(p^M, L^M) \equiv \arg\max_{p,L} pQ(p) - [W(L)L + r\,K(L, Q(p))]$, is $\delta^M < \delta_0^M$. More generally, for $\alpha \in [0, 1]$ one has that

$$\frac{\partial \delta^M}{\partial \alpha} < 0.$$

Therefore, policies weakening trade unions' bargaining power facilitate collusion.

B.3 Local labor markets and cross-market collusion

This section considers the model with independent product markets described in Section 4.2, with $n/2$ different single-product firms in each market.

Suppose that firms operating in market z only collude among each other (*within-market collusion*). Then, for any given δ and wage w_{-z} offered by firms in the other market, their optimal collusive scheme (p_z, w_z) maximizes the profit $\pi(p_z, w_z; w_{-z}, w_{-z})$ subject to the incentive-compatibility constraint $\delta \geq 1 - \pi(\cdot)/\pi^D(\cdot)$, where the highest deviation profit, similar to the baseline analysis, is given by $\pi^D(p_z, w_z; w_{-z}, w_{-z})$ (the functions $\pi(\cdot)$ and $\pi^D(\cdot)$ are defined in the Proof of Proposition 8).

Then, the within-market cartel outcome is defined as the solution to the fixed-point problem: $(p^S, w^S) \equiv \arg\max_{p_z, w_z} \pi(p_z, w_z; w^S, w^S)$, and it is sustainable for all $\delta \geq \delta^S$, where

$$\delta^S \equiv 1 - \frac{\pi(p^S, w^S; w^S, w^S)}{\pi^D(p^S, w^S; w^S, w^S)} < \delta_0^M.$$

For lower values of δ, the most profitable SPNE $(P^S(\delta), W^S(\delta))$ is again pinned down by the binding incentive-compatibility constraint and the optimality condition (10), where derivatives are taken with respect to variables in the own market z, imposing symmetry ($p_z = p$ and $w_z = w$ for $z = A, B$).

As within-market cartels do not internalize the cross-market externalities taking place through the labor market, w^S is too high from a cross-industry profit maximization viewpoint: A cross-market cartel would profitably set $(p^M, w^M) \equiv \arg\max_{p,w} \pi(p, w; w, w)$, with $w^M < w^S$ and, as hiring fewer workers makes it optimal to reduce production, $p^M > p^S$.

Firms selling their products in independent product markets thus have incentives to collude together. Under *cross market collusion*, the most profitable SPNE, denoted by

$(P^M(\delta), W^M(\delta))$, solves the following problem:

$$\max_{p,w} \pi(p, w; w, w)$$
$$\text{s.t. } \delta \geq 1 - \frac{\pi(p, w; w, w)}{\pi^D(p, w; w, w)}.$$

As cooperative and defection profits are unchanged, sustaining the more collusive cross-market cartel outcome requires a higher critical discount factor:[63]

$$\delta^M = 1 - \frac{\pi(p^M, w^M; w^M, w^M)}{\pi^D(p^M, w^M; w^M, w^M)} \in (\delta^S, \delta_0^M).$$

For all $\delta \in [0, \delta^M)$, the incentive-compatibility constraint binds, and the cross-market optimal collusive scheme satisfies the optimality condition

$$\frac{\partial \pi(\cdot)/\partial p_z}{\partial \pi^D(\cdot)/\partial p_z} = \frac{\partial \pi(\cdot)/\partial w_z + \partial \pi(\cdot)/\partial w_{-z}}{\partial \pi^D(\cdot)/\partial w_z + \partial \pi^D(\cdot)/\partial w_{-z}}. \tag{17}$$

The internalization of cross-market externalities, as reflected by the extra-terms on the right-hand side of Eqn. (17) relative to Eqn. (10), implies that $W^M(\delta) < W^S(\delta)$ for all $\delta > 0$. Whenever the incentive-compatibility constraint also binds under within-market collusion, however, colluding firms need to lower their price to offset the stronger incentives to deviate deriving from the more collusive wage level. Summing up:[64]

Proposition (cross-market collusion). *The most profitable SPNE $(P^M(\delta), W^M(\delta))$ with multiple (identical and independent) product markets is only achievable through cross-market collusion and is such that $W^M(\delta) < W^S(\delta)$ for all $\delta > 0$, and $P^M(\delta) > P^S(\delta)$ if and only if $\delta > \widetilde{\delta}$, where $\widetilde{\delta} \in (\delta^S, \delta^M)$.*

Labor market power in local labor markets creates the scope for cross-market collusion, strengthening the case for antitrust authorities' monitoring of collusive behavior. Interestingly, this broadening of collusive behavior does not necessarily translate into higher consumer prices: The presence of significant wage mark-downs unambiguously reveals that cross-market collusion is in place.[65]

[63] Still, $\delta^M < \delta_0^M$ as, by the same arguments outlined in Section 3.2, $\pi^D(p^M, w^M; w^M, w^M) < n\pi(p^M, w^M; w^M, w^M)$.

[64] The complete proof is omitted for brevity and available upon request.

[65] How would the above results change if markets are asymmetric? In the presence of significant asymmetries across markets, there may be values of δ such that the within-market cartel outcome is attainable in only some markets. In these circumstances, a cross-market collusive scheme can exploit the slackness of the incentive constraint in these markets to depress wage offers in all markets further. Then, prices can simultaneously rise in these markets where the incentive-compatibility constraint does not bind and drop in markets where the lower wage levels tighten the incentive constraint.

B.4 NPAs/NCAs as facilitating practices in perfectly competitive markets

This section considers the model of Section 5 assuming away product market oligopoly and employer differentiation, to derive the procollusive implications of NPAs and NCAs in the simplest setting.

Formally, suppose that firms can sell any quantity of their products at the exogenous, competitive price p_0 and are perfectly homogeneous from workers' viewpoint – i.e., $\xi_{i,j} \equiv \xi_j$ for all i. In this simple setting, firms' per-period (on-path) profit can be written as

$$\pi_i = \sum_{\tau=1,\ldots,T} (p - w_{\tau,i})\ell_{\tau,i},$$

where $\ell_{\tau,i}(\cdot)$ is firm i's labor force of age-τ workers (which will be characterized below) when each firm i offers the same wage $w_{\tau,i}$ to all workers of age τ (which holds, both on- and off-path, in the stationary equilibria considered below), and

$$p \equiv -r\frac{\partial K(\sum_\tau \ell_{\tau,i}, q^*(\sum_\tau \ell_{\tau,i}))}{\partial \ell},$$

with

$$q^*\left(\sum_\tau \ell_{\tau,i}\right) = \arg\max_q p\, q - \left[\sum_\tau w_\tau \ell_{\tau,i}(\cdot) - r\, K\left(\sum_\tau \ell_{\tau,i}, q\right)\right],$$

being a constant (i.e., p is independent of all $\ell_{\tau,i}$).[66]

No-poaching agreements. To simplify the exposition, suppose for the moment that workers are myopic. The dynamic game admits an equilibrium where, at any period and for any history of the game, no NPAs are signed and all firms offer $w = p$ to all available workers (i.e., all newcomers and all incumbent workers in the industry). Indeed, there are no profitable deviations from this equilibrium: At any t, unilaterally signing NPAs when other firms do not is inconsequential and, by a standard Bertrand-type argument, offering $w_{\tau,j} < p$ to any worker j of age τ implies not hiring it with probability one, while offering $w_{\tau,j} > p$ implies making a negative per-period surplus $p - w_{\tau,j} < 0$ from its hire. In this *competitive equilibrium*, firms make zero profits at each period.

[66] This result follows from the production function being CRS. Profit maximization with respect to q yields that $q^*(\sum_\tau \ell_{\tau,i})$ must satisfy the first-order condition $p_0 - r\frac{\partial K(\cdot)}{\partial q} = 0$, which, substituted into the profit, and using the fact that $K(\cdot)$ is homogeneous of degree one, gives the above expression for π_i. Finally,

$$\frac{\partial p}{\partial \ell_{\tau,i}} = -r\left[\frac{\partial^2 K(\cdot)}{\partial \ell_{\tau,i}^2} + \frac{\partial^2 K(\cdot)}{\partial \ell_{\tau,i}\partial q}\frac{\partial q^*(\cdot)}{\partial \ell_{\tau,i}}\right] = 0,$$

as, by the Implicit Function Theorem, $\partial q^*(\cdot)/\partial \ell_{\tau,i} = -\frac{\partial^2 K(\cdot)/\partial \ell_{\tau,i}\partial q}{\partial^2 K(\cdot)/\partial q^2}$, and, by the homogeneity of $K(\cdot)$, $(\partial^2 K(\cdot)/\partial \ell_{\tau,i}^2)(\partial^2 K(\cdot)/\partial q^2) = (\partial^2 K(\cdot)/\partial \ell_{\tau,i}\partial q)^2$.

By contrast, in the *monopsony equilibrium*, the wage offered to all available workers at every period would be $w^M \equiv \arg\max_w (p-w)TL(w) < p$, where $L(w) \equiv J\Xi(w_0 - w)$ is the *myopic labor supply* of each generation of workers.[67] As, in this simple model, firms are homogeneous from workers' viewpoint, the cartel outcome would feature the same wage w^M offered by all firms to all available workers over time, yielding per-firm profit $\pi(w^M) = (p - w^M)TL(w^M)/n$ at any period.

Suppose, first, that NPAs are not in place – e.g., they are banned by competition authorities. Then, starting from any stationary equilibrium with wage $w \in [w^M, p)$, the best deviation for any firm consists in slightly overcutting this wage offer to all available workers – i.e., all rivals' incumbent workers and all newcomers, which will then all join the deviating firm. By doing so, it obtains a profit

$$\pi^D(w) = (p - w)TL(w) = n\pi(w)$$

in the period of deviation. Afterward, however, firms will play the equilibrium where they offer $w = p$ to all available workers and so make zero profits, which therefore constitutes an optimal punishment.[68] Hence, absent NPAs, any stationary equilibrium wage $w \in [w^M, p)$ can be sustained as a SPNE of the dynamic game if and only if

$$\delta \geq \delta_0^M.$$

Consider, next, the case where NPAs are allowed, and are signed by all firms at any period along a stationary equilibrium path.[69] If any of the firms deviates and does not sign the NPA at some period t, all firms will immediately revert to a continuation equilibrium where they offer $w = p$ to all available workers starting from the same period t – i.e., all firms offer the competitive wage to newcomers and renegotiate at the competitive level the wage of their incumbent workers, so that the possibility of poaching rivals' incumbent workers cannot increase the deviating firm's profit above zero.[70] A firm contemplating a

[67] As seen in Section 3.1, wage discrimination would only increase the average wage needed to recruit any overall labor force. In this setting, there is no incentive to offer to available workers of any age τ a wage different from the statically optimal one w^M: Offering $w < w^M$ to a newcomer entails losing forever the profitable possibility of hiring it; and, once a worker is hired at w^M, it is pointless to increase its wage later on, and impossible to reduce it by the assumption of Pareto-improving renegotiation.

[68] As this equilibrium of the continuation game following the deviation is an optimal punishment no matter the wage the deviating firm has offered in the previous period and no matter how many workers it has employed, the optimal deviation is indeed the statically optimal one described above.

[69] In order for NPAs to have a bite, there must be some incumbent workers in the market, so the following analysis only applies for $t \geq 1$. However, one may consider that, at $t = 0$, only one generation of workers is already in the market, with the second generation coming in at $t = 1$, and so on, so that T generations of workers are available for all $t \geq T - 1$ (this does not affect the monopsony wage level w^M). This makes deviating in the first periods less profitable than in $t \geq T - 1$, implying that, both with and without NPAs, the critical discount factor is pinned down by the binding incentive-compatibility constraint at any $t \geq T - 1$.

[70] This is indeed a continuation equilibrium because, as seen above, offering $w = p$ to any worker is a best response for any firm when rivals do the same; moreover, by yielding a profit of zero to

deviation in period t shall thus sign the NPA and then deviate in the subsequent wage-offer game. In this case, the optimal deviation still consists in slightly overcutting the candidate equilibrium wage w; however, because of the binding NPA, this offer can only be made to newcomers and not to rivals' incumbent workers. As a result, given that firms revert to the continuation equilibrium where NPAs are not signed anymore and they offer $w = p$ to all available workers from period $t + 1$ onwards, the deviation profit is

$$\pi^D(w) = \frac{1}{n}(p - w)(T - 1)L(w) + (p - w)L(w) < n\pi(w).$$

Therefore, a stationary equilibrium with NPAs and wage $w \in [w^M, p)$ can be sustained as a SPNE of the dynamic game for all

$$\delta \geq \delta^M \equiv \frac{n-1}{n+T-1}, \quad \text{with} \quad \delta^M < \delta_0^M.$$

The above analysis implies that, for all $\delta \in [\delta^M, \delta_0^M)$, a ban on NPAs makes it impossible for firms to sustain the cartel outcome and so would result in higher equilibrium wages.[71]

Finally, the result that banning NPAs increases the critical discount factor to sustain collusive wages also holds if workers are farsighted and discount future profits at the common rate δ. Indeed, along any stationary equilibrium path, newcomers anticipate not getting higher wages in future periods. So, their decision to accept is the same, whether or not they are farsighted. Conversely, when a firm deviates by overcutting the candidate equilibrium wage, farsighted workers (provided they have *symmetric beliefs* or all offers are publicly observed) anticipate getting higher offers in the future because firms revert to the competitive equilibrium and are more willing to enter the industry (see below). This effect, however, is identical no matter whether NPAs are banned or not as, in either case, the competitive equilibrium played after the period of deviation entails that no NPAs are signed anymore and these workers obtain $w = p$.

Non-compete agreements. A myopic worker is always willing to sign a NCA without asking for compensation, given that this clause could only impact its wage in the future. Therefore, a meaningful analysis of NCAs requires considering farsighted workers (and *symmetric beliefs* or public offers). For simplicity of exposition, let me consider two-period lived workers ($T = 2$), and focus on the critical discount factor to implement the cartel outcome, i.e., the wage level w^M characterized above.

the deviating firm already in the deviation period, playing this continuation equilibrium constitutes an optimal punishment.

[71] Note that, once explicit, legally binding NPAs are banned, the possibility of firms tacitly agreeing not to poach each other's workers is valueless: Poaching rivals' workers is optimal only for a firm who wants to overcut the candidate equilibrium wage level to expand its labor force; the deviating firm would then also renege on the implicit NPA, which therefore cannot help firms to sustain collusive wage levels.

Under a ban on NCAs, for any deviation wage $w > w^M$, newcomers anticipate the wage war in the future period (i.e., that they will earn $w = p$ in the successive period) and so accept the deviating firm's offer if and only if $w + \delta p \geq (1 + \delta)(w_0 - \xi_j)$, yielding a labor supply $L^D(w) \equiv 1 - \Xi[w_0 - (w + \delta p)/(1 + \delta)] > L(w)$ for all w. As this labor supply is less elastic to wage relative to the on-path labor supply $L(w)$, the unconstrained monopsony wage would be lower, and so a deviating firm optimally slightly overcuts w^M *vis-à-vis* newcomers, besides with rivals' incumbent workers, and gets a deviation profit

$$\pi^D = (p - w^M)(L(w^M) + L^D(w^M)) > 2(p - w^M)L(w^M) = n\pi$$

in the deviation period, and zero afterward (because of wage war in the punishment phase for all available workers, who are all contestable). As a result, the critical discount factor to sustain the cartel outcome is strictly higher than δ_0^M.

Next, suppose NCAs are allowed, and consider a stationary equilibrium where, at all periods, all firms offer w^M, sign NCAs with all their workers, and do not make offers to rivals' incumbent workers. Then, as on-path workers do not expect to receive any attractive offer from their incumbent employer's rivals in the future, they are willing to sign these clauses at no extra compensation. Hence, a deviation consisting only of not offering NCAs to any worker is immaterial: The only relevant deviations are in wage offers. Newcomers anticipate that, by accepting NCAs from the deviating firm, they will earn the same wage w^M in the second period rather than the competitive wage p that they would get absent NCAs. Hence, one needs to distinguish two cases. First, if the deviating firm offers NCAs to newcomers, their labor supply will be $L(w)$ as in the myopic case. Then, the deviating firm optimally slightly overcuts w^M *vis-à-vis* newcomers, whom it employs at the same wage in the following period (no other profits are made from that period onwards). This deviation is not profitable if and only if

$$(p - w^M)\frac{L(w^M)}{n} + (p - w^M)L(w^M) + \delta(p - w^M)L(w^M) \leq \frac{1}{1 - \delta}2(p - w^M)\frac{L(w^M)}{n},$$

which is equivalent to $\delta \geq \delta_0^M$.

Second, if the deviating firm does not offer NCAs to newcomers, it is *de facto* providing them lifetime utility $w + \delta p$ when making an offer w, because of the wage-war in the punishment phase. In order to attract them, this must be higher than the lifetime utility $(1 + \delta)w^M$ offered by rivals (as they still offer the equilibrium contract, which features NCAs, and so would be able to employ them at w^M also during the punishment phase). As rivals offer the profit-maximizing lifetime utility, a deviating firm would optimally only slightly overcut it: $w = (1 + \delta)w^M - \delta p$; this implies that it again hires $L(w^M)$ and obtains a deviation profit

$$(p - w^M)\frac{L(w^M)}{n} + [p - ((1+\delta)w^M - \delta p)]L(w^M),$$

which, rearranged, is equal to the deviation profit in the previous case.

As a result, the cartel outcome can be sustained in a SPNE with NCAs for all $\delta \geq \delta_0^M$: Firms' possibility of signing NCAs thus facilitates collusion.

Chapter II

Fair Gatekeeping in Digital Ecosystems

This chapter is based on joint work with Jean Tirole (TSE). All authors contributed equally to this research.

1 Introduction

Dominant gatekeepers – the platforms controlling "core services" such as app stores, e-commerce, search, or social networking – are often suspected of charging excessive platform fees to business users (apps, merchants, advertisers) and/or practicing self-preferencing (favoring their own offerings when hybrid,[1] i.e., when competing in the markets they operate). This has led to numerous recent investigations and lawsuits concerning platforms such as Amazon, Apple, Booking, and Google;[2] similar questions will probably surface as AI-based platforms in tech and in health come to the fore.

Detecting "excessive fees" and "self-preferencing," the two prongs of the regulators' equity concern (e.g., in the EU Digital Market Act, DMA), is notoriously difficult or costly; and so regulators must pick their fights, which requires looking for smoking gun evidence that business users overpay or are discriminated against. Unfortunately, current regulations contain only broadly scripted prescriptions such as the DMA's requirement that access conditions be fair, reasonable and nondiscriminatory (FRAND). This leaves open the question of what "fair and reasonable" conceptually means, even putting aside the measurement issue. Our paper attempts to fill this void.

Why are policymakers preoccupied with business users' welfare in particular? After all, over twenty years of research have taught us that the "see-saw effect" in two-sided markets (that a price increase on one side increases the profitability of attracting users on the other side and induces a concomitant price decrease there) implies that antitrust analysis should consider the entire market and not just its business side. Similarly, the hypothesis of self-preferencing runs counter the Chicago School argument that a rich ecosystem brings product variety and lower prices, which can be monetized on the consumer side.

The paper builds a framework capable of accounting for existing business strategies and assessing regulation over a rich array of digital platform environments. It explains why there is a good reason to be preoccupied with equity for business users in the context of digital platforms: The important role played by two zero lower bounds (ZLBs) on core and app services (whose prices cannot be negative because of arbitrage and so are most

[1] Pure platform players (like Airbnb or Booking, which operate markets, but do not compete in them) cannot engage in self-preferencing, although they might enter into "sweet deals" with selected business users to the same effect.

[2] Examples of self-preferencing include the 2017 EU Google Shopping decision, the 2021 Google case in Italy (Android Auto did not accept an Enel's app that competed with Google Maps), and investigations into Amazon's prominent display of Amazon-branded goods and favoring its own logistics services (FBA).

Regarding excessive fees, several antitrust cases (Epic Games v. Apple; Spotify v. Apple; 2024 EU investigation of Apple and Google's non-compliance with the Digital Market Act) concern third-party apps trying to circumvent the 30% app store fee they deem unfairly high. The clampdown on most-favored-nation clauses similarly aims at capping access fees paid by merchants. Regulators may also directly set caps on access fees. Many local governments in the US introduced caps on the fees that food delivery platforms charge restaurants during the COVID-19 pandemic, and several of them then made these caps permanent. The major platforms (Uber Eats, Grubhub and DoorDash) typically charge a 30% fee, and most governments capped these fees to 15%.

often equal to 0)[3] in the setting of privately and socially optimal platform fees. Despite their importance for strategic behavior and policy and their endogeneity, the multi-sided-platform literature has mostly ignored ZLBs, or else posited free consumer access to the platform.

From a bird's eye view, platforms, whether app stores, e-commerce, OTAs, search engines, or social networks, all provide business users (apps, merchants, advertisers) with access to the consumers. Business users may thereby sell their goods or services and, importantly, receive ancillary benefits from attracting a consumer: Advertising revenues (content providers), data collection (most apps), fees collected from merchants selling their products through the app, or else the future profits attached to repeat purchases and upgrades; we capture these per-consumer ancillary benefits in a given app market by a variable $b > 0$. Ancillary benefits imply that the marginal cost is negative for digital goods, making the app ZLB particularly relevant (in contrast, $b < 0$ for most physical goods, whose production cost must be subtracted when computing the ancillary benefit). For digital goods, an incentive for self-preferencing arises when the platform is vertically integrated and makes more money by supplying, even for free, the good or app itself than by being paid for giving access, i.e., when $b > a$, where a is the access fee paid by the app or merchant to the platform. Even if there is no such foreclosure, low or nil access fees have a second drawback: they invite entry by me-too apps, that add little value to the ecosystem but extract a non-negligible share of it, because competition in the app market is hindered by the app ZLB.

While regulation may keep access fees low or nil, laissez-faire in contrast generates extractive access fees that squeeze business users; furthermore, the core ZLB, when relevant, blocks the see-saw effect and prevents consumers from benefitting from the squeeze. Such extractive fees both induce a suboptimal usage of apps and discourage their creation in the first place.

Section 2 derives a simple rule for the optimal regulation of access conditions. A "Pigouvian rule" ($\hat{a} = b$) allows the third-party apps to capture their contribution to the ecosystem, promoting the right level of innovation; it does so by pricing the unpriced positive externality (ancillary benefit) enjoyed by an app that receives access to the consumer. It also minimizes double marginalization in the set of access fees that do not induce self-preferencing.

Section 3 discusses the costs and benefits of various approaches to implementing the Pigouvian rule in the real-world context of heterogenous app sub-markets: regulatory measurement of b (perhaps triggered by an appeal against 'unfair' access fees), a con-

[3]Freely available apps include some of the most common third-party apps (e.g., payment apps such as PayPal, news aggregation apps such as Flipboard), as well as the competing in-house apps by Apple and Google (e.g., Apple Pay and Google Pay, Apple News and Google News, respectively). On the core side, most digital platforms, such as the major app stores, e-commerce platforms, search engines and social networks, grant free access to consumers.

straint on the distribution of access fees set by the platform, and elicitation from business users, as is already the case, as we show, for sponsored search and display ads.

Section 4 demonstrates that neither platform competition under consumer single-homing, nor app store competition triggered by regulations requiring Apple and Google to host competing app stores to encourage consumer multi-homing, solve the equity concern. Platform competition transfers some value from the platform to the consumer (the extent of such transfers is again limited by the core ZLB) but, provided platforms still control access to their consumers, does nothing to solve the equity concern for business users. App store competition is not effective if the dominant platform downlists multi-homing apps; but even if it is, the absence of access fee makes the platform too app friendly, because app stores will compete for superior app sales; put differently, app store competition requires levying the optimal access fee b, this time from the alternative third-party app stores rather than from now-disintermediated third-party business users. Section 5 reviews the relevant literature, and Section 6 concludes.

A benefit from our framework is that, despite its simplicity, it accounts for the rich diversity of digital environments. Table 1's illustrations are mostly drawn from the DMA's designated gatekeeper list.[4]

	App ZLB binding ($a < b$) Potential for self-preferencing	App ZLB non-binding ($a > b$) Potential for excessive fees
Core ZLB binding (no subscription fee)	• Limited gatekeeping ability ○ 3rd party payments $\Rightarrow a = 0$ (Epic's, Spotify's attempts at disintermediation) ○ Organic search $\Rightarrow a = 0$ (Google shopping) ○ App-store competition $\Rightarrow a = 0$ (if effective!) • Regulatory cap (DMA?) • Freemium model • Uniform access fee/heterogenous benefit	• App stores, in-app purchases ($b > 0$) • E-commerce ($b < 0$) ○ Amazon (hybrid player) ○ Booking (pure players) • Ad-supported media/services ($b < 0$) ○ Facebook, TikTok (display ads) ○ Google (sponsored search ads)
Core ZLB non-binding (positive subscription fee or physical device)	• Vertically integrated app-store	• Videogame platforms ($b > 0$)

Table 1. Taxonomy of digital platforms.

1. Low access fees. The left column captures alternative drivers of small (often zero) access fees, implying that the platform makes little or no money on access and therefore is incentivized to self-preference. First, the platform may have limited gatekeeping ability.

[4]These illustrations include all the DMA's 7 designated gatekeepers and most of their 24 core platform services.

This is the case if apps can disintermediate, making it impossible to levy a revenue-based fee. Such disintermediation is at the core of the recent Epic and Spotify high-profile cases. Similarly, in the case involving an alleged demotion of competing shopping services by Google's search engine, Google could not make money on access while it could profit from the consumer's using Google Shopping.[5] The third illustration of a limited gatekeeping capability is provided by app-store competition when the consumer multi-homes across competing app stores (this illustration is for the moment theoretical, as the EU still has to effectively force Apple and Google to open their operating systems to alternative app stores). This example is less of a concern than the previous two, since self-preferencing, if known, will lead consumers to choose alternative app stores.

Besides disintermediation, regulatory caps (e.g., a free-access requirement) is a second factor of low access fees. A third and more subtle one is provided by the popular freemium model, if the independent app is discovered (in its free version) on the platform, but upgrades are purchased through a direct channel. Finally, an incentive for self-preferencing exists on a subset of app markets when the platform charges a uniform fee (unconstrained by the previous factors), but ancillary benefits are heterogeneous across app markets; even tough a will then be high, it is still the case that $a < b$ for some app markets.

2. Unconstrained fees. The right column in contrast illustrates gatekeepers that make more money on access than on supplying the app itself ($a > b$). The concern then is that the platform selects excessive fees. In some illustrations, apps are digital and have positive ancillary benefits ($b > 0$). In other illustrations, a physical good or service creates a positive cost for the merchant ($b < 0$). This is the case for e-commerce platforms, whether hybrid like Amazon or pure player like Booking, and for ad-supported media/services, whether they get revenue from display ads (Facebook, TikTok, LinkedIn) or sponsored search ads (Google).

3. Binding core ZLB? Last, many (although not all) platforms provide access to consumers for free: Access or self-provision of services make consumers sufficiently profitable on the business side and so the platform benefits from attracting the maximum number of consumers. But some platforms do not face a core ZLB. A device may be sold at a loss (videogame consoles) or at a profit (Apple's iPhone), but in either case the price is unconstrained and the platform always wants to raise it to reflect a more valuable ecosystem. Because of the potency of the see-saw effect in this context, excessive fees are less of a concern than when the core ZLB binds, though they still inefficiently hamper innovation in the app segment.

[5] Similarly, highly visible apps that consumers can access through the browser for free (connecting to one's bank, Deliveroo, Amazon), known as hybrid apps, also exhibit $a = 0$. The app store then has a limited gatekeeping ability.

2 Access pricing by a gatekeeping platform

2.1 Basic framework

Consider a two-sided digital platform (e.g., an app store) that connects sellers of digital goods (hereafter, apps) with a mass 1 of consumers. Digital goods entail negligible marginal costs of production and distribution. Rather, their usage by consumers brings ancillary benefit $b \geq 0$ for the app providers, such as advertising revenues, consumers' data that can be monetized, fees collected from merchants selling their products through the app, etc. Hence, their opportunity cost is negative. App providers face a zero lower bound constraint because negative prices are subject to arbitrage: Bots and uninterested consumers may take advantage of the payment for usage, and yet bring no profit for merchants and advertisers and provide valueless data.[6]

Multiple (a mass 1 of) app markets coexist on the platform. We will analyse a representative app market for expositional simplicity; the multiplicity of app markets only serves to better motivate the "platform pivotality" assumption below, but is otherwise inessential. In the representative app market, two apps compete for the platform's customers.[7] Consumers have a unit demand each, and derive utility v_i when using app $i = 1, 2$, with $v_1 = v$, and $v_2 = v + \Delta$ if the app 2 provider sinks an investment cost $\gamma > 0$ and $v_2 = v$ otherwise. The development cost γ is drawn from a smooth distribution $G(\gamma)$, with density $g(\gamma)$ and increasing hazard rate $G(\cdot)/g(\cdot)$ on $\mathbb{R}^+$, and it is privately observed by the app 2 provider.

Welfare-detrimental platform behavior can occur only when there is a single superior app, and this app is owned by an independent developer. Therefore, without loss of insight for the equity question, we assume that only an independent provider can develop a superior app. The platform in contrast may either own app 1 (the platform is then hybrid), or be a pure platform (app 1 is also independently owned). The platform business model is endogenous, and vertical integration takes place by introducing another inferior app, which is able to drive app 1 out of the market.[8]

We consider a gatekeeping platform that charges consumers a fixed access price and adopts an agency business model: third-party app providers pay a unit access fee $a \geq 0$

[6] Alternatively, non-negative price constraints arise because of technical difficulties in operationalizing negative prices – as in the following quote from the Stigler Report (p. 30) *"It is possible that a digital market has an equilibrium price that is negative; in other words, because of the value of target advertising, the consumer's data is so valuable that the platform would pay for it. But the difficulty of making micropayments might lead a platform to mark up this negative competitive price to zero."*

[7] There can be more than two apps, as when there is a fringe of previous generation (i.e., inferior) apps. The case where there is only one app is captured in our notation below by $v = 0$.

[8] If both apps are owned by third-party providers, it is without loss of generality to assume that only app 2 can innovate: since Bertrand competition among symmetric apps dissipates profits, at most one app innovates in a pure-strategy equilibrium. The results are unchanged if the platform vertically integrates by making a take-it-or-leave-it takeover offer to app 1's provider; making a takeover offer to a superior app instead is never profitable for the platform.

for distributing their apps and set their prices.[9] Figure 1 depicts the (most interesting) case in which the platform operates a hybrid marketplace and app 2 sells a superior version.

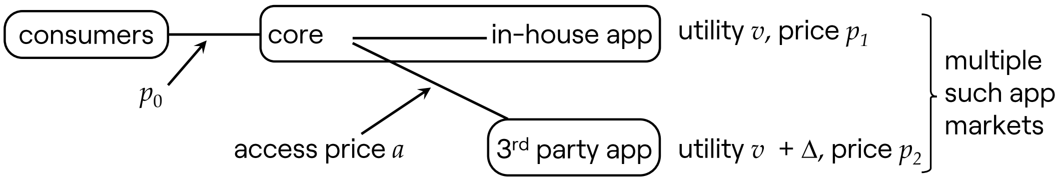


Figure 1: Two-sided market.

Let $x_i = 1$ if consumers buy app $i = 1, 2$, and $x_i = 0$ otherwise. To avoid the standard "openess problem," we assume that consumers buy an app when indifferent between doing so and not buying at all and, if $v_2 = v + \Delta$, they select the superior app when indifferent between the two apps. When instead $v_1 = v_2$, we make the standard tie-breaking assumption that consumers select both apps with the same probability in case of price equality (this assumption is not crucial to the results). In the following, p_0 denotes the consumers' access price to the platform and p_i the price of each app i. The profit of app provider $i = 1, 2$ (or division if owned by the platform) is

$$\pi_i \equiv x_i(p_i + b - a).$$

A pure-player platform's profit is

$$\pi_0 = p_0 + a(x_1 + x_2),$$

whilst a hybrid platform's profit can be written as the profit it would make as a pure platform, plus the extra profit its app division makes if it captures the app market – i.e., it is given by $\pi_0 + \pi_1$.

Left unmonitored, the platform can, if it wants, make apps less attractive (e.g., through downlisting). It then chooses $\{\delta_i \geq 0\}_{i=1,2}$ so that the value of app i for consumers becomes $v_i - \delta_i$. Strict inequalities correspond to non-price foreclosure.[10] In the case of a hybrid platform, self-preferencing corresponds to policy $\{\delta_1 = 0, \delta_2 > 0\}$. We will employ "non-price foreclosure" and "self-preferencing" indifferently in the hybrid context. The concept of "non-price foreclosure" is broader – to the extent that the hybrid

[9] Appendix B.2 shows that, provided $\Delta > b$, our insights are unchanged considering instead ad-valorem access fees. The reason is that with an ad-valorem fee a pure-player platform can capture Δ, which the superior app passes through to consumers, but cannot capture b. Therefore, if $\Delta < b$, vertical integration combined with self-preferencing is always optimal.

[10] When the platform finds it optimal to foreclose, it will reduce the attractiveness of a single app so that the other app is selected by consumers. In practice this can be accomplished by downlisting the foreclosed app. Hence, our large set of foreclosing options involves no loss of generality.

platform won't handicap its own app – as it applies also to the pure-platform case. In contrast, regulatory monitoring of equal access forces the platform to select $\delta_i = 0$ for $i = 1, 2$.

As summarized in Figure 2, the timing is as follows: (1) The access fee a is set, by the platform or by regulation; (2) app 2 developer privately observes its development cost $\gamma \sim G(\cdot)$ and decides whether to introduce a superior version of its app; (3) the platform decides whether to vertically integrate and to foreclose the third-party app(s); (4) the platform and the apps simultaneously set prices $\{p_0, p_1, p_2\}$;[11] finally, (5) consumers choose whether to patronize the platform and, if so, which app to purchase.

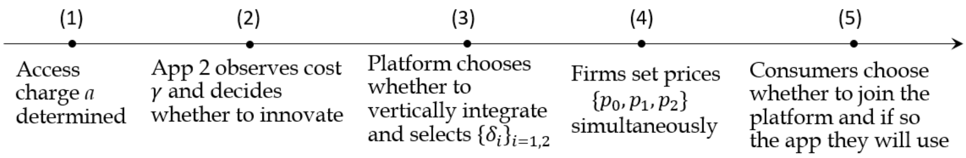


Figure 2: Timing.

The simultaneity of price choices may give rise to a multiplicity of pure-strategy equilibria, as is familiar in Nash demand games – e.g., in the literature on tying (Choi and Stefanadis, 2001, Carlton and Waldman, 2002). We will make the reasonable assumption that the platform is "pivotal" for consumers' participation:

Definition (platform pivotality). An equilibrium of the pricing subgame (stage (4)) exhibits platform pivotality if a third-party app maximizes its profit taking the mass of consumers present on the platform as given; that is, independent app providers do not perceive themselves as pivotal for the consumers' decision of whether to join the platform.

With many app markets, the platform pivotality assumption is innocuous when the platform faces a downward-sloping demand (as in Section 2.4.1) because the value and price of an individual third-party app in one app market has a negligible impact on consumers' overall utility from access to the platform. Unless otherwise stated, we will focus on the following equilibria (insights are not much affected by this focus; Appendix B.1 characterizes all equilibria):

Definition (equilibrium). An equilibrium of the pricing subgame is a set of pure strategies that (i) are undominated and (ii) satisfy platform pivotality.

The equilibrium profit of app i provider if there is no foreclosure is denoted by $\pi_i^*(a)$. Let us introduce some further definitions:

[11] That a is set first is natural if this access fee is regulated. Under laissez-faire, the platform needs to commit to a for some time to attract sellers (in reality, access fees charged by major digital platforms are stable over time). Put differently, the timing allows the platform to squeeze a superior third-party app, but not to hold it up. As standard tie breaking conditions, we suppose that in case of indifference, app 2 decides to introduce the superior version, whereas the platform does not vertically integrate or foreclose.

Definition (competitive neutrality). The access fee a is competitively neutral in a range $[\underline{a}, \bar{a}]$ if, in this range, (i) the platform has no incentive to use non-price foreclosure (even if it can), and (ii) the equilibrium profits and the allocation $\{x_i\}_{i\in\{1,2\}}$ are independent of a over the range.

Definition (fairness and squeeze). If the platform does not foreclose it, a superior app receives its fair share of its contribution to the ecosystem if $\pi_2^*(a) = \Delta$ and it is squeezed if $\pi_2^*(a) < \Delta$.

While at this stage the notions of "fair reward" and "squeeze" are only definitions, we will show that the fair values of the access fee maximize social welfare.

Definition (zero lower bounds). The app i's zero lower bound (ZLB) binds if $p_i^* = 0$. The core ZLB binds if $p_0^* = 0$.[12]

The remainder of this section is organized as follows. Section 2.2 characterizes the subgame-perfect equilibrium for any level of the access fee. Section 2.3 characterizes and contrasts privately and welfare optimal levels of the access fee. Extending the model to heterogeneous consumers' valuations for the platform's core services and the superior app's quality, Section 2.4 derives a Pigouvian rule for access fee regulation.

2.2 Impact of the access fee

2.2.1 Equilibrium prices in the absence of self-preferencing

We first assume that non-price foreclosure is prevented through regulatory monitoring. Note that, in both cases of a pure and a hybrid platform, both apps have the same opportunity cost of selling $(a - b)$ and do not charge below this opportunity cost in an equilibrium in undominated strategies, except an in-house app when constrained by consumers' willingness to pay for it. When it sells a superior version, the non-pivotal third-party app charges consumers a mark-up Δ over the price of the inferior app, unless it is constrained by consumers' willingness to pay. Therefore, in any subgame following app 2's innovation, the unique equilibrium outcome (absent non-price foreclosure) takes four configurations as a increases:

(1) *Muted app competition.* When the opportunity cost is negative $(a - b < 0)$, app 1 cannot charge an app price p_1 below 0 due to the app ZLB and therefore sets $p_1^* = 0$, while app 2 is priced at $p_2^* = \Delta$. The superior app does not feel the full competitive pressure from app 1, and so makes supranormal profit $(\pi_2^*(a) = \Delta + (b - a) > \Delta)$. The consumers obtain surplus $v > 0$ and so $p_0^* = v$.

(2) *Access fee neutrality.* If the opportunity cost is non-negative $(a - b \geq 0)$, the standard Bertrand equilibrium outcome in undominated strategies has $p_1^* = a - b$ (the

[12] If app 2 introduces a superior version, $p_2^* > 0$ always holds in the basic model: a "superior app ZLB" is never binding). This ZLB may bind in extensions of the basic model, for instance if apps adopt a freemium business model, or the demand for a superior app is downward sloping (Section 2.4.2).

app ZLB does not bind) and $p_2^* = (a-b)+\Delta$, so that $\pi_2^*(a) = \Delta$ (fair reward): A change in the access fee increases one-for-one both apps' opportunity cost. Access fee pass-through is feasible in our model as long as consumers keep purchasing the app; using the fact that consumers are in equilibrium indifferent between the two apps, $p_0^* = v - (a-b)$, and so it must be that $a - b \leq v$. By making charging the consumer or app 2 for access perfect substitutes, pass-through causes the access fee to be neutral in this region. The neutrality region exhibits the familiar "see-saw property" of two-sided-market theory, in which an increase in the merchant fee translates (in our case one-for-one) into a decrease in the consumer fee.

(3) *Superior app's squeeze.* When $a > b + v$, an inferior third-party app is not viable and the superior app can no longer apply mark-up Δ over the opportunity cost: $p_2 = (a-b) + \Delta > v + \Delta$. So, to sell to consumers on the platform it must lower its mark-up and sell at $p_2^* = v+\Delta$. Because the consumers do not benefit from apps, the core ZLB is binding ($p_0^* = 0$) and the access fee is no longer neutral, as app 2 developer must absorb its increase to keep customers. App 2 developer's margin is squeezed ($\pi_2^*(a) < \Delta$, with $\pi_2^*(a)$ strictly decreasing in a and $\pi_2^*(b+v+\Delta) = 0$), and the platform appropriates, at least in part, the superior app's contribution to the ecosystem.

(4) *Superior app's exit.* When $a > b+v+\Delta$, the superior app would have to sell at a price below the fee paid to the platform minus the ancillary benefit. It is then excluded from the app market. Such price foreclosure benefits neither the gatekeeper nor app 2 developer.

Lemma 1 (retail prices in the absence of self-preferencing). *Suppose that app 2 sells a superior version and the platform cannot use non-price instruments to foreclose. Whether the platform is pure or hybrid, the equilibrium outcome is unique. Because consumers are homogeneous, their surplus is extracted ($p_0^* + p_2^* = v + \Delta$) and $\pi_0^*(a) + \pi_2^*(a) = b + v + \Delta$ (and $\pi_1^*(a) = 0$) for all $a \leq b + v + \Delta$. Furthermore,*

(1) *when $a < b$:*

$$\begin{cases} \textit{App ZLB: } p_1^* = 0 \quad \textit{and} \quad p_2^* = \Delta, \\ \textit{Supranormal app profit:} \quad \pi_2^*(a) = \Delta + (b-a) > \Delta, \end{cases}$$

(2) *when $b \leq a \leq b + v$:*

$$\begin{cases} \textit{Passthrough: } p_1^* = a - b \quad \textit{and} \quad p_2^* = p_1^* + \Delta, \\ \textit{Fair reward:} \quad \pi_2^*(a) = \Delta, \end{cases}$$

(3) *when $b + v < a \leq b + v + \Delta$:*

$$\begin{cases} \textit{Core ZLB } (p_0^* = 0)\textit{: } p_2^* = v + \Delta \ (\textit{and } p_1^* = v \textit{ in hybrid platform case}), \\ \textit{Superior app's squeeze:} \quad \pi_2^*(a) = b + v + \Delta - a < \Delta, \end{cases}$$

(4) *when* $a > b + v + \Delta$*:*

$$\begin{cases} \textit{Superior app's exit: } p_1^* = v \textit{ in hybrid platform case; } \pi_i^*(a) = 0\,\forall i \textit{ in pure platform case,} \\ \textit{Price foreclosure: } \quad \pi_2^*(a) = 0. \end{cases}$$

Proof of Lemma 1. Consider a pure or hybrid platform. Suppose that app 2 is viable ($a \leq b + v + \Delta$), hence it serves consumers in equilibrium. Prices p_1 above $a - b \geq 0$ cannot be equilibrium prices. (a) Either the superior app is not constrained by users' willingness to pay ($p_2 < v + \Delta$) and then $p_2^* = p_1 + \Delta$ by platform pivotality. The consumers are indifferent between the two apps and so app 1 could gain $a - b > 0$ by slightly lowering its price. By this reasoning, the app ZLB is binding for $a < b$. (b) Or, if the superior app is constrained by users' willingness to pay ($p_2 = v + \Delta$), then $p_1 \geq v$ for app 2 to win the market and so app 1 is out of the market. This is the case if and only if $a - b > v$: Otherwise app 1 could charge $p_1 = a - b(+\varepsilon)$, take the market and gain relative to the presumed equilibrium behavior. In this region of the access fee, the in-house app's undominated price is $p_1^* = v$, because it ensures the ecosystem viability if app 2 unexpectedly charges $p_2 > v + \Delta$. Similarly, if app 2 is not viable ($a > b + v + \Delta$), $p_1^* = v$ because any lower price would reduce the hybrid platform's chances of gaining more than the price foreclosure profit through the access fee if app 2 unexpectedly charges $p_2 \leq v + \Delta$. Conversely, prices p_1 strictly below $a - b$ are dominated by price $a - b$ irrespective of app 1 ownership for all $a \leq b + v$, because $p_1 < a - b$ would make app 1 regret having won the consumer if app 2 charged an unexpectedly high price. Finally, $p_0^* = v_{i^*} - p_{i^*}^*$ where $i^* \in \{1, 2\}$ is the app consumers patronize in equilibrium. □

The same arguments imply that, if the app 2 provider does not introduce a superior version (i.e., $v_1 = v_2$), perfect Bertrand competition in the app market (constrained by the app ZLB) implies equilibrium prices $p_1^* = p_2^* = \max\{a - b, 0\}$ whenever the apps are not constrained by consumers' willingness to pay ($a - b \leq v$); for $a > b + v$, third-party apps are not viable whereas a hybrid platform serves consumers at $p_1^* = v$ – i.e., the equilibrium prices are as in Lemma 1 with $\Delta = 0$.

Remark (the link between ECPR and the ZLBs). The equilibrium characterization in Lemma 1 unveils a simple connection between the ZLBs and Baumol and Willig's ECPR rule for a vertically integrated firm providing access to a rival:

Definition (ECPR level). The access fee is below (equal to, above) the Baumol-Willig efficient component pricing rule level if a is smaller than (equal to, higher than) the unit profit, $p_1 + b$, lost by the hybrid platform when a third-party app attracts a consumer.

Corollary 1 (ECPR). *In equilibrium, the access fee is*

- *below the ECPR level* ($a < p_1^* - (-b)$) *if and only if the app ZLB binds* ($a < b$)*;*
- *at the ECPR level* ($a = p_1^* - (-b)$) *if and only if no ZLB binds* ($b \leq a \leq b + v$)*;*

- *above the ECPR level ($a > p_1^* - (-b)$) if and only if the core ZLB binds* ($a > b + v$).

2.2.2 Platform business model and self-preferencing

Let us augment the platform's strategy space by letting it engage in non-price foreclosure (the platform is left unmonitored), and consider its business model choice.

Proposition 1 (vertical integration and self-preferencing). *The platform benefits from vertical integration combined with self-preferencing if and only if $a < b$.*

Proof of Proposition 1. First, we argue that the platform has no incentive to engage in non-price foreclosure ($\delta_i > 0$ for some $i = 1, 2$) when both apps are independent. This is clearly the case if both apps have value v, as symmetric competition keeps prices at the lowest possible level and reducing both apps' values to the same extent leaves app prices unchanged but implies that the platform needs to reduce p_0. If app 2 sells a superior version, suppose without loss of generality that $v_2 - \delta_2 > v_1 - \delta_1$. If app 2 is not constrained by consumers' willingness to pay, an increase in δ_2 has no impact on consumer surplus from apps as it decreases the superior app's markup and its value for the consumers by the same amount. In contrast, an increase in δ_1 reduces app competition and hurts the consumers, so the platform could reduce δ_1 and raise p_0 to the same extent (keeping consumer membership constant), thereby increasing its profit. When app 2 is constrained by consumers' willingness to pay, $p_0^* = 0$ and so the platform's profit equals $\pi_0^* = a$ for all (δ_1, δ_2).

Second, consider a hybrid platform. Without loss of generality, we can assume that $\delta_1 = 0$ and either $\delta_2 = 0$ (no foreclosure) or $\delta_2 = v_2$ (full foreclosure). Intuitively, the platform's choice determines which among the in-house and third-party apps the consumers will select. In the former case, making the third-party app worthless involves no loss of generality. In the latter case, as seen above, picking $\delta_2 > 0$ does not benefit the platform. When foreclosing, the platform can achieve profit $v + b$ – i.e., the value it creates on a stand-alone basis. When not foreclosing, if the third-party app sells a superior version, the platform makes profit $\pi_0^*(a) = v + \min(a, b)$ if there is no squeeze and more when there is a squeeze; if instead $v_2 = v$, the platform makes profit $\pi_0^*(a) + \pi_1^*(a) = v + [\min(a, b) + b]/2$. Therefore, in either case, a hybrid platform engages in self-preferencing if and only if $a < b$.

Finally, as seen in Section 2.2.1, app 1's ownership does not affect the prices that consumers pay in the equilibrium without foreclosure. Hence, provided app 2 is viable, vertical integration is profitable if and only if the hybrid platform has incentives to engage in self-preferencing – i.e., for all $a < b$. Note that the platform also finds it optimal to vertically integrate when app 2 is not viable – i.e., for all $a > b + v$ (resp. $a > b + v + \Delta$) if $v_2 = v$ (resp. $v_2 = v + \Delta$) – to avoid the Pareto-dominated price-foreclosure outcome.[13] □

In a nutshell, for $a < b$ (i.e., when the app ZLB binds), a hybrid platform does not have enough skin in the game to want to give access to its rival. Then, vertical integration enables the platform to reap the benefit $b - a$ from foreclosing the third-party app provider. Without taking a stance on their desirability, we note that existing (GDPR, DMA) and forthcoming regulations aim at restricting the use of data and thereby reduce the ancillary benefit b. Such a decrease in b reduces the incentive for self-preferencing, keeping the access fee constant.

Extending the model to app-provider-specific ancillary benefits, Appendix B.3 shows that the end of current arrangements under which the platform shares with its apps their data would increase the incentive for self-preferencing, as this scenario reduces the ancillary benefit for the platform only conditional on a third-party app making a sale, increasing its opportunity cost to let it serve consumers. Such a move would thus have to be accompanied with increased regulatory monitoring and/or reduced regulatory pressure on access-fee setting.

2.2.3 App development

The app 2 provider introduces a superior version of its app if and only if the extra-profit it makes by offering a superior app covers its investment cost γ. Absent innovation, app 2 always makes zero profit, as either it is foreclosed or it engages in perfect Bertrand competition with positive opportunity cost. Figure 3 depicts how the superior app's profit varies with the access fee; the innovation is introduced whenever this profit exceeds the investment cost.

Lemma 2 (innovation). *The innovation always takes place when socially optimal if and only if the access fee lies in the competitive neutrality region $a \in [b, b+v]$, at which a superior app receives its fair share of its contribution to the ecosystem.*

Proof of Lemma 2. The app 2 provider introduces a superior version of its app if and only if $\pi_2^*(a) \geq \gamma$ and the platform has no incentive to foreclose it. Therefore:

- For $a < b$, the anticipation of foreclosure gives the app 2 provider no incentive to innovate for any $\gamma > 0$.

- For $a \in [b, b+v]$, the innovation takes place if and only if it is socially optimal – i.e., $\Delta \geq \gamma$.

[13] Alternatively, for $a \in (b+v, b+v+\Delta)$ this outcome can be avoided by a hybrid platform by committing to innovation-contingent access fees or engaging in ex-post Pareto-improving renegotiation, capping a at $b+v$ if app 2 does not sell a superior version. In practice, this can be accomplished by charging a lower fee to a third-party seller whose revenue does not exceed a certain threshold ($b+v+\Delta$ in our model), as the major app stores actually do.

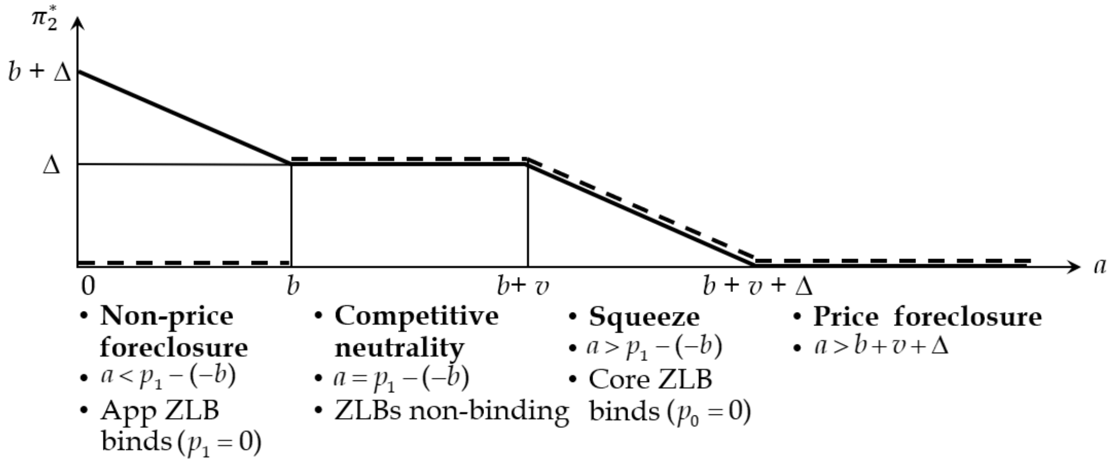


Figure 3: Superior app's profit (gross of the development cost γ). The dashed line represents the profit when self-preferencing is feasible and the full line the profit when it is not.

- For $a \in (b+v, b+v+\Delta]$, a socially optimal innovation with development cost $\gamma \in (b+v+\Delta-a, \Delta]$ would not be undertaken.
- For $a > b+v+\Delta$, a superior app is not viable and so the app 2 provider does not innovate for all $\gamma > 0$. □

An independent developer's innovation creates social value Δ while entailing a cost γ. Therefore, the innovation has positive net value if and only if $\Delta \geq \gamma$. As app 2 has private incentives to innovate if and only if $\pi_2^*(a) \geq \gamma$, guaranteeing a third-party provider a fair reward for its contribution to the ecosystem is needed to ensure the socially efficient level of innovation. By contrast, both the anticipation of price or non-price foreclosure and of margin squeezes inefficiently dampen a third-party developer's incentive to innovate.

2.3 Platform-optimal vs welfare-optimal access pricing

We define social welfare as the sum of consumer net surplus and the firms' profits, net of app 2 developer's cost of innovation.

Proposition 2 (optimal access fees).

(i) Profit-maximizing access fee. *The platform's profit is maximized at* $a^* \in (b+v, b+v+\Delta)$, *given by*

$$a^* = b + v + \frac{G(b+v+\Delta-a^*)}{g(b+v+\Delta-a^*)}.$$

(ii) Welfare-optimal access fees. *Social welfare is maximized for* $a \in [b, b+v]$, *i.e. if and only if a superior app receives a fair reward for its contribution to the ecosystem.*

Proof of Proposition 2. (i) No matter whether app 2 sells a superior version, for $a < b$ the platform makes profit $v + b$ by vertically integrating and foreclosing app 2, which coincides with its profit in the neutrality region $a \in [b, b + v]$ and in the price-foreclosure region $a > b + v + \Delta$. For $a \in (b + v, b + v + \Delta)$, innovation takes place with probability $G(b + v + \Delta - a)$, in which case the platform gets a, else the platform gets the price-foreclosure profit $v + b$ through vertical integration. Hence, the platform's expected profit is $v + b + G(b + v + \Delta - a)[a - (v + b)]$, which is maximized at $a^* > b + v$ characterized above.

(ii) Consumer surplus is always extracted by the platform through the core price and, as the platform can always ensure the ecosystem viability through vertical integration, ex-post social welfare is maximized whenever there is no price ($a > b + v + \Delta$) or non-price foreclosure ($a < b$) of a superior app. While margin squeeze, occurring for $b + v < a < b + v + \Delta$, has a purely redistributional effect conditional on a superior app being introduced, it reduces welfare by yielding suboptimal investments (Lemma 2). As a result, ex-ante social welfare is maximized if and only if $a \in [b, b + v]$, which correspond to the levels of the access fee at which a superior app developer receives its fair reward (Lemma 1). □

A squeezed third-party app provider has a suboptimal incentive to develop its app. The impact of the access fee on the richness of the ecosystem is accounted for by the platform, but incompletely so. As a result, an inefficiently low amount of innovation takes place under laissez-faire. Capping, by regulation, the access fee to any level in the competitive neutrality region (i.e., $a \in [b, b + v]$) is needed to maximize social welfare. Considering independent developers' innovation incentives thus unveils a natural link between fair access pricing and welfare maximization.

Remark (is there a trade-off between consumer surplus and innovation?). The call for third-party developer reward Δ was made from the point of view of efficiency/welfare maximization. A consumer standard might seem to lead to a social demand for some "taxation" of innovation in the form of a squeeze on app profits, provided that the increase in access fee is passed through to consumers via a reduction in the core price. But there is here no trade-off between consumer and innovator surplus because there is no pass through (the squeeze region coincides with the core ZLB one); and so the consumer standard leads to no specific conclusion on the innovation front.

This conclusion is robust to platform or app store competition (Section 4), because the core ZLB again binds and so there is no pass-through to consumers. Conversely, by raising the price of the superior app, high access fees can trigger an excessive substitution toward the inferior app or reduce consumer participation on the platform, resulting in consumer harm (Section 2.4). As we shall see, in all these environments there is no competitive neutrality region and so the unique welfare-maximizing access fee is $\hat{a} = b$.

Remark (monitoring of self-preferencing and excessive innovation). The welfare-optimal access fees are $a \in [b, b+v]$ even if self-preferencing is prevented through regulatory monitoring. In this case, the supranormal profit made by a superior third-party app implies that it has socially excessive incentives to innovate for $a < b$: Because as app 1 is selected by one half of consumers when app 2 does not introduce the superior version, a socially inefficient innovation with development cost $\gamma \in (\Delta, \Delta + (b-a)/2]$ would be introduced.

One may in general be suspicious of concerns about excessive innovation. Yet, this possibility is natural in the digital economy, precisely because of the existence of an ancillary benefit $b > 0$. A me-too innovation in the app segment, bringing along a small improvement $\Delta = \varepsilon$ in app quality, allows the innovator to corner the app market.[14]

Remark (Physical goods and devices). Our analysis also applies to platforms hosting sellers of physical products (e.g., e-commerce platforms such as Amazon and eBay) or services (e.g., OTAs such as Booking and Expedia) that entail positive marginal costs c_a. With ancillary benefit $b^\dagger > 0$, and letting $b \equiv b^\dagger - c_a$, the region where a superior app makes supranormal profits (if the platform is a pure-player or self-preferencing is monitored) or is foreclosed is empty whenever the adjusted ancillary benefit is negative ($b < 0$). The platform-optimal level of the access fee is still a^*, which strictly exceeds the welfare-optimal levels $a \in [\max\{b, 0\}, b+v]$.

The analysis also generalizes to accommodate devices (smartphone, laptop, or game console) needed to connect to the platform. Assume that the device brings stand-alone value v_d, the same for all consumers,[15] and let c_d denote its production cost. If the device is "cheap" relative to its stand-alone value, that is if $v_d \geq c_d$, consumers own it regardless of whether there is a competitive original equipment manufacturers (OEM) sector or a monopoly platform vertically integrated into device manufacturing, and so the foregoing analysis is literally unchanged.[16]

In the case of a "costly" device, i.e. if $v_d < c_d$, the results depend on whether the platform is vertically integrated into device manufacturing (or, equivalently, it can subsidize device manufacturers). If the device (e.g., an Android-powered smartphone) is manufactured by a competitive OEM industry, it is sold at cost (i.e., at price c_d), implying that consumer surplus from the apps must exceed $c_d - v_d > 0$. For that reason, the equilibrium cannot be in the squeeze region. Conversely, when the device is produced by the platform (e.g., iPhone), the core ZLB can be circumvented by subsidizing the device: the platform

[14]Similarly, assuming that entry can occur at value proposition v (no innovation relative to the current generation), me-too entry will occur for $a < b$ as long as the entry cost is smaller than $(b-a)/n$, where n is the number of active apps: The ancillary benefit is thus dissipated.

[15]For example smartphones can be used for "non-gatekeeping purposes" such as taking pictures and making calls; likewise desktops have other usages than supporting services intermediated by a gatekeeper.

[16]This is the case also if v_d represents the value of platform's core services (independently of app consumption), as these are digital goods and so $c_d = 0 < v_d$.

can squeeze a superior app and bundle device and app store at a price v_d. As a result, allowing vertical integration into device manufacturing and (at least "some") squeeze may be necessary for the ecosystem viability (see Appendix B.4 for the complete analysis).[17]

2.4 Consumer heterogeneity and Pigouvian regulation

This section allows consumers to differ in the overall demand for the platform (Section 2.4.1) or in their demand for a superior app (Section 2.4.2).[18] Beside showing the robustness of our insights, these extensions introduce a meaningful distinction between the social and the consumer welfare standard, and derive a Pigouvian principle that underlies optimal access fee regulation in more general environments. Concretely, we show that the regulated access fee should coincide with the ancillary benefit associated with app distribution:

$$\hat{a} = b.$$

The reason why this can be interpreted as the Pigouvian level of the access fee is that b represents an unpriced externality that is internalized when app suppliers are charged $\hat{a} = b$. The proofs of the results of this section are omitted for brevity and can be found in Appendix A.

2.4.1 Elastic platform demand

Assume that consumers directly derive utility from the core service, independently of apps. Their willingness to pay for the core service, v_c, is heterogeneous, has wide support (in $\mathbb{R}$), and is distributed according to a smooth cdf $F(v_c)$ with density $f(v_c)$ and decreasing (inverse) hazard rate $\rho(v_c) \equiv [1 - F(v_c)]/f(v_c)$. A negative value of v_c corresponds to a learning or an opportunity cost.

If the platform is hybrid and there is no superior app (or if it is foreclosed), both the core and app ZLBs bind ($p_0^* = p_1^* = 0$) whenever

$$\arg \max_{\{p_0+p_1\}} (p_0 + p_1 + b)[1 - F(p_0 + p_1 - v)] \leq 0 \iff b \geq \rho(-v),$$

because the cutoff is given by $v_c^* + v = p_0 + p_1$. As this is a novel feature compared with the foregoing analysis (where at most one ZLB bound), we restrict attention to this region of parameters:

Proposition 3 (elastic platform demand). *Augment the basic model by adding a consumer utility from core services, v_c, distributed according to a smooth cdf $F(v_c)$ with*

[17] The possibility that some squeeze be needed to ensure viability applies not only to the case of costly physical devices but also when entry costs in the core segment are very large, and so the viability of the platform is not a foregone conclusion: see Appendix B.5.

[18] If consumers differ only in the ancillary benefit they provide to the app they patronize, the foregoing analysis remains unchanged, with b reinterpreted as the average per-consumer benefit.

density $f(v_c)$ in $\mathbb{R}$, and decreasing (inverse) hazard rate $\rho(v_c) \equiv [1-F(v_c)]/f(v_c)$. Suppose that $b \geq \rho(-v)$, so that the core and app ZLB both bind ($p_0^* = p_1^* = 0$) in the hybrid platform case when a superior app does not exist or is foreclosed.

(i) The platform engages in vertical integration combined with self-preferencing if and only if $a < b$.

(ii) Consumer surplus and social welfare are maximized at $\hat{a} = b$. In contrast, under laissez-faire, $a^* \geq \hat{a}$.

For all $a \leq b$, the unique equilibrium features the same app prices of the basic model, so that consumers are indifferent between the two apps, no matter whether app 2 sells a superior version. Therefore, letting a third-party app serve consumers does not increase their participation but reduces the platform's unit revenue for all $a < b$. As a result, the platform makes strictly more by vertically integrating and foreclosing any third-party app, which crowds out innovation incentives. Conversely, at $\hat{a} = b$ the platform has no incentive to foreclose and so a third-party app innovates if and only if $\Delta[1 - F(-v)] \geq \gamma$ – i.e., whenever it is socially optimal taking as given consumers' participation on the platform at the foreclosure level (a non-pivotal innovative app cannot expand consumer participation, as it would never charge $p_2 < \Delta$).

For $a > b$, a continuum of equilibria that satisfy conditions (i)-(ii) of our equilibrium definition exist in the hybrid platform case. In particular, there exists a strictly increasing function $\bar{p}(a)$, with $\bar{p}(b) = 0$ and $\bar{p}(a) < a - b$ for $a > b$, such that for any $p_1 \in [0, \bar{p}(a)]$, $\{p_1, p_2^* = p_1 + \Delta\}$ is an equilibrium whenever $p_2 \leq v_2 = v + \Delta$; the core ZLB binds ($p_0^* = 0$) in all equilibria.[19] Yet, any of these equilibria yields strictly lower social welfare relative to the equilibrium emerging for $\hat{a} = b$: Either $p_1 = 0$ and so a superior third-party app's margin is squeezed ($p_2^* + b - a = \Delta + b - a < \Delta$ for all $a > b$), which benefits the platform (hence, $a^* > \hat{a}$) but implies inefficiently low innovation incentives; or app prices are increasing in the access fee and, as the binding core ZLB prevents such price increases to be offset by a lower consumer access price, also consumer participation is inefficiently reduced. In the latter case, also consumer surplus is uniquely maximized at $\hat{a} = b$.

Remark (core ZLB and the rich ecosystem argument). The platform's incentive to provide a rich ecosystem rather than extract the business users' surpluses through squeezes and self-preferencing depends on whether, at the margin, the platform can monetize the ecosystem on the consumer side. The core ZLB is a simple and robust reason why such

[19] This is because the platform may want to set app prices p_1 that would be "too low" (dominated) from the point of view of the narrowly-construed app-market-i profit, as a low app price attracts more consumers to the platform. Thus, the concept of undominated strategy must be interpreted at the multi-product level for the hybrid platform, and it here fails to select a unique equilibrium. Such below-cost pricing would not emerge in equilibrium if the inferior app were independently owned, and it implies that app 2 is not viable if it does not sell a superior version.

monetization may be infeasible.[20] One might argue that platforms can break the core ZLB constraint by offering free software and services (Amelio and Jullien, 2012). Yet, if raising quality of service comes at an increasing cost, there is only partial relaxation of the constraint. Moreover, bundling may not relax the core ZLB at all, as an inferior in-house app commands a zero price regardless, and so it makes no difference whether it is bundled with the core or marketed separately.

2.4.2 Elastic demand for the superior app

This section introduces heterogeneity with respect to the perceived extra quality of a superior app.[21]

Proposition 4 (elastic demand for the superior app). *Suppose that consumers have heterogeneous valuations Δ for the extra value brought about by a superior app (distributed according to a smooth cdf $H(\Delta)$ with support $\mathbb{R}^+$ and a decreasing inverse hazard rate). The platform engages in vertical integration combined with self-preferencing if and only if $a < b$. Consumer surplus and social welfare are uniquely maximized at $\hat{a} = b < a^*$.*

By encouraging an excessive consumption of the in-house app, a hybrid platform's below-opportunity-cost pricing in the competitive segment, which emerges in the unique Nash equilibrium for all $a > b$,[22] harms both consumers and the superior third-party app provider, inefficiently weakening innovation incentives. As a result, optimal access fee regulation must follow the Pigouvian principle.[23]

[20] But it is not the only reason; suppose, e.g., that, in contrast with our specification, the inframarginal consumers value the app store more that the marginal ones. Suppose consumers get utility from two services: a non-platform one (say, pictures and calls for an iPhone) and, for a subset of them only, the app store. Then, a marginal improvement in app store quality does not induce the platform to increase its price on the device, as this improvement is valued only by inframarginal users. There is de facto a lower bound, with similar implications as a core ZLB, but it is not 0. This second reason for the absence of pass-through is reminiscent of Spence's (1975) observation that a monopolist's incentive to (over or under) supply quality depends on the relationship between the marginal and the average consumer's willingness to pay for quality, where "quality" in our context can be understood as "low superior app's price".

[21] With a continuum of app markets, we assume that for a given consumer Δ is the same across app markets.

[22] The mechanism is similar to the one at play in Chen and Rey (2012), who provide a rationale for loss leading in the retailing industry. By pricing the competitive good (that is, the in-house app) below cost, and raising the price for the monopolized good (that is, consumers' access price) accordingly, the platform: (i) maintains the total price charged to consumers with low (extra-) willingness to pay for the superior app (corresponding to one-stop shoppers in Chen-Rey), who buy the in-house app; (ii) increases the margin earned on those with higher willingness to pay, who buy the superior app (Chen-Rey's multi-stop shoppers) in the monopolized segment; and (iii) induces the superior app to reduce its price (hence, squeezes its margin). By contrast, a third-party seller would price app 1 above the opportunity cost. This implies that the platform always has an incentive to vertically integrate in order to exert competitive pressure on the superior app.

[23] Unlike in the previous versions of the model, here this result would hold even if the superior third-party app pre-exists the access fee policy; a lower access fee would instead be socially optimal if vertical integration is not an option or self-preferencing can be monitored, so as to further reduce double marginalization and spur innovation.

3 Implementation of the Pigouvian rule

Our analysis, which calls for Pigouvian regulation ($\hat{a} = b$), posits a representative app market.[24] Appendix B.1 shows that the theory easily generalizes to an arbitrary number of heterogeneous app markets: Letting the ancillary benefit, the inferior app value and the competitive advantage of the superior app in market k be denoted b^k, v^k, and Δ^k, respectively, the platform and welfare optimal access fees in each market k are[25]

$$a^{k*} = b^k + v^k + \frac{G^k(b^k + v^k + \Delta^k - a^{k*})}{g^k(b^k + v^k + \Delta^k - a^{k*})} > \hat{a}^k = b^k,$$

with $G^k(\cdot)$ and $g^k(\cdot)$ denoting the cdf and pdf of the development cost γ_k in market k. As is the case for optimal taxes in public finance, the theoretical benchmark – here the ancillary benefit obtained when the app acquires a customer – must be supplemented with an empirical methodology to measure the relevant data. A weak spot of the DMA is its limited guidance regarding both the theoretical benchmark and its implementation. It contains broadly scripted conditions[26] and alludes to FRAND (Fair, Reasonable and Non-Discriminatory) access fees. To go beyond such general statements, the regulator may either engage in information collection or elicit this information from the parties.

In the first approach, the regulator estimates the ancillary benefit b or the existence of an "unfair downlisting" (which requires measuring Δ). Measuring the ancillary benefit is the path taken in the EU for capping the merchant fees for card payments. The investigation of unfair downlisting (self-preferencing) has been undertaken in a few recent academic papers on Amazon's vertical integration.[27] Note that the detection of unfair downlisting speaks to the self-preferencing question, but does not address the excessive-fees one.

The heterogeneity of app markets however hinders either endeavor. For instance, app categories differ substantially in terms of the ancillary benefit their distribution generates: there are data-poor and data-rich markets – e.g., social media and food delivery apps

[24] Or, equivalently, a mass 1 of ex-ante identical app markets $k \in [0, 1]$, with the development cost γ_k being i.i.d. across markets. Indeed, by the law of large numbers, the probability $G(\pi_2^*)$ with which app 2 innovates in the representative app market (almost surely) equals the fraction of markets where the innovation is introduced, implying that expected profits and welfare are as in the basic model.

[25] Even under laissez-faire, the platform may not finely tailor the access fee to the specific app market (as we will see, an exception to this rule is search or display advertising). This is for at least two reasons. The first is the complexity cost: The platform would have to define individual app markets and estimate the profit-maximizing fee in each of them. The second is related to commitment: a very-fine-grained policy may discourage innovation in the app market (or equivalently the porting of apps to the particular platform). In such circumstances, the platform may prefer a uniform policy (such as the app stores' familiar 30% cut) to a fine-grained one. But, in the class of uniform fees, it is still the case that the platform's optimal fee exceeds the socially optimal one.

[26] E.g., *"The gatekeeper shall not engage in any behaviour that undermines effective compliance with the obligations of Articles 5, 6 and 7"* (Article 13(4)).

[27] See Farronato et al. (2023), Lee-Musolff (2024), and Waldfogel (2024) for recent studies, and Etro (2024) for a survey of some earlier studies.

sell much more personal data to third-party advertisers than videoconferencing apps.[28] The industry has private information about these values that is hardly available to the regulator. The alternative is to elicit the value of the benefit b from the industry, or possibly combine both approaches. This section explores these paths.

3.1 Eliciting the information from the platform

We first consider an elicitation of ancillary benefits from the platform. To examine how the regulator's limited knowledge of market-specific ancillary benefits affects access fee regulation in the simplest possible model, we consider the best-case scenario in which the regulator knows their cumulative distribution $K(b)$ in the population of app markets. A necessary condition for the Pigouvian access fee to be implemented in all markets is that the distribution of (observed) access fees be equal to the distribution of benefits.

Proposition 5 (impossibility of elicitation from a hybrid platform). *Suppose the regulator knows the distribution $K(b)$ of ancillary benefits and lets the platform choose $(a^k)_k$ subject to the constraint that the distribution of access fees mimics that of benefits (i.e., follows $K(a)$). Then, if self-preferencing cannot be monitored, setting $a^k = b^k$ for all k is not incentive-compatible for a hybrid platform.*

Proof of Proposition 5. Take two markets k' and k'' such that $v^{k'} \leq b^{k''} - b^{k'} < v^{k'} + \Delta^{k'}$. If the platform sets $a^k = b^k$ for $k \in \{k', k''\}$, in equilibrium it obtains profit $v^k + b^k$ from each of these markets: it obtains b^k from either the access fee or its in-house app distribution, and appropriates consumer net value from this app market, v^k, through p_0. By setting instead $a^{k'} = b^{k''}$ and $a^{k''} = b^{k'}$, and foreclosing the third-party app in market k'', the hybrid platform still obtains profit $v^{k''} + b^{k''}$ in the higher-b market (i.e., consumer net value plus benefit from in-house app distribution in market k''), but now makes a larger profit $a^{k'} = b^{k''} > b^{k'} + v^{k'}$ from the lower-b market with probability $G^{k'}(v^{k'} + \Delta^{k'} + b^{k'} - b^{k''}) > 0$, because in market k' app prices are in the squeeze region if app 2 innovates. □

If self-preferencing cannot be monitored, the Pigouvian rule is not implementable in all markets even if the regulator knows the distribution of b, and so can require that a and b have the same distribution $K(\cdot)$.[29] The reason is that, rather than charging $a^k = b^k$ in all markets, a hybrid platform can profitably charge higher fees in markets where b is lower, so as to squeeze superior third-party developers' margins in these markets, and foreclose

[28] See https://www.pcloud.com/it/invasive-apps.

[29] Whether the regulator observes (v^k, Δ^k) or not is immaterial for Proposition 5. Note also that the impossibility result holds a fortiori if the regulator sets a global access fee cap, which would be a less stringent regulation. If vertical integration is not an option, or under monitoring of foreclosure, the Pigouvian rule can instead be implemented by delegating fee setting to the platform under the constraint that the distributions of a and b be the same; but the assumption that the regulator knows the distribution of ancillary benefits is a strong one.

independent providers in markets where b is higher, where it is constrained to set lower fees, which allows it not to lose profit in these markets. Thus, under no monitoring of self-preferencing, market-specific fees cannot be enforced under asymmetric information, and the regulator faces a trade-off between preventing foreclosure of providers in high-b markets and allowing margin squeeze in low-b ones, implying that innovation incentives are always inefficiently dampened.

Remark (off-path measurement and appeals). Measuring b systematically to overcome this impossibility result would imply considerable costs and delays. At best one can allow appeals that hopefully will not be frequent if the incentive scheme is designed properly. Suppose that the platform chooses the access conditions $\{a^k, \delta_2^k\}$ (of which the regulator observes only a^k) and, after prices are set, a superior third-party app chooses whether to appeal "against a high access fee" (an inferior app 2 has nothing to gain from an appeal). In this appeal procedure, the authority observes a noisy, but unbiased, measure $\tilde{b}^k$ of the ancillary benefit, with cdf $R(\tilde{b}^k)$ such that $\int_{\mathbb{R}} \tilde{b}^k \mathrm{d}R(\tilde{b}^k) = b^k$. If $a^k > \tilde{b}^k$, then the access fee is assessed to be unfair, and the defendant (the platform) must pay a fine $\tau(a^k - \tilde{b}^k)$ (with $\tau > 0$) to the plaintiff (the third-party app); and vice versa if $a^k \leq \tilde{b}^k$.

The outcome of the appeal procedure interferes neither with the platform's choice of whether to foreclose the third-party app, nor with prices chosen by the firms before the appeal. This implies that the third-party app will appeal whenever $\int_{\mathbb{R}} \tau(a^k - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k) > 0 \Leftrightarrow a^k > b^k$. So if $\tau \geq 1$, the platform does not gain from inflating the access fee beyond b^k. The possibility to levy sufficiently large fines to the platform if it loses the appeal also discourages it from building a reputation for engaging in foreclosure *after* being challenged by an app: see Appendix B.6 for a detailed analysis.

3.2 Eliciting the information from business users

Section 3.1's impossibility result hinged on the assumption that fee setting is delegated to the platform. We now reverse the roles in access fee setting. Section 3.2.1 argues that an elicitation from third-party sellers is the norm for ad-supported platforms that award ad-slots through auctions, and shows that our analysis extends to display and search advertising. Building on these insights, Section 3.2.2 explores elicitation from third-party apps in our app store model.

3.2.1 Elicitation through auctions: Ad-supported platforms

When analyzing ad-supported media and services, one should think of ads as giving merchants access to the consumer. In this respect, Facebook or Google Search are not that different from Amazon or Booking. Our "app providers" can be third-party sellers advertising their products on the platform – e.g., display advertising on social media platforms (Facebook, TikTok) or search advertising on search engines (Google, Microsoft

Bing). An interesting feature is that the access fee, a, is here elicited from business users through an ad-auction rather than set by a platform or a regulator.

Display and search advertising can thus be modeled by slightly extending our basic setting. Consider N identical markets, indexed by $k = 1, \ldots, N$, each comprising two sellers whose product values are as in the basic model, and a unit mass of consumers. Each consumer is interested in only one product category[30] – i.e., derives values $(v_1^{k^*}, v_2^{k^*})$ from the products in market k^* and no value from products in other markets – which is unknown to the firms and possibly to the consumer as well: The platform showing ads (in the case of display advertising) and also the consumer conducting the search (in the case of search advertising) know only that the consideration set includes N markets, but cannot identify the desired product k^* prior to the consumer accessing the platform.

The timing goes as follows: (1) The platform commits to displaying consumers a number m of ad-slots; (2) seller 2 in each market decides whether to sink a cost γ to introduce a superior product (investment decisions are publicly observed); (3) slots are auctioned off by the platform through a *uniform price auction* (which ensures the Vickrey-Clarke-Groves outcome): the m highest bidders are selected and pay the highest losing bid per-click, hereafter $a(m)$; (4) the platform can engage in (vertical integration combined with) self-preferencing – i.e., as in the basic model, it can reduce consumers' perceived value of any third-party product; (5) the platform and the displayed sellers set their product prices; and (6) consumers observe their valuation and the prices of the m displayed products and decide whether to patronize the platform and which products to purchase – hence, each click leads to a sale.[31]

Note that imposing a reserve price equal to the platform-optimal access fee a^* in Proposition 2 would allow the platform to directly replicate the laissez-faire outcome of the basic model: even when the number of slots is not restricted (i.e., the platform displays all sellers who pay the reserve price), only superior sellers are willing to pay $a^* \in (v+b, v+b+\Delta)$ per-sale (therefore, they are able to charge consumers the monopoly price $p_2^* = v + \Delta$).

Even in the absence of reserve prices, however, the platform can squeeze innovative sellers by restricting the number of slots, as in Prat and Valletti (2022).[32] To see this, let $n \in \{0, ..., N\}$ denote the number of sellers that introduce the superior version. If $m < n$, competition for ad-slots among superior sellers yields $a(m) = b + v + \Delta$, and so they cannot recoup their investment cost. As a result, for all $m < N$ the unique symmetric equilibrium of the investment subgame is in mixed strategies: each seller 2 is

[30] The results also hold if each consumer is interested in all N markets.

[31] Our results hold irrespective of whether the pool of displayed sellers is common knowledge at the pricing stage, and immediately extend to any constant click-through and conversion rates.

[32] See also Ichihashi and Smolin (2024) for a rationing of ad-slots that is contingent on prices offered by sellers, and Janssen et al. (2024) for the assignment of products to sponsored positions and the obfuscation of the organic positions' informational content.

indifferent and invests with probability $\rho(m)$, increasing in m. By contrast, for $m = N$ any superior seller ensures a slot by outbidding inferior ones, who are willing to bid up to $a(m) = b + v$ per-sale for a slot. This outcome yields a fair reward (each superior seller gets Δ per consumer, because $p_2^* = v + \Delta$ when $a = b + v$), and so ensures the socially efficient level of innovation.[33] Yet, the platform might gain strictly more by setting some $m^* < N$: doing so implies that with some probability $m^* < n$ and so all superior sellers are fully squeezed; this *intensive margin gain* may offset the *extensive margin loss* from not showing to some consumers their desired product. The following result is proved in Appendix A:

Proposition 6 (ad-auctions). *Any superior seller receives its fair share for its contribution to the ecosystem if a slot is auctioned off without reserve price for each market ($\hat{m} = N$). In contrast, the platform may find it profitable to restrict the number of available slots (i.e., it auctions $m^* \leq N$ slots).*

Note that while the Pigouvian outcome is achieved for $\hat{m} = N$, the resulting access fee is now $a = v + b$ rather than b. This is because seller competition is now *for* the market rather than *in* the market. As a result, both under laissez faire and in the socially optimal outcome, all displayed sellers charge the monopoly price ($p_i^* = v_i$) and so the core ZLB binds ($p_0^* = 0$), as is typically the case for ad-supported platforms.

In advertising markets, the challenge for regulators is not to measure b in order to identify excessive access fees, but rather to ensure that gatekeepers do not design mechanisms that allow them to squeeze third-party sellers in similar ways – e.g., through reserve prices, fixed fees in managed ad campaigns (see Bergemann et al., 2024), or by restricting the number of ad slots.

3.2.2 Access fee setting by third-party apps

When the number of slots is not restricted, as is the case for app stores, the challenge for an elicitation from third-party sellers is that they want the lowest possible access fee ($a = 0$); we saw that for digital goods the absence of access fee encourages the development of me-too apps, whose main purpose is to steal value from existing apps and which create little value for the consumer. To prevent this possibility, the regulator may refrain from monitoring foreclosure.

Let us go back to the app store model with possibly heterogeneous app markets indexed by k, and examine access fee setting by third-party apps subject to the threat

[33] This outcome indeed arises for all $m \in [N, 2N)$, also if $b < 0$ (provided the total value $b + v$ is positive). In contrast, not restricting at all the number of slots ($m \geq 2N$) implies no competition for slots, hence $a(m) = 0$. Then, as in the basic model, the platform has incentives to vertically integrate and foreclose superior sellers if $b > 0$ (socially excessive innovation would instead prevail in the pure-player platform case or if foreclosure is monitored). Similarly, in the case of organic search, where de facto $a = 0$, the search engine has incentive to engage in self-preferencing in markets where it is present as a competitor to third-parties and $b > 0$, as, e.g., in the Google Shopping and Google Flight cases.

of foreclosure. In each market k, after each independent app i proposes a unit access fee a_i^k, app 2 decides whether to innovate. Then, the (pure or hybrid) platform selects $\{\delta_i^k\}_{i=1,2}$ – i.e., it decides to give access to no, one or the two app providers in each market. Finally, the platform and the (non foreclosed) app providers set their prices $\{p_0, p_1^k, p_2^k\}$, and consumers make their consumption choice.

Proposition 7 (information-light implementation). *The welfare-optimal outcome is implemented in all markets by letting all third-party app providers pick their access fee subject to the threat of foreclosure.*

Proof of Proposition 7. We establish this result separately for the two cases of a hybrid and a pure platform:
(i) *Hybrid platform*: When the platform is hybrid and can engage in self-preferencing, app 2 always makes no profit if it does not sell a superior version. As a result, the socially optimal level of innovation obtains if and only if $\pi_2^{k*} = \Delta^k$ with the innovation (fair reward). This outcome can be attained by just eliciting the access fee from the third-party app in each app market k: If foreclosure is not monitored, then choosing $a^k \in \left[b^k, b^k + v^k\right]$ is optimal for the third-party app, as it is foreclosed for $a^k < b^k$ and squeezed for $a^k > b + v^k$.
(ii) *Pure platform*: Let the two third-party apps in market k propose access fees a_1^k and a_2^k. We claim that $a_1^k = a_2^k = v^k + b^k$: Any inferior app knows that it can win consumers if and only if it is the more rewarding app from the platform's standpoint; it cannot afford paying more than $v^k + b^k$, though. A superior app must then bid $v^k + b^k$ as well, as otherwise it would be foreclosed by the platform, which would bring it a higher access fee. The platform lets both apps operate, and app prices are $p_i^k = v_i^k$ for all $i = 1, 2$ and k (hence, the core ZLB binds). Hence, any inferior app makes zero profit, whilst a superior app obtains its fair reward ($\pi_2^{k*} = v^k + \Delta^k + b^k - (v^k + b^k) = \Delta^k$), which yields the socially efficient level of innovation. ■

Although this result is encouraging, it relies on the platform maximizing its profit in each app market. Yet, the fact that platforms are engaged in a variety of B2B relationships across apps and across time gives them the possibility to build a reputation for toughness, or, put differently, to extract higher access fees through predation (by adopting behaviors that do not maximize their short-term profit). This may be the case if the superior app determines the access fee, as in the present model. The platform may downlist this app when the latter offers a socially optimal access fee but refuses to "self-squeeze" (offer above $v^k + b^k$ in our model). Such downlisting would "teach a lesson" to the app developer or, more to the point, its colleagues. How can such predation be prevented? One possibility is to ensure that the access fee is not determined by a superior app, so the latter cannot be pressured to self-squeeze. This is the case in the auction model of Section 3.2.1, provided that the platform does not unduly restrict the number of slots so as to induce *competition* *for the market* among superior providers in distinct app markets. Another approach is to introduce an appeal procedure, similar to the Remark in Section 3.1, to ascertain an "unfair downlisting" (which requires measuring Δ rather than b).

4 Contested bottlenecks

Does platform competition eliminate the scope for access fee regulation? To answer this question, a prior query is "does platform competition promote multi-homing?"; for, it is well known that consumer single-homing on an intermediary provides this intermediary with the monopoly of access to the consumer, regardless of whether it had to compete with other intermediaries to enlist the consumer. Accordingly, the intermediary is a "gatekeeper" or a "bottleneck" pursuant to acquiring the consumer, and can sell access to this consumer at a monopoly price (a^* in the basic model), with the negative consequences that we described earlier. On the other hand, whether a platform captures single-homing consumers could depend on the fee a, high fees potentially inducing high app prices; so we need to look into the mechanics of competition to become the bottleneck.

Even in the presence of competition among several intermediaries, consumer multi-homing may not emerge for at least two reasons: (a) the intermediary is associated with a costly device (few people have both an iPhone and a Samsung), and (b) habit formation and familiarity imply that consumers may multi-home in membership but single-home in usage (most consumers use solely Google search even though also Bing is available on any browser, or systematically consult Booking even though they have as easy an access to Expedia).

In this section, we assume that independent apps multi-home (most popular third-party apps are available on both Apple's and Google's app stores: Bresnahan et al., 2015). By convention and to illustrate the two polar cases, we talk about "competing platforms" when consumers single-home (Section 4.1) and "competing app stores" when consumers multi-home on rival intermediaries (Section 4.2), as envisioned by the DMA (for expositional simplicity, we consider a representative app market as in the basic model). The proofs of the results of this section are relegated to Appendix A.

4.1 Platform competition

Consider $N \geq 2$ (symmetric) competing platforms, indexed by j (Figure 4). In the representative app market, each platform may own an inferior app (valued $v_1 = v$ by consumers); app 2 multi-homes on all platforms, and is valued $v_2 \in \{v, v + \Delta\}$ – i.e., sinking the development cost γ allows it introduce the superior version of its app on all platforms. Let $U^j \equiv u^j - p_0^j$ denote consumers' net value from access to platform j's ecosystem, where $u^j \equiv \max\{v_1 - p_1^j,\ v_2 - p_2^j,\ 0\}$, and $\{p_0^j, p_1^j, p_2^j\}$ are consumers' access price and app $i = 1, 2$ prices on platform j, respectively. To analyse platform competition in the starkest way, we consider perfect competition. That is, we suppose that all consumers patronize only the platform offering the highest net value U^j. As a tie-breaking condition, we assume that platforms offering the same utility split equally the demand, though this does not affect our results.

The timing is the same as with a single platform: (1) The access fees $\{a^j\}$ are set either by the platforms or by regulation; (2) app 2 decides whether to introduce the superior version; (3) the platforms decide whether to vertically integrate and select the apps' realized quality advantage $\{\delta_1^j, \delta_2^j\}$; (4) the platforms and the apps select their prices $\{p_0^j, p_1^j\}$ and p_2^j;[34] (5) consumers choose their platform, and their app on that platform. We can skip the foreclosure decision (the choice of $\{\delta_i^j\}_{i=1,2}$) because under perfect platform competition, a platform has no incentive to degrade its ecosystem by foreclosing a superior app even if $a^j < b$.

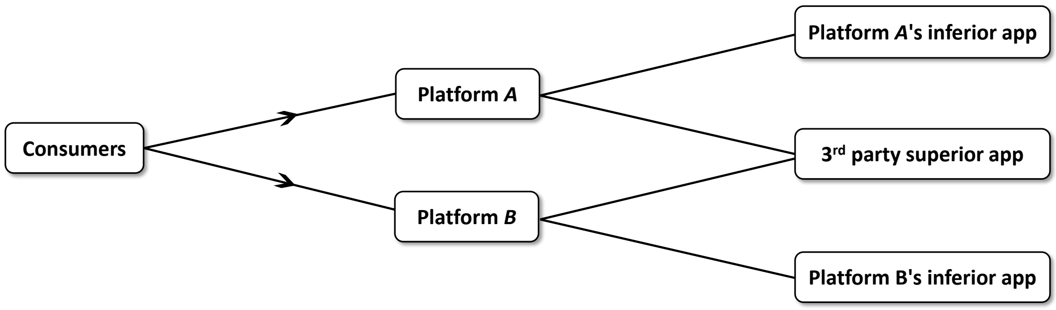


Figure 4: Competing platforms ($j = A, B$) under consumer single-homing.

In equilibrium, all platforms vertically integrate and offer the same net utility $U^* = v$ to consumers, and the core ZLB binds. The presence of platform competition forces platforms to offer the inferior app at zero even if $a^j > b$, which is only possible under vertical integration. Indeed, because in equilibrium consumers are indifferent between the two apps, any candidate equilibrium $p_1^j > 0$ would give room for a platform j to vertically integrate and undercut its rivals. Then, a superior third-party app optimally sets price $p_2^j = \Delta$ and is squeezed if $a^j > b$ for all j. The analysis is similar to that with a monopoly platform in which the core ZLB binds; indeed, it is optimal for each platform to squeeze a superior app. The only difference with the monopoly platform case is a transfer of value v from the platform to the consumers.

Proposition 8 (platform competition). *Consider $N \geq 2$ identical competing platforms, indexed by j.*

[34] It is straightforward to check that, in this simple model, whether a platform's decisions in stages (1) and (3) are observed by rival platforms before the pricing stage is immaterial to the results. Note that, in this context, platform pivotality requires any third-party app to take consumers' demand on each platform as given when setting its price.

(i) Laissez-faire. *In the laissez-faire equilibrium, all platforms vertically integrate, both ZLBs are binding ($p_0^* = p_1^* = 0$), and a superior app sets $p_2^* = \Delta$ on all platforms. All platforms select access fee $a^* = b + G(\Delta + b - a^*)/g(\Delta + b - a^*)$ and make profit a^*/N each. Consumers receive net surplus v each, and a superior third-party app is squeezed.*
(ii) Access fee regulation. *The welfare optimal access fee is $\hat{a} = b < a^*$, yielding per-platform profit b/N and a fair reward to a superior third-party app; consumers still receive net surplus v each.*

The laissez-faire result aligns with the conventional wisdom in platform economics[35] that the multi-homing side does not benefit from platform competition, while the single-homing one (the competitive bottleneck) does, because the platform is the gatekeeper for users on the single-homing side: Platform competition allows consumers to get positive net surplus v. The novel feature of our framework is that perfectly competing platforms collectively earn a positive profit, $(b+G(\Delta+b-a^*)/g(\Delta+b-a^*)$, under laissez-faire. The first component of this profit is the ancillary benefit from app distribution; the second component is part of the value brought about by a superior app developer, which is extracted through the access fee squeeze.[36] Both revenues are not competed away by price competition because of the core ZLB.

The latter conclusion would not hold if platforms are vertically integrated into device manufacturing (or subsidize external device manufacturers under exclusive dealing): competing platforms would then pass through to consumers, via a below-cost price of their devices, the profits earned by squeezing a superior app through the access fee a^*, implying that access fees above the Pigouvian level benefit consumers. However, under the welfare-oriented criterion, the optimal access fee is still $\hat{a} = b$.

4.2 App store competition on a platform

The DMA and the proposed Open App Markets Act require Apple and Google to guarantee third-party app stores' access to their respective devices. These alternative paths from business users to consumers are meant to discipline the currently monopolistic app stores and bring higher quality to consumers and lower fees to business users.[37] As the regulatory texts are silent as to the access conditions, we look at a benchmark in which

[35] See Caillaud and Jullien (2003), Armstrong (2006), Armstrong and Wright (2007) and, more recently, Teh et al. (2023). Armstrong and Wright (2007) explore the implications of a ZLB constraint on the access price charged to the single-homing side, which competing platforms would like to subsidize.

[36] This result hinges on the assumption that platforms can vertically integrate into the app segment: If the inferior app were offered by a (single- or multi-homing) third-party provider, $a^* = b$ would prevail in the laissez-faire equilibrium, which would eliminate the scope for regulation. The reason is that, as non-pivotal third-party apps set their prices as in the basic model, any access fee $a^j > b$ would be passed through to consumers, implying $U^j < v$ and no customer for platform j.

[37] Scott Morton et al. (2024) argue that the Apple's App Store offers poor-quality discovery and curation, and that rival app stores could innovate in the two dimensions and further offer lower fees to app providers.

third-party app stores must be given free access to the platform. Does the availability of competing app stores on a single device eliminate the scope for access fee regulation?

We now have a sequence of "platforms," so we must clarify the terminology. In the following, "platform" will keep designating the gatekeeper to the consumer, whilst "app stores" will be the entities interacting with business users: see Figure 5. Consider a monopoly platform, hereafter denoted by A, vertically integrated into device manufacturing. Let v_d and p_0 denote, respectively, the device stand-alone value and price. On its app store, whose access is priced at p_0^A,[38] consumers can find an inferior, in-house or third-party, app valued v and a potentially superior third-party app valued $v_2 \in \{v, v + \Delta\}$, depending on whether its provider sinks the investment cost γ to introduce a superior version, at prices p_1^A and p_2^A respectively. A's in-house app store faces competition from a third-party app store B priced at p_0^B, where consumers can find the respective inferior (in-house or third-party) app, bringing value v, at price p_1^B, and the same, multi-homing third-party app available on A's store at price p_2^B.

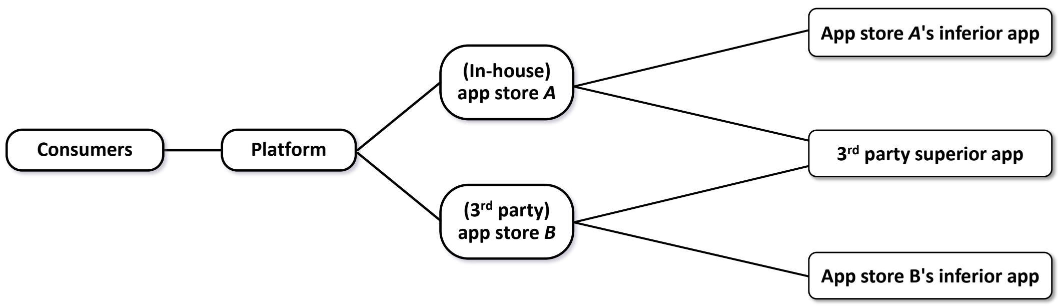


Figure 5: Competing app stores under consumer multi-homing.

Suppose consumers multi-home across app stores that they can access for free (which is always the case in equilibrium).[39] Then, a superior third-party app would serve all consumers on the least expensive platform: App stores de facto engage in Bertrand competition for a superior app, which dissipates their profits – i.e., nil access fees prevail in equilibrium.

Proposition 9 (**app store competition**). *Suppose that the regulator mandates app store competition on devices, with app stores enjoying free access to the device, and that consumers multi-home on app stores on their device.*

[38] When multiple app stores compete for consumers on the same device, its vertically integrated manufacturer is forced to unbundle its two core products (the device and the app store), charging two different prices. In what follows, we refer to the app stores as the core products.

[39] If instead consumers always single-home (because of, e.g., habit formation, or else each downloads at most one app store), then, as all app stores are equally constrained by the core ZLB ($p_0^A \geq 0$ and $p_0^B \geq 0$) the analysis is as in Section 4.1 (with the only difference that the monopoly manufacturer appropriates consumer surplus charging $p_0 = v + v_d$ for the device; this value would instead be appropriated by consumers in the presence also of platform competition). Pigouvian regulation is thus still needed to fairly reward a superior third-party app provider.

- *Because a superior app steers the consumers to the app store with the lowest access fee, Bertrand competition among pure-player or hybrid app stores induces them to charge nothing for consumer access ($p_0^* = 0$) and to levy no fee on third-party apps ($a^* = 0$). A superior app then makes supranormal profit $\Delta + b$.*

- *The Pigouvian access fee ($\hat{a} = b$, where now b is a floor rather than a cap) maximizes welfare by avoiding over-investment in the app market.*

In this simple model, the fair outcome can alternatively be achieved by allowing the platform to levy on third-party app stores a unit access fee α for each app sold through their stores. As Bertrand competition among app stores with opportunity cost α implies that they will in turn charge $a = \alpha$ to the third-party apps, setting by regulation $\hat{\alpha} = b$ maximizes welfare. Thus, whether for apps or for app stores, the proper concept of FRAND access pricing boils down to the Pigouvian principle.

This conclusion would be supported also by a consumer surplus standard in a model where consumers have heterogeneous valuations v_d for the device, because A would react to the reduced profitability of the app store (due to competition) by increasing the device price p_0 (as in Anderson and Bedre-Defolie, 2024b).

5 Relevant literature

(a) Foreclosure. The paper offers new insights on incentives for foreclosure. Because our focus is on gatekeeping, foreclosure of independent apps cannot arise here from the platform's desire to erect "apps barriers to entry" for alternative platforms (Carlton and Waldman, 2002). The aftermarket (Farrell and Klemperer, 2007) and loss-leader (e.g, Lal and Matutes, 1994; Chen and Rey, 2012) incentives to favor internal production over external ones are also absent in our framework. Another rationale for foreclosure stems from regulation: see the literature on the essential facility doctrine (e.g., Hart and Tirole, 1990; Rey and Tirole, 2007), on access pricing for one-sided markets (e.g., Laffont and Tirole, 1994) and on telecom and payment card markets (e.g., Armstrong, 1998; Laffont et al., 1998; Rochet and Tirole, 2002, 2011). In this literature, the access fee to the essential facility is capped, as well as the (equivalent of) the core price; ZLBs play no role. We have contributed to this literature by showing that in digital ecoystems the validity of Chicago School's "rich ecosystem argument"[40] requires an access fee that

[40] The old Chicago School critique of foreclosure theory can be stated for the platform context in the following way: "*Aside from efficiency motives, a hybrid platform (the monopoly segment) has no incentive to foreclose a third-party app (an independent player in the competitive market): A rich ecosystem benefits consumers in two ways, product variety and enhanced competition, and allows the platform to raise its consumer price to extract the associated increase in consumer surplus.*"
This argument is akin to several others recommending a focus on price levels, but non on price structures: in public utility regulation (delegation of individual prices to utilities under the umbrella of a price cap solely aimed at reducing the overall price level), in the antitrust of two-sided markets (the see-

exceeds the ancillary benefit of attracting consumers on the app.

(b) Platform is not a gatekeeper. The literature has studied the regulation of platform fees when the consumer and the merchant can transact through multiple channels: the platform and another channel (direct purchases, other platforms, other payment methods in the case of a payment platform). Because the consumer chooses the channel, the welfare analysis is naturally grounded in the externalities associated with this choice.

Some contributions suppose that the merchant offers the same price regardless of the channel (there is a most-favored-nation, MFN, clause); the merchant's revenue from a sale is then channel-independent, which does not mean that its markup is. The merchant may enjoy a convenience benefit from the platform channel, as in Rochet and Tirole (2011): A card payment may dominate cash and cheque in terms of expediency, fraud prevention, accounting, or absence of hold up. The socially optimal access fee corrects for externalities of consumer channel choice upon merchants, and the socially optimal access fee (which in payment networks is at least partially passed through by issuers to consumers) is equal to the merchant benefit from a card usage; this internalization principle is the so-called *tourist test*. In Gomes and Mantovani (2025), the platform creates an informational and a convenience benefits for consumers; in particular, the platform offers products that they were unaware of. This improved-opportunities benefit of the platform is internalized by consumers. But, consumers' access to the platform being assumed free, they do not directly reward the platform for it, which is a problem if the platform is created only if sufficiently profitable. The platform however can charge consumers indirectly through the competing merchants' access fee, then passed through to consumers. Gomes and Mantovani show that, provided the presence of the platform does not increase aggregate sales, the welfare-maximizing access fee equals the sum of the two benefits it brings. In both papers, $a^* > \hat{a}$.

Alternatively, there may be no MFN (Wang and Wright, 2025). Prices are lower on the platform if it displays tougher merchant competition than the direct sale channel. The consumers may then choose to transact through the platform not because they prefer this channel, but because the latter lowers merchants' markups, at least in part a redistributive effect. The privately optimal fee may now fall short of the socially efficient one, which equals the platform's marginal cost of implementing the transaction plus the amount by which the platform, by intensifying seller competition, decreases the merchants' margins. Again, the merchants' pass-through of the access fee is key to restoring proper consumer incentives.

In contrast with these papers, which hinge on consumers' choice of channel to interact with merchants, we assume that consumers single-home, whether there is platform competition or not: the platform is a "gatekeeper". The set of potential externalities

saw argument and the concomitant recommendation of looking at a single market), and in authorities' agnostic stance with regards to (second- and third-degree) price discrimination.

under consideration is then rather different: (a) a vertically integrated platform may use non-price instruments to prevent consumers from accessing the best product; (b) the platform may jeopardize the existence of superior third-party apps by squeezing them through a high access fee; (c) the third-party apps enjoy supranormal profit when the app ZLB binds. The welfare-maximizing access fee is then equal to the opportunity cost for the platform of letting third-party sellers serve consumers, rather than to the benefits it brings to one or both sides of the market.

(c) Platform presence in app markets. A number of recent papers examine platforms' incentive to vertically integrate, and the welfare effect of this vertical integration, in the presence of foreclosure and/or imitation concerns: see Anderson and Bedre-Defolie (2024a,b), Etro (2021, 2023), Gutièrrez (2021), Hagiu et al. (2022) and Zennyo (2022). Yet, these works, as the ones on platform fees' regulation, assume non-negative opportunity costs (i.e., rule out an app ZLB) and do not consider access pricing on the consumer side.[41] To be certain, one may argue that the widespread assumption that platforms grant free access to consumers in these papers reflects a core ZLB.[42] However, they do not connect the validity of the underlying assumption with the level of seller access fees.

Another closely related contribution to our paper is Choi and Jeon (2021). They show that tying may help a firm circumvent a non-negative price constraint in the tied (complementary) product market that prevents it from squeezing superior sellers in that market. Zero lower bounds do not usually emerge in standard models (e.g., Choi and Stefanadis, 2001, Carlton and Waldman, 2002), which assume that the tied market involves a positive marginal cost. Unlike in this literature on tying, which does not consider access pricing, in our paper margin squeeze of superior third-party sellers by the platform does not necessarily occur via below-cost pricing in the tied (competitive) good market, but primarily via fees: In this case, it is the core ZLB, rather than the ZLB in the tied market (the app ZLB in our terminology), that binds.

6 Conclusion

Gatekeeping platforms control businesses' access to us. Policymakers dealing with platform access have met with the difficulty that welfare analyses in two-sided markets are generally ambiguous. The see-saw effect, and its distant parent, the Chicago school rich ecosystem argument, hold that self-preferencing and high access fees, by degrading the ecosystem and making it unattractive to the consumer side, do not benefit the platform.

[41]By considering access pricing both on consumer and seller side, our work relates to the literature on optimal pricing by two-sided platforms pioneered by Armstrong (2006), Caillaud and Jullien (2003) and Rochet and Tirole (2003, 2006). This literature however is not concerned with hybrid platforms and mostly ignores ZLB constraints.

[42]In other papers on hybrid platforms, including Etro (2023) and Padilla et al. (2022), app stores are bundled with physical devices, so that consumers are always charged a positive price.

Relatedly, capping access fees for business users leads to higher prices on the consumer side. This paper argues that this logic does not apply to the zero-lower-bounds environment of digital markets.

The *core ZLB* (the impossibility for digital platforms to charge negative access prices to consumers) creates incentives for harmful behaviors:

1. The see-saw effect no longer operates. Benefits from a better ecosystem are not passed through to consumers as the platform is reluctant to raise prices. This gives the platform incentives to maximally extract the surplus of business users through high access fees. Extractive access fees create a double marginalization and induce a suboptimal usage of innovative apps. They furthermore discourage the creation of apps. A binding core ZLB might therefore be a smoking gun that high access fees are detrimental.

2. The core ZLB is more likely to bind if there is platform competition, or, in its absence, a high elasticity of consumer demand for the platform. It is less likely to bind if a costly device is part of the bottleneck.

The *app ZLB* (the infeasibility of negative app prices) limits competition in the app markets and generates two inefficiencies:

3. The greater profit made in-house relative to providing access (which arises when $a < b$) creates incentives for self-preferencing, all the more so, the larger the ancillary benefit b (e.g., goods are digital rather than physical) relative to the access fee a. Unfortunately, antitrust watchdogs find it notoriously difficult to discern and demonstrate self-preferencing. A binding app ZLB might therefore be a smoking gun that the platform has incentives to engage in self-preferencing.

4. Low or zero access fees dissipate value by inviting business stealing by me-too apps, that add little value to the ecosystem but extract a non-negligible share of it.

Overall, the argument for capping access fees and more generally enforcing equitable access to gatekeeping platforms is definitely stronger in the presence of ZLBs. In this respect, we concur with the spirit of recent regulatory developments. The latter however remain nebulous when it comes to specific recommendations, and the occasional invocation of the need for “fair, reasonable and non-discriminatory” terms is not helpful. The paper provides guidance for policy-making:

5. The Pigouvian rule ($\hat{a} = b$) discourages self-preferencing and thereby spares intrusive assessment of whether access conditions are actually fair; it also provides app developers with a fair return and therefore a proper incentive to innovate; finally, it minimizes double marginalization conditional on intrusive regulation being infeasible or too costly.

Despite these clear theoretical messages, meeting the empirical challenge of regulating platforms' access policies remains as difficult as it is essential. The task of answering whether a 10% or 30% merchant fee is appropriate is marred with asymmetric information. We made real progress on the question of how to implement the theoretical benchmark; but we feel that more work is necessary to properly tame the gatekeeping platforms while not preventing them from offering innovative services to consumers and businesses alike. As new AI-based platforms are entering the e-commerce, search, and health markets, this question should remain a priority.

References

Amelio, A., & Jullien, B. (2012). Tying and freebies in two-sided markets. *International Journal of Industrial Organization, 30*(5), 436–446.

Anderson, S. P., & Bedre-Defolie, Ö. (2024a). Hybrid platform model: Monopolistic competition and a dominant firm. *The RAND Journal of Economics, 55*(4), 684–718.

Anderson, S. P., & Bedre-Defolie, Ö. (2024b). App platform model. Unpublished manuscript.

Armstrong, M. (1998). Network interconnection in telecommunications. *Economic Journal, 108*(448), 545–564.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics, 37*(3), 668–691.

Armstrong, M., & Wright, J. (2007). Two-sided markets, competitive bottlenecks and exclusive contracts. *Economic Theory, 32*, 353–380.

Bergemann, D., Bonatti, A., & Wu, N. (2024). How do digital advertising auctions impact product prices? *Review of Economic Studies.* https://doi.org/10.1093/restud/rdae087

Bresnahan, T., Orsini, J., & Yin, P. L. (2015). Demand heterogeneity, inframarginal multi-homing, and platform market stability: Mobile apps. In *Proc. 9th IDEI-TSE-IAST Conf. Econ. Intellectual Property, Softw. Internet.*

Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *The RAND Journal of Economics, 34*(2), 309–328.

Carlton, D. W., & Waldman, M. (2002). The strategic use of tying to preserve and create market power in evolving industries. *The RAND Journal of Economics, 33*(2), 194–220.

Chen, Z., & Rey, P. (2012). Loss leading as an exploitative practice. *American Economic Review, 102*(7), 3462–3482.

Choi, J. P., & Jeon, D. S. (2021). A leverage theory of tying in two-sided markets with nonegative price constraints. *American Economic Journal: Microeconomics, 13*(1), 283–337.

Choi, J. P., & Stefanadis, C. (2001). Tying, investment and dynamic leverage. *The RAND Journal of Economics, 32*(1), 52–71.

Etro, F. (2021). Product selection in online marketplaces. *Journal of Economics & Management Strategy, 30*(3), 614–637.

Etro, F. (2023). Hybrid marketplaces with free entry of sellers. *Review of Industrial Organization, 62*(2), 119–148.

Etro, F. (2024). E-commerce platforms and self-preferencing. *Journal of Economic Surveys, 38*(4), 1043–1543.

Farrell, J., & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. In M. Armstrong & R. Porter (Eds.), *Handbook of Industrial Organization* (Vol. 3, pp. 1967–2072). Elsevier.

Farronato, C., Fradkin, A., & MacKay, A. (2023). Self-preferencing at Amazon: Evidence from search rankings. *American Economic Review: Papers and Proceedings, 113*, 239–243.

Gomes, R., & Mantovani, A. (2025). Regulating platform fees under price parity. *Journal of the European Economic Association, 23*(1), 190–235.

Hagiu, A., Teh, T. H., & Wright, J. (2022). Should platforms be allowed to sell on their own marketplaces? *The RAND Journal of Economics, 53*(2), 297–327.

Hart, O., & Tirole, J. (1990). Vertical integration and market foreclosure. *Brookings Papers on Economic Activity: Microeconomics, 1990*, 205–286.

Ichihashi, S., & Smolin, A. (2024). Buyer-optimal algorithmic consumption. Unpublished manuscript.

Janssen, M., Jungbauer, T., Preuss, M., & Williams, C. (2024). Search platforms: Big data and sponsored positions. Unpublished manuscript.

Laffont, J. J., Rey, P., & Tirole, J. (1998). Network competition: I. Overview and nondiscriminatory pricing; II. Price discrimination. *The RAND Journal of Economics, 29*(1), 1–37, 38–56.

Laffont, J. J., & Tirole, J. (1994). Access pricing and competition. *European Economic Review, 38*(9), 1673–1710.

Lal, R., & Matutes, C. (1994). Retail pricing and advertising strategies. *Journal of Business, 67*(3), 345–370.

Lee, K. H., & Musolff, L. (2024). Entry into two-sided markets shaped by platform-guided search. *Econometrica*, R&R.

Padilla, J., Perkins, J., & Piccolo, S. (2022). Self-preferencing in markets with vertically integrated gatekeeper platforms. *The Journal of Industrial Economics, 70*(2), 371–395.

Prat, A., & Valletti, T. (2022). Attention oligopoly. *American Economic Journal: Microeconomics, 14*(3), 530–557.

Rey, P., & Tirole, J. (2007). A primer on foreclosure. In M. Armstrong & R. Porter (Eds.), *Handbook of Industrial Organization* (Vol. 3, pp. 2145–2220). Elsevier.

Rochet, J. C., & Tirole, J. (2002). Cooperation among competitors: Some economics of payment card associations. *The RAND Journal of Economics, 33*(4), 549–570.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association, 1*(4), 990–1029.

Rochet, J. C., & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics, 37*(3), 645–667.

Rochet, J. C., & Tirole, J. (2011). Must-take cards: Merchant discounts and avoided costs. *Journal of the European Economic Association, 9*(3), 462–495.

Scott Morton, F., Crémer, J., Dinielli, D., Heidhues, P., Kimmelman, G., Monti, G., O'Grady, M., Podszun, R., & Schnitzer, M. (2024). Access pricing for app stores under the Digital Markets Act. Yale Tobin Center for Economic Policy.

Spence, A. M. (1975). Monopoly, quality, and regulation. *Bell Journal of Economics, 6*(2), 417–429.

Teh, T. H., Liu, C., Wright, J., & Zhou, J. (2023). Multi-homing and oligopolistic platform competition. *American Economic Journal: Microeconomics, 15*(4), 68–113.

Waldfogel, J. (2024). Amazon self-preferencing in the shadow of the Digital Markets Act. Unpublished manuscript.

Wang, C., & Wright, J. (2025). Regulating platform fees. *Journal of the European Economic Association, 23*(2), 746–783.

Zennyo, Y. (2022). Platform encroachment and own-content bias. *Journal of Industrial Economics, 70*(3), 684–710.

Appendix A: Omitted proofs

Proof of Proposition 3. We distinguish two cases, depending on whether the access fee is lower or larger than the ancillary benefit.

$\boldsymbol{a < b}$. Whether the platform is vertically integrated or not does not matter; intuitively, app 1 is available at price 0 and the platform cannot boost consumer membership by reducing its price. We can therefore focus on the vertically integrated case. As we saw in the text, $b \geq \rho(-v)$ implies that under self-preferencing $p_0 = p_1 = 0$ and the platform's profit is $b[1 - F(-v)]$. Can the platform do better by granting access to a superior third-party app? The unique price equilibrium is $p_1^* = 0$ and $p_2^* = \Delta$. For, if $p_2 > \Delta$, the platform could charge $p_1 = (p_2 - \Delta) - \varepsilon$ and obtain $p_1 + b > a$ per-consumer, without this affecting consumer surplus from apps, hence p_0. Hence, the utility of the consumers from apps is v, and the platform's profit if the third-party app serves the app market is $(p_0 + a)[1 - F(p_0 - v)] < (p_0 + b)[1 - F(p_0 - v)] \leq b[1 - F(-v)]$.

Similarly, if app 2 does not introduce a superior version, symmetric Bertrand competition with negative opportunity costs implies that the unique equilibrium is $p_1^* = p_2^* = 0$, yielding platform profit $[p_0+(a+b)/2][1-F(p_0-v)] < (p_0+b)[1-F(p_0-v)] \leq b[1-F(-v)]$. Therefore, no matter whether app 2 sells a superior version, the platform finds it optimal to engage in vertical integration combined with self-preferencing for all $a < b$.

$\boldsymbol{a \geq b}$. If app 2 does not introduce a superior version, it is never viable in equilibrium. This is because the platform's profit from letting an inferior app serve consumer at the marginal cost, $(p_0 + a)[1 - F(a - b + p_0 - v)]$, is maximized at $p_0 = 0$ and lower than the foreclosure profit, as the two are equal at $a = b$ and $\frac{\partial}{\partial a}\{a[1 - F(a - b - v)]\}\big|_{a=b} = 1 - F(-v) - bf(-v) < 0$ for $b > \rho(-v)$. Therefore, if app 2 does not invest, the platform gains strictly more by vertically integrating and setting $p_0^* = p_1^* = 0$ (price-foreclosure) for all $a > b$.

Consider the subgame where app 2 has invested. In the hybrid platform case, consider the following candidate equilibrium:

$$\{p_0^* = 0, p_1, p_2^* = p_1 + \Delta\},$$

with $p_1 \in [0, v]$. This constitutes an equilibrium – i.e., it satisfies conditions (i) and (ii) of our equilibrium definition – whenever app 2 is viable (i.e., $p_1 + \Delta + b \geq a$) for all p_1 such that

$$[1 - F(p_1 - v)]a \geq [1 - F(-v)]b,$$

which defines an interval $[0, \overline{p}(a)]$, where $\overline{p}(a) \in (0, a - b)$ is increasing in a.

To see this, note that given $p_1 \in [0, v]$, a viable non-pivotal third-party app optimally charges $p_2^* = p_1 + \Delta$. Then,

$$\arg\max_{p_0}(p_0 + a)[1 - F(p_0 + p_1 + v)] = 0 \iff a \geq \rho(p_1 - v),$$

which is implied by $a > b$ and (the inverse hazard rate being decreasing) $p_1 \geq 0$; so, $p_0^* = 0$. Since increasing p_1 has no effect on the equilibrium profits (consumers still buy the third-party app), the only possible deviations to consider are to $\tilde{p}_1 < p_1$. Any such deviation implies that the in-house app is sold. Then, the optimal deviation is to $\tilde{p}_1 = 0$, yielding the foreclosure profit $[1 - F(-v)]b$. It follows that p_1 is an equilibrium price for the inferior app if and only if $[1 - F(p_1 - v)]a \geq [1 - F(-v)]b$. Hence, $p_1 = 0$ is always an equilibrium, but it is not unique: The larger a, the larger the upper bound on p_1, denoted by $\bar{p}(a)$, that can be sustained as an equilibrium. Still, $b > \rho(-v)$ implies that $\bar{p}(a) < a - b$. Moreover, for given a, platform's equilibrium profit decreases in p_1.

Note that here the concept of equilibrium in undominated strategies does not help selecting an equilibrium. Compare an equilibrium price p_1 and consider an alternative price $\hat{p}_1$. A price $\hat{p}_1 < p_1$ (if any) increases platform profit if price p_2 (not necessarily the equilibrium price) is such that app 1 has the market regardless of $\hat{p}_1$ or p_1; it decreases platform profit if price p_2 is such that app 1 is selected under $\hat{p}_1$, but not under p_1, as $\hat{p}_1 + b < a$. Similarly, a price $\hat{p}_1 > p_1$ decreases platform profit if price p_2 (not necessarily the equilibrium price) is such that app 1 has the market regardless of $\hat{p}_1$ or p_1; it increases platform profit if price p_2 is such that app 1 is selected under p_1, but not under $\hat{p}_1$, as $p_1 + b < a$.

Therefore, the hybrid platform gains at least the foreclosure profit for all $a \geq b$, and it can gain strictly more for some $a^* > b$, depending on which equilibrium is played. E.g., in the equilibrium with $p_0^* = p_1^* = 0$ and $p_2^* = \Delta$ for all $a \leq b + \Delta$, as app 2 innovates if and only if $(\Delta + b - a)(1 - F(-v)) - \gamma \geq 0$, the platform's expected profit is $[1 - F(-v)]\{b + (a - b)G[(\Delta + b - a)(1 - F(-v))]\}$, which is maximized at $a^* > b$ such that $a^* = b + G[(\Delta + b - a^*)(1 - F(-v))]/g[(\Delta + b - a^*)(1 - F(-v))]$.

Welfare-optimal access fee. As app 2 does not make any profit when it does not introduce a superior version, it decides to innovate if and only if $a \geq b$ (no self-preferencing) and $(p_1 + \Delta + b - a)[1 - F(p_1 - v)] \geq \gamma$. At $\hat{a} = b$, the unique equilibrium features $\{p_0^* = p_1^* = 0, p_2^* = \Delta\}$. Hence, innovation takes place if and only if $\Delta[1 - F(-v)] \geq \gamma$. As a non-pivotal innovative app never charges $p_2 < \Delta$, it cannot expand consumer participation relative to the foreclosure level, $1 - F(-v)$. For $a > b$, innovation is inefficiently dampened: either $p_1 = 0$ and so socially efficient innovations with $\gamma \in ((\Delta + b - a)[1 - F(-v)], \Delta[1 - F(-v)]]$ are not undertaken, or p_1 is increasing in a and so socially efficient innovations with $\gamma \in (\Delta[1 - F(-v - p_1)], \Delta[1 - F(-v)]]$ are not undertaken.

Ex-post social welfare equals the value under foreclosure, $\int_{-v}^{+\infty}(b + v_c + v)\mathrm{d}F(v_c)$, if the innovation is not undertaken, and $\int_{-u}^{+\infty}(b + v_c + v + \Delta)\mathrm{d}F(v_c) - \gamma$ if the innovation is introduced, where u is consumer net utility. Given that $u \leq (v+\Delta) - p_2 \leq v$, social welfare is maximized at $u = v$, which requires $p_2 = \Delta \iff p_1 = 0$. As noted earlier, this is the unique equilibrium outcome if and only if $\hat{a} = b$.

Similarly, ex-post consumer surplus equals the value under foreclosure, $\int_{-v}^{+\infty}(v_c + v)\mathrm{d}F(v_c)$, if the innovation is not undertaken, and $\int_{-u}^{+\infty}(v_c + v - p_1)\mathrm{d}F(v_c) - \gamma$ if the innovation is introduced, hence it is also maximized at $u = v$ and $p_1 = 0$, which is the unique equilibrium outcome if and only if $\hat{a} = b$. □

Proof of Proposition 4. If no innovation takes place, consumers are homogeneous and so the equilibrium of the basic model obtains, implying that the app 2 provider always makes zero profit, whereas the platform obtains the foreclosure profit $v + b$. In what follows, we consider the subgames following the introduction of a superior version by app 2.

Consider, first, a hybrid platform. For $p_0 + p_1 \leq v$, all consumers buy one app. Since a consumer with type Δ prefers the third-party app if and only if $\Delta \geq p_2 - p_1$, letting $H(\cdot)$ denote the cdf of consumers' type Δ, firms' profits in the representative market are

$$\pi_0^* + \pi_1^* = p_0 + a + H(p_2 - p_1)(p_1 + b - a),$$

and

$$\pi_2^* = [1 - H(p_2 - p_1)](p_2 + b - a).$$

Since its profit is increasing in p_0, the platform optimally sets $p_0 = v - p_1 \in [0, v]$, so that all consumers buy one app, and those buying the in-house app are left with no surplus. By doing so, it achieves a higher profit compared with the one attainable setting prices so that $p_0 + p_1 > v$.[43] For any given a, denoting by $h(\cdot)$ the pdf of consumers' type on $\mathbb{R}^+$ and by $\rho_\Delta(\tilde{\Delta}) \equiv [1 - H(\tilde{\Delta})]/h(\tilde{\Delta})$ the inverse hazard rate, which we assume decreasing, we have:

Lemma (Equilibrium prices and self-preferencing). *Suppose app 2 sells a superior version. There are two thresholds $(\underline{a}, \overline{a})$, with $b < \underline{a} < b + v < \overline{a}$, such that, in the hybrid platform case:*

[43] For $p_0 + p_1 > v$, only the third-party app is bought in equilibrium, and firms' profits are

$$\pi_0 + \pi_1 = \pi_0 = [1 - H(p_0 + p_2 - v)](p_0 + a),$$

and

$$\pi_2 = [1 - H(p_0 + p_2 - v)](p_2 + b - a).$$

For the deviation $p_0 = v - p_1$ (so p_2 remains the same): $p_0 + p_2 - v = p_2 - p_1$, and so

$$\pi_0^* + \pi_1^* = [1 - H(p_0 + p_2 - v)](p_0 + a) + H(p_0 + p_2 - v)(b + v) > \pi_1.$$

Therefore, $p_0^* = v - p_1$ is set so that all consumers access the platform in equilibrium.

- *For $a \leq \underline{a}$, the inferior app ZLB binds ($p_1^* = 0$) and $p_0^* = v$; also the superior app ZLB binds ($p_2^* = 0$) for $a \in [0, b - \rho_\Delta(0)]$, p_2^* is instead strictly positive and increasing in a for $a \in (b - \rho_\Delta(0), \underline{a}]$.*

- *For $a \in (\underline{a}, \overline{a})$: $0 < p_1^* < a - b < p_2^*$, and $p_0^* = v - p_1^* > 0$, with $(p_2^* - p_1^*)$ and firms' profits being constant when a varies.*

- *For $a \geq \overline{a}$, the core ZLB binds ($p_0^* = 0$) and $p_1^* = v < p_2^*$, with p_2^* being strictly increasing in a.*

If left unmonitored, the platform engages in self-preferencing if and only if $a < b$.

Proof of Lemma. The first-order conditions with respect to app prices are as follows

$$\frac{\partial[\pi_0^* + \pi_1^*]}{\partial p_1} = -h(p_2 - p_1)(p_1 + b - a) - 1 + H(p_2 - p_1) = 0 \iff a - b - p_1 = \rho_\Delta(p_2 - p_1), \quad (1)$$

and

$$\frac{\partial \pi_2^*}{\partial p_2} = -h(p_2 - p_1)(p_2 + b - a) + 1 - H(p_2 - p_1) = 0 \iff p_2 - (a - b) = \rho_\Delta(p_2 - p_1). \quad (2)$$

As Δ is distributed on $\mathbb{R}^+$, $p_2^* \geq p_1^*$ in any equilibrium, with strict inequality whenever the superior app ZLB does not bind. First, consider an equilibrium where $p_1^* = 0 \leq p_2^*$. By (1), this is the case if and only if

$$\left.\frac{\partial[\pi_0^* + \pi_1^*]}{\partial p_1}\right|_{p_1=0} \leq 0 \iff a - b \leq \rho_\Delta(p_2) \leq p_2 - (a - b) \iff p_2 \geq 2(a - b), \quad (3)$$

where the second inequality uses (2), which holds with equality as long as $p_2 > 0$. Hence, in equilibrium $p_1^*(a) = p_2^*(a) = 0$ if $\frac{\partial[\pi_0^* + \pi_1^*]}{\partial p_1}|_{p_1=p_2=0} \leq 0$ and $\frac{\partial \pi_2^*}{\partial p_2}|_{p_1=p_2=0} \leq 0$, which gives $a < b - \rho_\Delta(0)$. In turn, from (2),

$$p_2 \geq 2(a - b) \iff a - b \leq \rho_\Delta(2(a - b)), \quad (4)$$

which, as the LHS (resp. RHS) is increasing (resp. decreasing) in a, is satisfied if and only if $a \leq \underline{a}$, with $\underline{a} > b$. The platform's profit is

$$\pi_0^*(a) + \pi_1^*(a) = v + a + H(p_2^*)(b - a).$$

For $a \in [0, b - \rho_\Delta(0)]$, as $p_1^* = p_2^* = 0$ and $H(0) = 1$ (i.e., all consumers buy the third-party app), $\pi_0^*(a) + \pi_1^*(a) = v + a < v + b$. For $a \in (b - \rho_\Delta(0), \underline{a}]$, $p_2^* > 0$, and, by the implicit function theorem,

$$\frac{\partial[\pi_0^* + \pi_1^*]}{\partial a} = h(p_2^*)\frac{\partial p_2^*}{\partial a}(b - a) - H(p_2^*) + 1 > 0 \iff \frac{\partial p_2^*}{\partial a}(a - b) < \rho_\Delta(p_2^*) = p_2^* - (a - b),$$

which is satisfied for all $a < b$, as $\frac{\partial p_2^*}{\partial a} > 0$ (since $p_2^* > 0 > a - b$), and for $a \in [b, \underline{a}]$ as well, by (3), as $\frac{\partial p_2^*}{\partial a} < 1$ (the monotone hazard rate assumption implies $\frac{\partial p_2^*}{\partial a} \in (0, 1)$). Therefore, we can conclude that $\frac{\partial[\pi_0^* + \pi_1^*]}{\partial a} > 0$ for all $a \in [0, \underline{a}]$. Given that $\pi_0^*(b) + \pi_1^*(b) = v + b$, it then follows that non-price foreclosure is optimal for the platform if and only if $a < b$.

Next, consider an equilibrium where $p_2^* > p_1^* \in (0, v)$. In this equilibrium, (1)-(2) imply

$$a - b - p_1 = \rho_\Delta(p_2 - p_1) = p_2 - (a - b) \iff p_1 + p_2 = 2(a - b). \tag{5}$$

As $p_2^* > p_1^*$, it must be $p_1^* < a - b < p_2^*$. Using (5), (1) rewrites as

$$a - p_1 - b = \rho_\Delta(2(a - p_1 - b)). \tag{6}$$

As the LHS (resp. RHS) is decreasing (resp. increasing) in p_1, this equilibrium exists if and only if

$$p_1^* > 0 \iff a - b > \rho_\Delta(2(a - b)) \iff a > \underline{a},$$

and, using (2),

$$p_1^* < v \iff a - b - v < \rho_\Delta(2(a - b - v)) \iff a < \overline{a},$$

where, comparing the two above inequalities, it follows that $\overline{a} > \underline{a}$. From (6) it follows that $p_1^* - a$ is constant varying a. Since $p_1 + p_2 = 2(a - b)$ is equivalent to $p_1 - a = a - p_2 - 2b$, this implies that $a - p_2^*$ is constant in a as well, and so also $p_2^* - p_1^*$ does not vary with a. This shows a neutrality result: $\pi_0^*(a) + \pi_1^*(a) = [1 - H(p_2^* - p_1^*)](v - p_1^* - a) + H(p_2^* - p_1^*)(b + v)$ is independent of a in this range. However, $\pi_0^*(a) + \pi_1^*(a) > v + b$ since $p_1^* < a - b$.

Finally, we consider an equilibrium where $p_1^* = v < p_2^*$ (and so $p_0^* = 0$). By (1) and (2), this is the case if and only if

$$\left.\frac{\partial[\pi_0 + \pi_1]}{\partial p_1}\right|_{p_1 = v} \geq 0 \iff a - b - v \geq \rho_\Delta(p_2 - v) = p_2 - (a - b) \iff a - b - v \geq \rho_\Delta(2(a - b - v)),$$

which holds if and only if $a \geq \overline{a}$, with $\overline{a} > b + v$ implying $p_1^* = v < a - b$. The platform's profit is

$$\pi_0^*(a) + \pi_1^*(a) = H(p_2^* - v)(b + v - a) + a.$$

We then have:

$$\frac{\partial[\pi_0^* + \pi_1^*]}{\partial a} = h(p_2^* - v)\frac{\partial p_2^*}{\partial a}(b + v - a) - H(p_2^* - v) + 1,$$

where, by the monotone hazard rate assumption, $\frac{\partial p_2^*}{\partial a} \in (0, 1)$ is characterized using the implicit function theorem. □

Platform business model and innovation. If app 1 was independently owned, it would make positive expected profit $\pi_1^* = H(p_2 - p_1)(p_1 + b - a)$ by setting a price $p_1 > a - b$ if $b < a < b + v$. Vertical integration allows the platform to capture app 1's profit – i.e., a hybrid platform earns π_1^* on top of the pure platform profit π_0^*. By revealed preference, the result that p_1 is below the opportunity cost under vertical integration implies that the hybrid platform earns strictly more than the sum $\pi_0^* + \pi_1^*$ in the equilibrium without vertical integration. Hence, the platform has strict incentives to vertically integrate both when integration takes place by introducing another inferior app – as, this app featuring a lower price than the independent inferior app, it takes all its consumers – or by making a takeover offer to the independent provider of app 1. Finally, an independent app 1 provider would not be viable if $a > b + v$, implying that a fortiori the platform vertically integrates in this case. As a result, when app 2 sells a superior version, the platform finds it optimal to vertically integrate for all values of a.

Hence, app 2 earns nothing if it does not innovate, whereas by innovating it gets nothing for $a < b$ (self-preferencing region) and the profit π_2^* characterized above, minus the investment cost γ, for $a \geq b$. As a result, app 2 innovates with probability $G[(1 - H(p_2^* - p_1^*))(p_2^* + b - a)]$ if and only if $a \geq b$.

Platform-optimal vs welfare-optimal access fees. Therefore, platform's expected profit equals $v + b$ for all $a < b$, and simplifies as

$$v + b + G[(1 - H(p_2^* - p_1^*))(p_2^* + b - a)][1 - H(p_2^* - p_1^*)](a - b - p_1^*),$$

for $a \geq b$. Since, at $a = b$, $p_1^* \equiv 0 < p_2^*$, the derivative of this expected profit at $a = b$ equals $G[(1 - H(p_2^*))p_2^*][1 - H(p_2^*)] > 0$, which implies that $a^* > b$.[44]

Similarly, expected social welfare equals $W(a) = v + b$ for $a < b$, and

$$W(a) = v + b + G[(1 - H(p_2^* - p_1^*))(p_2^* + b - a)] \int_{\Delta \geq p_2^* - p_1^*} \Delta \, \mathrm{d}H(\Delta),$$

[44]The result $a^* > b$ also holds if the superior app pre-exists the access fee setting: indeed, the platform's profit conditional on the innovation being introduced is maximized at $a^* > \overline{a}$. To see this, note that for $a \geq \overline{a}$, we have

$$\frac{\partial[\pi_0^* + \pi_1^*]}{\partial a} > 0 \iff a - b - v < 2(p_2^* + b - a) + \frac{h'(p_2^* - v)}{h(p_2^* - v)}(p_2^* + b - a)^2.$$

At $a = \overline{a}$, $p_2^* = 2(a - b) - v$. Substituting into the above inequality and simplifying gives

$$\frac{h'(2(a - b - v))}{h(2(a - b - v))}(a - b - v) = \frac{h'(2(a - b - v))}{h(2(a - b - v))}\rho_\Delta(2(a - b - v)) > -1,$$

where the equality follows from the definition of $\overline{a}$. This inequality is always satisfied as it is equivalent to the assumption of decreasing inverse hazard rate. Therefore, the platform's equilibrium profit is still increasing at $a = \overline{a}$, and so $a^* > \overline{a} > b$.

which is strictly larger than $v+b$, for all $a \geq b$. We have that

$$\frac{\partial W(a)}{\partial a} = - g(\cdot)\left[h(\cdot)(p_2^* + b - a)\frac{\partial[p_2^* - p_1^*]}{\partial a} + [1 - H(\cdot)]\left(1 - \frac{\partial p_2^*}{\partial a}\right)\right]\int_{\Delta \geq p_2^* - p_1^*} \Delta \, \mathrm{d}H(\Delta) + \\ - G(\cdot)h(\cdot)(p_2^* - p_1^*)\frac{\partial[p_2^* - p_1^*]}{\partial a} \leq 0 \quad \forall a \geq b,$$

with strict inequality at $a = b$, implying that, when non-price foreclosure cannot be monitored, $\hat{a} = b$ maximizes expected social welfare.

Finally, as $p_0^* + p_1^* = v$, consumers purchasing the platform's in-house app have zero surplus, and expected consumer surplus equals $S(a) = 0$ with foreclosure ($a < b$), and

$$S(a) = G[(1 - H(p_2^* - p_1^*))(p_2^* + b - a)]\int_{\Delta \geq p_2^* - p_1^*} [\Delta - (p_2^* - p_1^*)] \, \mathrm{d}H(\Delta) > 0,$$

if there is no foreclosure (i.e., for all $a \geq b$). We have that

$$\frac{\partial S(a)}{\partial a} = - g(\cdot)\left[h(\cdot)(p_2^* + b - a)\frac{\partial[p_2^* - p_1^*]}{\partial a} + [1 - H(\cdot)]\left(1 - \frac{\partial p_2^*}{\partial a}\right)\right]\int_{\Delta \geq p_2^* - p_1^*} [\Delta - (p_2^* - p_1^*)] \, \mathrm{d}H(\Delta) \\ - G(\cdot)[1 - H(\cdot)]\frac{\partial[p_2^* - p_1^*]}{\partial a} \leq 0 \quad \forall a \geq b,$$

with strict inequality at $a = b$, implying that, when non-price foreclosure cannot be monitored, $\hat{a} = b$ also maximizes expected consumer surplus.[45] □

Proof of Proposition 6. Consider a subgame following innovation by $n \in \{0, ..., N\}$ sellers. For any number m of available slots, the undominated strategy equilibrium of the uniform price auction is as follows.

Lemma (equilibrium bid). *Denoting by $n \leq N$ the number of app markets where seller 2 innovates, the equilibrium bid paid by the m winners of the ad-auction is:*

$$a(m) = \begin{cases} v + b + \Delta & \text{if } m < n \\ v + b & \text{if } n \leq m < 2N \\ 0 & \text{if } m \geq 2N \end{cases}$$

Proof of Lemma. We consider three distinct cases:

[45]The same results also hold if the superior app pre-exists the access fee policy: in this case, consumer surplus and social welfare are as above with $G(\cdot) \equiv 1$; hence, they are both lower under foreclosure (implying that, if non-price foreclosure cannot be monitored, it must be that $a \geq b$) and decreasing in the relative price $p_2^* - p_1^*$, which is a weakly increasing function of a (strictly so for $a \in [b, \underline{a}]$). The result that expected social welfare and consumer surplus are strictly decreasing at $a = b$ implies that, if foreclosure can be monitored (irrespective of whether the superior app pre-exists the access fee policy) the regulator would set $\hat{a} < b$.

(i) If $m < n$, there are fewer slots than superior sellers; the Bertrand logic applies. Superior apps' competition implies that all superior apps bid $v + b + \Delta$, which then is the equilibrium bid: inferior sellers cannot afford to pay $a(m) = v + b + \Delta$ per-sale, and any displayed superior seller can set the monopoly price $p_2^* = v + \Delta$ and receives an extra b, and so makes zero profit. A superior seller cannot increase its profit neither by bidding $a < a(m)$ (it would lose the auction for sure), nor by bidding $a > a(m)$ (as serving consumers would come at a loss).
(ii) If $n \leq m < 2N$, all superior sellers win with probability 1 by slightly overbidding the inferior sellers' bid, $v+b$, i.e., the per-consumer revenue any inferior seller makes if selected (this per-consumer revenue is captured only if the superior seller in the same market is not selected, which never happens on-path; however, "pay-per-click" implies that only bidding $v + b$ is an undominated strategy for an inferior seller). Therefore, $a(m) = v + b$. Inferior sellers, if selected, charge $p_1^* = v$ and so make zero profits regardless of whether a superior seller in their market exists. Any superior seller's price is $p_2^* = v + \Delta$ whether or not it faces a competitor. Hence, the pricing behavior is unchanged no matter whether the pool of selected sellers is common knowledge at the pricing stage. All superior sellers are displayed, set $p_2^* = v + \Delta$ and make a profit Δ per-sale.
(iii) If $m \geq 2N$, all products are always shown, and so all sellers bid zero: hence, $a(m) = 0$. Equilibrium prices are then $p_1^* = 0$ and $p_2^* \in \{0, \Delta\}$ depending on whether product 2 is superior, so that a superior seller makes profit $\Delta + b$ per-sale. □

Anticipating this, platform's behavior is as follows:

Lemma (self-preferencing). *The platform engages in (vertical integration combined with) self-preferencing if and only if $m \geq 2N$.*

Proof of Lemma. Consumer access price p_0^* equals consumer net surplus from the displayed slots.

- If $m \geq 2N$, all products are shown and consumers receive net surplus v from the preferred market, hence $p_0^* = v$, which, as $a(m) = 0$, coincides with the platform's profit. By serving consumers directly – i.e., offering an inferior product in each market and self-preferencing it – the platform can instead appropriate $v + b$, and so (vertical integration combined with) self-preferencing is privately optimal.

- For all $m < 2N$, consumers get no net surplus from the displayed products (as $p_i^* = v_i$ for all sellers), and so $p_0^* = 0$, implying that the platform's profit coincides with the revenues from fees. As the platform obtains at least $v + b$ through fees from any third-party sale, it has no incentive to steer consumers to its in-house products (hence, no incentives to vertically integrate either). □

We next characterize the symmetric equilibrium of the innovation stage as a function of the sellers' development cost $\gamma > 0$.

Lemma (innovation). *For a given number m of available slots, the innovation stage admits a unique symmetric equilibrium. If $\gamma \leq \Delta/N$, each seller 2 innovates with probability $\rho^*(m) \in [0,1]$, with:*

(i) $\rho^*(0) = 0$,

(ii) $\frac{d\rho(m)}{dm} > 0$ *for all* $m \in [0, N]$,

(iii) $\rho^*(m) = 1$ *for all* $m \in [N, 2N)$,

(iv) $\rho^*(m) = 0$ *for all* $m \geq 2N$.

If instead $\gamma > \Delta/N$, no seller innovates.

Proof of Lemma. An inferior seller always makes zero profit; a superior one obtains profit Δ/N (as each consumer selects it with probability $1/N$) if and only if $m < 2N$ (otherwise, the platform engages in self-preferencing) and the number n of superior sellers in equilibrium satisfies $n \leq m$, and zero profit otherwise. Therefore, there is never innovation if $\gamma > \Delta/N$. For $\gamma \leq \Delta/N$, no seller innovates in the trivial case $m = 0$ or if $m \geq 2N$, whereas all sellers 2 innovate if $m \geq N$; if $m \in (0, N)$, seller 2 in market i innovates if and only if

$$\Pr[n_{-i} \leq m-1] \cdot \Delta/N - \gamma \geq 0,$$

where $n_{-i} \in \{0, \ldots, N-1\}$ is the number of its innovative rivals. A symmetric mixed strategy equilibrium in which each seller 2 innovates with probability $\rho \in (0,1)$ is obtained by solving the indifference condition:

$$\sum_{j=0}^{m-1} \binom{n_{-i}}{j} \rho^j (1-\rho)^{n_{-i}-j} = \frac{N\gamma}{\Delta}.$$

Uniqueness and monotonicity in m of the solution follow from the properties of the binomial distribution. □

We are now ready to prove the statement of Proposition 6.

Welfare-optimal number of slots. Setting $\hat{m} = N$ (indeed, any $m \in [N, 2N)$) implies that all sellers 2 innovate if and only if $\Delta/N \geq \gamma$ and no seller innovates otherwise, which is socially efficient. Note that, as each consumer derives value from each market $k = 1, ..., N$ with probability $1/N$, the socially efficient number of slots guarantees each innovative seller its fair reward Δ/N. Conversely, when innovation is socially optimal, it is inefficiently dampened for all $m < N$ (as each seller innovates only with probability

$\rho^*(m) < 1$) or $m \geq 2N$ (because the anticipation of foreclosure gives no incentives to innovate).[46]

Platform-optimal number of slots. By contrast, the platform may make a larger profit by setting $m^* < N$.[47] As seen above, the platform makes profit $v + b$ by setting any $m \in [N, 2N)$ (and also for $m \geq 2N$ provided it can vertically integrate and engage in self-preferencing). If $\gamma \leq \Delta/N$, choosing $m < N$ implies that the number of superior sellers is binomially distributed with success probability $\rho^*(m)$, and the platform's expected profit is

$$m/N \left\{v + b + \Pr[n > m] \cdot \Delta\right\}.$$

Whether this is larger than $v + b$ depends on the parameters.

To see this, consider the simplest example with $N = 2$ markets. For $m = 1$, we have $\rho^* = 1 - 2\gamma/\Delta$ and so the platform's profit equals

$$\frac{1}{2}\left[v + b + \Delta(\rho^*)^2\right] = \frac{1}{2}\left[v + b + \frac{(\Delta - 2\gamma)^2}{\Delta}\right],$$

which is larger than $b + v$ if and only if

$$\Delta > 2\gamma + \frac{b+v}{2} + \sqrt{\left(\frac{b+v}{2}\right)^2 + (b+v)2\gamma},$$

which concludes the proof. □

Proof of Proposition 8. Suppose app 2 sells a superior version. As $u^j \equiv \max\{v - p_1^j,\ v+\Delta-p_2^j,\ 0\}$, and the superior app always charges $p_2^{j*} \geq \Delta$, we have $u^j \in [0, v]$ for all j. As $p_0 \geq 0$ and consumers' outside option is zero, also $U^j \in [0, v]$. Next, take two platforms j' and j'' and suppose they offer different utility levels to consumers, $v \geq U^{j'} > U^{j''} \geq 0$, with $U^{j'} = \max_j\{U^j\}$ so that platform j' has a strictly positive market share. Then, platform j'' would face no demand and make zero profit. By vertically integrating and foreclosing the third-party app, and setting prices $p_0^{j''} + p_1^{j''} \leq v - U^{j'}$, it would offer utility $U^{j''} \geq U^{j'}$ and make a positive profit.

As a result, all platforms must offer the same utility U^* in equilibrium. Hence, their profit is $\frac{1}{N}(p_0^j + p_1^j + b)$ with foreclosure, with $p_0^j + p_1^j = v - U^*$, and $\frac{1}{N}(p_0^j + a^j)$ without

[46] In the latter case, if self-preferencing is monitored or vertical integration is not an option for the platform, then excessive innovation would prevail, i.e., all sellers 2 would innovate even for some $\gamma > \Delta/N$, and so $\hat{m} = N$ would still be optimal.

[47] This possibility result hinges on the restriction to symmetric equilibria in the innovation stage. Indeed, for all $m \in (0, N)$, if $\gamma \leq \Delta/N$ the innovation stage also admits asymmetric pure-strategy equilibria in which a number $n = m$ of sellers innovate. This is because if seller 2 in market k expects m (resp. $m - 1$) rivals to innovate, it has no (it has) incentives to do so, given that an innovative seller gets profit Δ/N if and only if $n \leq m$ and zero otherwise. Selecting any such equilibrium implies that the platform's profit is always maximized at $m^* = \hat{m} = N$, as restricting the number of slots reduces the amount of sales without yielding a positive probability of margin squeeze.

foreclosure, with $p_0^j = v + \Delta - p_2^{j*} - U^*$. If $p_0^j > 0$ for some j, then, no matter whether it forecloses or not the third-party app, platform j would find it optimally to deviate, charging a slightly lower access price to consumers to serve all demand. Therefore, in equilibrium $p_0^{j*} = 0$ for all j.

Whenever its rivals are expected to provide $U^* = v$ in equilibrium, any platform j has no profitable deviation to $U^j \neq U^*$: offering $U^j < v$ drives its profit to zero, and, as shown above, it is never possible to provide $U^j > v$. As $U^* = v$ can always be provided by vertically integrating, foreclosing the third-party app, and setting $p_0^j = p_1^j = 0$, it follows that an equilibrium where $U^* = v$ always exists. We next characterize the corresponding subgame perfect equilibrium prices for any given $(a^j, \delta_2^j = 0)_{j=1,\ldots,N}$ (no foreclosure). Suppose that in equilibrium $p_1^j > 0$ – this is necessarily the case if platform j is a pure player and $a^j > b$ (with two independent apps, equilibrium prices are as in the basic model). Then, as $p_2^{j*} = \min\{p_1^j + \Delta, v + \Delta\}$ whenever app 2 is viable, $u^j = \max\{v - p_1^j, 0\} < v$. Given that rival platforms offer higher value $U^* = v$, the considered platform makes no profit. It has therefore a strictly profitable deviation: It can vertically integrate, set $p_1^j = 0$ and thus, by selling its in-house app, offer value $U^j = v$ to consumers, so as to attract some of them and make positive profits. Hence, the app ZLB binds: $p_1^{j*} = 0$ for all j and $(a^j, \delta^j = 0)_{j=1,\ldots,N}$. Anticipating this, the (non-foreclosed) third-party seller must set $p_2^{j*} = \Delta$ to sell its app. It optimally does so whenever selling its app yields positive profit on platform j (i.e., as long as $\Delta + b - a^j \geq 0$). This implies that each platform has an incentive to vertically integrate for all $a^j > b$. For $a^j \leq b$, app prices are $p_1^{j*} = 0$ and $p_2^{j*} = \Delta$ no matter whether the platform is vertically integrated; still, platforms do not foreclose as otherwise they could be undercut by a deviating platform offering access to the superior app.

By the same arguments, if app 2 does not introduce a superior version, the platform always vertically integrates and sets $p_1^{j*} = 0$, and it engages in non-price foreclosure if $a^j < b$ and in price foreclosure if $a^j > b$. Therefore, app 2 does not make any profit if it does not innovate (in this case, each platform makes the foreclosure profit b/N), whereas it makes profit $\Delta + b - a^j$ from each platform j's consumer if it innovates (recall that for all $a^j \leq b + \Delta$ app prices are $p_1^{j*} = 0$ and $p_2^{j*} = \Delta$). As each platform is selected by a share $1/N$ of consumers, app 2 introduces the innovation if and only if $\Delta + b - \frac{1}{N}\sum_{j=1}^N a^j \geq \gamma$. Therefore, platform j's expected profit is

$$\pi^j = \frac{1}{N}\left[b + (a^j - b)G\left(\Delta + b - \frac{\sum_{k=1}^N a^k}{N}\right)\right].$$

Maximizing it with respect to a^j immediately yields the unique symmetric equilibrium

$$a^* = b + \frac{G(\Delta + b - a^*)}{g(\Delta + b - a^*)},$$

with $a^* \in (b, b+\Delta)$. Hence, $\pi_0^* = \frac{1}{N}\left[b + G(\Delta + b - a^*)/g(\Delta + b - a^*)\right]$ exceeds the foreclosure profit $\frac{b}{N}$. By contrast, socially efficient investment incentives require that $\hat{a} = b$. $\square$

Proof of Proposition 9. As in Proposition 8, no matter whether app 2 sells a superior version, app store competition implies that each app store provides net utility $U^* = v$ to consumers in equilibrium. Hence, it must be $p_0^{j*} = p_1^{j*} = 0$ for all a^j and app stores j (implying that platform j would vertically integrate for all $a^j > b$), including the in-house app store. As all app stores can be accessed for free, consumers multi-home.

Moreover, as in Proposition 8, if app 2 does not introduce a superior version, app store j engages in (vertical integration combined with) non-price foreclosure if $a^j < b$ and in price foreclosure if $a^j > b$. Without innovation, for all access fees the equilibrium thus features the foreclosure profits, i.e. b/N for each app store and 0 for app 2.

If app 2 introduces a superior version, an equilibrium where it is foreclosed by all app stores cannot exist under laissez-faire: Given that consumers would prefer to buy the superior app at any price $p_2^j \leq \Delta$, any app store would deviate by granting access to the superior app at an access fee $a^j > b/N$, promoting app 2 innovation and increasing the profit conditional on innovation taking place. By contrast, starting from a candidate equilibrium where the superior app is given access by its rivals, a deviating app store would lose all its consumers if foreclosing.

Because the superior app can sell to all consumers on any app store j at any price $p_2^j \leq \Delta$, it will optimally sell at a price (slightly below) Δ on the app store charging the lowest access fee: this app store attracts all sales, and the superior app makes a profit $\pi_2 = \Delta + b - \min_j a^j$. Because of this, the unique (symmetric) laissez-faire equilibrium features $a^{j*} = 0$ for all j.[48] Indeed, in any candidate equilibrium with positive access fees, all app stores must set the same access fee a^* to attract some sales; but then each app store would profitably slightly undercut a^*: this deviation has a negligible impact on the probability $G(\Delta + b - \min_j a^j)$ with which app 2 innovates, but allows the deviating app store to capture a positive access fee on all (rather than a fraction $1/N$ of) consumers.

Then, under laissez-faire a superior app makes profit $\pi_2^* = \Delta + b$, while app stores make zero profits. As a result, app 2 innovates for all $\gamma \leq \Delta + b$, implying that socially inefficient innovations with development cost $\gamma \in (\Delta, \Delta + b]$ will be introduced. By contrast, social welfare is maximized whenever $\pi_2^* = \Delta$, which is the case if and only if $a^j = b$ for all j.

Finally, both under laissez-faire and under the Pigouvian rule, consumers get net value v from the app stores and v_d from the device, and so the monopoly platform optimally

[48] More precisely, when at least two app stores set $a^{j*} = 0$, the others could set any positive access fee. Yet, in any such equilibrium the superior app, selling only through the most convenient app stores, pays no access fee.

sets $p_0 = v + v_d$. □

Appendix B: Extensions and robustness

B.1 Heterogeneous app markets

This section generalizes the basic model to multiple, heterogeneous app markets, indexed by k. The following analysis applies to an arbitrary finite number or to a continuum (mass 1) of markets. Let the ancillary benefit, the inferior app value and the competitive advantage of the superior app in market k be denoted b^k, v^k, and Δ^k; $G^k(\cdot)$ and $g^k(\cdot)$ denote the cdf and pdf of the development cost γ_k of app 2 in market k. Third-party providers are different across app markets. We allow for more generality the access fee to be app-market contingent (a^k).

Proposition (heterogeneous app markets). *Suppose that multiple heterogeneous app markets, indexed by k and each described as the representative app market of the basic model, coexist on the platform.*

- *The platform engages in vertically integration combined with self preferencing in any market k where $a^k < b^k$.*
- *Under laissez-faire, the platform's profit is maximized by setting a market-specific access fee a^{k*}, defined by*

$$a^{k*} = b^k + v^k + \frac{G^k(b^k + v^k + \Delta^k - a^{k*})}{g^k(b^k + v^k + \Delta^k - a^{k*})};$$

  this value strictly exceeds the welfare-optimal values of the access fee in market k, given by $a^k \in [b^k, b^k + v^k]$.

Proof of Proposition. To start with, we show that the pricing stage admits a unique equilibrium, with app prices as in Lemma 1 in each market k. Let us consider general valuations $v_i^k - \delta_i^k$ for $i = 1, 2$ in each app market k. Slightly abusing notation, in what follows we denote $\Delta^k = 0$ if app 2 in market k does not sell a superior version.

Assume, without loss of generality, that $v_2^k - \delta_2^k \geq v_1^k - \delta_1^k$, i.e. $\Delta^k - \delta_2^k + \delta_1^k \geq 0$. Let us first show that we can assume, also without loss of generality, that $p_1^k \leq \max(a^k - b^k, 0)$. Either $v_1^k - \delta_1^k < p_1^k$ and app 1 is out of app market k; then charging $\max(a^k - b^k, 0)$ cannot do worse than charging p_1^k, regardless of whether app 1 is owned by the platform or an independent app provider (and in the case of vertical integration, may increase consumers' surplus, which benefits the platform, which can raise p_0). Or $v_1^k - \delta_1^k \geq p_1^k$; because app 2 wins the market, if app 2 is owned by an independent app provider, platform pivotality (condition (ii) of our equilibrium definition) implies that app 2 takes as given consumers' participation on the platform and so optimally sets p_2^k such that consumers are indifferent between the two apps: $p_2^k = p_1^k + \Delta^k - \delta_2^k + \delta_1^k$. If $p_1^k > a^k - b^k > 0$, app 1 can charge $p_2^k - (\Delta^k - \delta_2^k + \delta_1^k) - \varepsilon < p_1^k$, win the market and make a strictly higher profit; furthermore, if app 1 is owned by the platform, the consumers' utility from app market k slightly increases.

Next suppose that $p_1^k < a^k - b^k$ (which requires $a^k > b^k$). Let us show that price p_1^k is strictly dominated for an independent owner of app 1 by price $a^k - b^k$. If $\tilde{p}_2^k \equiv p_2^k - (\Delta^k - \delta_2^k + \delta_1^k)$ lies below p_1^k, whether app 1 is priced at p_1^k or $a^k - b^k$ makes no difference; but if it lies above p_1^k, charging p_1^k rather than $a^k - b^k$ implies a loss of $a^k - b^k - p_1^k > 0$ if app 1 is owned by an independent provider; in this case, the only undominated behavior involves $p_1^k = a^k - b^k$.

Finally, suppose that the platform owns app 1. The same arguments imply that $p_1^k = \max\{a^k - b^k, 0\}$ is the unique price in an equilibrium in undominated strategies if the non-pivotal app 2 is viable – i.e., $a^k \leq b^k + v^k + \Delta^k - \delta_2^k$. Otherwise, app 2 charges any $p_2^k > v^k + \Delta^k - \delta_2^k$ and the unique undominated price for the in-house app is $p_1^k = v^k$: reducing p_1^k and increasing p_0 to the same extent would in fact reduces platform profit if app 2 were to set $\tilde{p}_2^k \leq v^k + \Delta^k - \delta_2^k$.

Consumers' utility from accessing the app store is

$$U \equiv \sum_k u^k - p_0,$$

with u^k being the utility obtained from app market k:

$$u^k \equiv \max\{v_1^k - p_1^k,\ v_2^k - p_2^k,\ 0\}.$$

As seen above, irrespective of the ownership of the inferior app in each market k, equilibrium app prices $\{p_1^{k*}, p_2^{k*}\}$ are as in the basic model. Then, p_0 is set so as to satisfy consumers' participation constraint with equality ($U = 0$):

$$p_0^* = \sum_k u^{k*}.$$

Hence, denoting $x_i^k = 1$ if app i is sold in market k, platform's profit writes

$$p_0^* + \sum_{\{k:\, x_1^k=1\}} (p_1^{k*} + b^k) + \sum_{\{k:\, x_2^k=1\}} a^k = \sum_{\{k:\, x_1^k=1\}} \pi^k(x_1^k = 1) + \sum_{\{k:\, x_2^k=1\}} \pi^k(x_2^k = 1),$$

where

$$\pi^k(x_1^k = 1) \equiv v_1^k + b^k, \qquad \pi^k(x_2^k = 1) \equiv v_2^k - p_2^{k*} + a^k,$$

are the per-market profits with and without foreclosure, respectively (inclusive of the revenues from optimally setting consumers' access price).

Consider, first, markets where app 2 sells a superior version. If not foreclosed, the superior third-party app in market k makes

$$\pi_2^{k*} = p_2^{k*} + b^k - a^k.$$

In any market k where $a^k < b^k$, absent foreclosure, consumers purchase the superior app at $p_2^{k*} = \Delta^k$ and obtain utility $u^{k*} = v^k$. As this is the same utility that they would obtain under foreclosure and $p_1^{k*} = 0$, it follows that by foreclosing superior rivals in any such market the platform can charge the same access price p_0^* to consumers, but obtains higher unit revenues $b^k > a^k$. Therefore, vertical integration combined with self-preferencing occur for all $a^k < b^k$. In any market k where $a^k \in [b^k, b^k + v^k]$, absent foreclosure, consumers purchase the superior app at $p_2^{k*} = a^k - b^k + \Delta^k$ and obtain utility $u^{k*} = v^k - (a^k - b^k) > 0$. From any such market, the platform obtains profit $\pi^k(x^k = 1) = v^k + b^k = \pi^k(x^k = 0)$, and so is indifferent between foreclosing or not. The third-party seller gains $\pi_2^{k*} = \Delta^k$. In any market k where $a^k \in (b^k + v^k, b^k + v^k + \Delta^k]$, absent foreclosure, consumers purchase the superior app at $p_2^{k*} = v^k + \Delta^k < a^k - b^k + \Delta^k$ and obtain utility $u^{k*} = 0$. From any such market, the platform obtains profit $\pi^k(x^k = 1) = a^k > v^k + b^k = \pi^k(x^k = 0)$, and so is strictly better off than under foreclosure. The superior app is squeezed: $\pi_2^{k*} = v^k + \Delta^k + b^k - a^k < \Delta^k$. The platform vertically integrates, and the price foreclosure outcome arises, in markets where the superior app is not viable, i.e. where $a^k > b^k + v^k + \Delta^k$.

In markets where app 2 does not sell a superior version, $p_1^{k*} = p_2^{k*} = \max\{a^k - b^k, 0\}$ for all $a^k \leq b^k + v^k$. As above, the platform gains from vertical integration combined with self-preferencing in markets where $a^k < b^k$, whereas it is indifferent for $a^k \in [b^k, b^k + v^k]$. For $a^k > b^k + v^k$, both third-party apps are not viable, hence the platform makes $\pi^k(x_1^k = 1)$ by vertically integrating (price foreclosure). Therefore, for all a^k, app 2 in market k does not make any profit without introducing a superior version. Innovation therefore takes place if and only if $\pi_2^{k*} \geq \gamma$, i.e., with probability $G^k(p_2^{k*} + b^k - a^k)$ if and only if $a^k \geq b^k$ (no self-preferencing). Overall, the platform always gains $b^k + v^k$, unless app 2 introduces a superior version and it is squeezed. The expected platform's profit is then

$$\sum_k b^k + v^k + G^k(v^k + \Delta^k + b^k - a^k),$$

which is maximized at a^{k*} given in the statement, implying that the core ZLB binds under laissez-faire. By contrast, social welfare is maximized when investment incentives are socially efficient: $\pi_2^{k*} = \Delta^k$ – i.e., for $a^k \in [b^k, b^k + v^k]$. □

Remark (other Nash equilibria). We have focused throughout on the unique equilibrium outcome satisfying undominated strategies and platform pivotality. When $a^k > b^k$ and app 2 sells a superior version in market k, there are other Nash equilibria satisfying platform pivotality, in which app 1 in market k charges a price p_1^k below its opportunity cost $a^k - b^k$ – i.e., $p_1^k \in [0, a^k - b^k)$ and, as above, $p_2^* = \min\{p_1^k + \Delta^k, v^k + \Delta^k\}$ provided that app 2 makes a non-negative profit, and p_0^* is set to extract consumers' net utility across app markets in the considered equilibrium.[49] These equilibria with below-cost pricing involve a squeeze of the superior app (i.e. $\pi_2^{k*}(a^k) < \Delta^k$ for all $a^k > b^k$). As a result, $\hat{a}^k = b^k$ is the only access fee that yields a fair reward and is not subject to the multiplicity issue provided that independent apps do not perceive themselves as pivotal (part (ii) of our equilibrium definition).[50]

Remark (elicitation from the platform under monitoring of foreclosure). This setting with multiple heterogeneous app markets has been employed in Section 3 to study the implementation of the Pigouvian rule. In particular, Proposition 5 has established the impossibility of elicitation from the platform. Let us observe that this impossibility result does not hold if self-preferencing can be monitored or if vertical integration is not an option. A sketch of the proof goes as follows. Let $W^* \equiv \mathbb{E}[W^{k*}] = \mathbb{E}[v^k + b^k + \Delta^k]$ and $\Delta \equiv \mathbb{E}[\Delta^k]$ (we again denote $\Delta^k = 0$ if app 2 in market k does not innovate). Truth-telling yields platform payoff $W^* - \Delta$. A deviation yields platform payoff $\mathbb{E}[W^{k*}\mathbb{1}_{\{a^k \leq v^k + b^k + \Delta^k\}} - \pi_2^k(a^k)]$. If $a^k > v^k + b^k + \Delta^k$, the app-$k$ market disappears, which hurts the platform; if $a^k < b^k$, the platform leaves a supranormal rent to the app ($\pi_2^k(a^k) > \Delta^k$), which reduces the platform payoff unless this allows the platform to squeeze the superior app in some market l. But even in the latter case, the platform loses: if $b^k > b^l$ and $b^l + v^l \leq b^k \leq b^l + v^l + \Delta^l$, a permutation ($a^k = b^l$ and conversely) leads to $\pi_2^l(b^k) + \pi_2^k(b^l) = [v^l + b^l + \Delta^l - b^k] + [\Delta^k + b^k - b^l] = [\pi_2^k(b^k) + \pi_2^l(b^l)] + v^l$. The equilibrium outcome (although not the equilibrium strategies) is unique.

B.2 Ad-valorem access fees

Throughout the paper we considered for simplicity linear (per-unit) access fees. Here we show that our results are robust when considering instead ad-valorem fees (which are more often employed in reality): For each app sold by a third-party seller at price p_i, the platform gets tp_i and the seller $(1-t)p_i$, with $t \in [0,1]$. Let us first consider the hybrid platform case.

Lemma (hybrid platform). *For any ad-valorem access fee $t \in [0,1]$, in the hybrid platform case the equilibrium has the following features:*

[49] At prices below $a^k - b^k$, app 1 would lose money if the superior app were to raise its price and surrender the market to app 1; therefore, such equilibria are ruled out by the requirement (i) of the equilibrium definition that dominated strategies be eliminated.

[50] Finally, for all values of the access fee, there are also Nash equilibria in which p_2^k is again low, but for another reason: App 2 could internalize the consumers' participation constraint, i.e. perceive itself as pivotal (for example all app 2 providers would charge nothing if $p_0 = \sum_k v_2^k$); hence the need for condition (ii) of the equilibrium definition as well.

1. *If app 2 does not sell a superior version, or if it does but $b > \Delta$, the platform is better off foreclosing the third-party app for all t.*

2. *If app 2 sells a superior version and $b \leq \Delta$:*

   - *For $t \in [0, \frac{b}{\Delta})$: $p_1^* = 0$ (the app ZLB binds), $p_2^* = \Delta$, and $p_0^* = v$; hence, $\pi_0^*(t) = v + t\Delta < v + b$, and so the platform is better off foreclosing the third-party app (self-preferencing);*
   - *For $t \in \left[\frac{b}{\Delta}, \frac{b+v}{v+\Delta}\right]$, $p_1^* = \frac{t\Delta - b}{1-t}$, $p_2^* = \frac{\Delta - b}{1-t}$, and $p_0^* = v - \frac{t\Delta - b}{1-t}$; for any such t, $\pi_0^*(t) = v + b$ and $\pi_2^*(t) = \Delta$ (competitive neutrality);*
   - *For $t \in (\frac{b+v}{v+\Delta}, 1]$: $p_1^* = t(v + \Delta) - b$, $p_2^* = v + \Delta$, and $p_0^* = 0$ (the core ZLB binds); hence, $\pi_0^*(t) = t(v + \Delta) > v + b$ and $\pi_2^*(t) = (1 - t)(v + \Delta) + b < \Delta$ (superior app squeeze).*

Proof of Lemma. App 2 always makes positive per-consumer profit $(1 - t)p_2 + b$, and the platform prefers selling its own app as long as $tp_2 < b + p_1$. If app 2 does not sell a superior version, perfect Bertrand competition implies $p_2^* = p_1^* \leq (1 - t)p_2^* + b$, and so the app ZLB binds: $p_1^* = p_2^* = 0$ for all t. As the platform does not collect fees from app 2 sales, it finds it optimal to foreclose app 2 and so gets $v + b$.

If app 2 sells a superior version, given the app ZLB constraint, equilibrium app prices are

$$p_1^* = \max\{0, tp_2^* - b\} \quad \text{and} \quad p_2^* = \min\{p_1^* + \Delta, v + \Delta\}.$$

As long as app providers are unconstrained by consumers' willingness to pay, the equilibrium app prices are

$$\begin{cases} p_1^* = \frac{t\Delta - b}{1-t}, p_2^* = \frac{\Delta - b}{1-t} & \text{if } t \geq \frac{b}{\Delta} \\ p_1^* = 0, p_2^* = \Delta & \text{if } t < \frac{b}{\Delta} \end{cases}$$

Hence, for $t < \frac{b}{\Delta} \in (0, 1)$, $p_0^* = v$, and foreclosure is optimal (this result holds for all $t \in [0, 1]$ when $b > \Delta$): $\pi_0^*(t) = v + t\Delta < v + b \iff t < \frac{b}{\Delta}$. For $t \geq \frac{b}{\Delta}$, $p_0^* = v + \Delta - p_2^* = v + \frac{b - t\Delta}{1-t}$, and so $\pi_0^*(t) = v + b$ and $\pi_2^*(t) = v$. This is the equilibrium outcome as long as $p_2^* < v + \Delta$, which requires $t < \frac{b+v}{v+\Delta}$. For $t \geq \frac{b+v}{v+\Delta}$, $p_2^* = v + \Delta$, and so $p_1^* = t(v + \Delta) - b \in (v, p_2^* - \Delta)$ and $p_0^* = 0$. In this case, $\pi_0^*(t) = t(v + \Delta) > v + b$ and $\pi_2^*(t) = (1 - t)(v + \Delta) + b < \Delta$. □

Under ad-valorem fees, the platform can capture Δ, which is charged by the superior app, but cannot capture b. As a result, foreclosure is always optimal if $b > \Delta$; this is always true when app 2 does not sell a superior version – i.e., $\Delta = 0$. If instead $b \leq \Delta$, the equilibrium characterization with a hybrid platform mirrors the one under unit fees: for low (resp. high) values of the access fee, the app (resp. core) ZLB binds, and the platform is strictly better off foreclosing (resp. not foreclosing) a superior third-party app; for intermediate values of t, no ZLB binds, and the neutrality result holds.

Unlike in the basic model, however, this neutrality region does not exist in the pure platform case. This is because, unlike under linear fees, here an inferior third-party app provider has a lower opportunity cost, and thus is a tougher competitor to a superior seller, relative to the platform. This result has important implications on the platform's business model choice:

Lemma (platform business model). *If the platform is a pure player, for all $t \in [0,1]$, $p_1^* = 0$ and $p_2^* = \Delta$ (resp. $p_2^* = 0$) if app 2 sells (resp. does not sell) a superior version. Therefore, the platform gains from vertical integration combined with self-preferencing: (i) for all $t \in [0,1]$ if either $b > \Delta$ or $b \leq \Delta$ but app 2 does not innovate; (ii) only for $t < \frac{b}{\Delta}$ if $b \leq \Delta$ and app 2 sells a superior version. In the latter case, the superior app is squeezed for all $t > \frac{b}{\Delta}$.*

Proof of Lemma. If app 1 is also provided by a third-party seller, it finds it optimal to sell if and only if $(1-t)p_1 + b \geq 0$, and so $p_1^* = \max\{0, -\frac{b}{1-t}\} = 0$. As then $p_2^* = \min\{p_1 + \Delta, v + \Delta\}$, we obtain that, for all t, $p_1^* = 0$ and $p_2^* = \Delta$, with $\Delta = 0$ if app 2 does not innovate, and so $p_0^* = v$.

Then, as above, for all $t \in [0,1]$ the platform does not make access fee revenues if app 2 does not innovate, implying that it has incentives to vertically integrate and foreclose app 2 in these circumstances.

If, in contrast, app 2 innovates, it makes profit $\pi_2^*(t) = (1-t)\Delta + b$, whereas the pure-player platform obtains $\pi_0^*(t) = v + t\Delta$. Therefore:

- For all $t < \frac{b}{\Delta}$ (again, always if $b > \Delta$), the superior app makes a supranormal profit. The platform has incentives to vertically integrate, by acquiring the inferior app at a negligible price, and foreclose the superior app;
- For $t = \frac{b}{\Delta}$, the superior app obtains its fair reward, and the platform has no strict incentives to vertically integrate;
- For all $t > \frac{b}{\Delta}$, the superior app is squeezed, and the platform is strictly better off by operating as a pure-player platform. □

As app 2 always makes zero profit if it does not innovate, the superior version is introduced if and only if $t \geq \frac{b}{\Delta}$ (no self-preferencing) and $(1-t)\Delta + b \geq \gamma$. Accordingly, it is easy to derive the following results:

Proposition (optimal access fees). *Suppose $b < \Delta$. Then:*

(i) Profit-maximizing access fee. *The platform's profit is maximized at*

$$t^* = \min\left\{\frac{b}{\Delta} + \frac{1}{\Delta}\frac{G((1-t^*)\Delta + b)}{g((1-t^*)\Delta + b)}, 1\right\};$$

(ii) Welfare-optimal access fee. *Social welfare is maximized at* $\hat{t} = \frac{b}{\Delta} < t^*$*, at which a superior app receives a fair reward for its contribution to the ecosystem.*

Proof of Proposition. By the results in the Lemma above, the platform's gains the foreclosure profit $v + b$ for all $t \leq \frac{b}{\Delta}$. For $t > \frac{b}{\Delta}$, the platform still vertically integrates and gets $v+b$ if app 2 does not innovate, otherwise it gets a higher profit $t\Delta + v$ by remaining a pure player. As in this region app 2 innovates if and only if $(1-t)\Delta + b \geq \gamma$, the platform's expected profit equals $v + b + G((1-t)\Delta + b)[t\Delta - b]$, which is maximized for t^* given in the statement.

As consumer surplus is always extracted by the platform through the access price, ex-post social welfare is simply $W^* = b+v+\Delta x_2$, with $\Delta = 0$ if app 2 does not innovate, and so is maximized whenever there is no self-preferencing vis-à-vis a superior app: $t \geq \frac{b}{\Delta}$. Ex-ante expected social welfare maximization then requires the innovation to take place only when it is socially optimal – i.e., if $\Delta \geq \gamma$. This is equivalent to $\pi_2^*(t) = \Delta$ (fair reward), and so the welfare optimal fee is $\hat{t} = \frac{b}{\Delta} < t^*$. □

Note that the welfare optimal access fee is such that $p_2^*(\hat{t}) = \Delta$, so that the platform obtains $\hat{t}p_2^*(\hat{t}) = b$ from distributing a superior third-party app. Hence, optimal access fee regulation still follows the Pigouvian principle.[51]

B.3 Asymmetric ancillary benefits

We assumed for simplicity that all providers reap the same benefit from app distribution. This need not be the case. First, the platform may obtain a share of benefits when a third-party app is sold, as when Google shares data with independent apps in its ecosystem. Second, the benefits from app distribution may depend on provider-specific features. We therefore allow the ancillary benefit to take the form $b^{\dagger}_{ix_i}$ for app i, where $x_i = 1$ if the consumer chooses app i and $x_i = 0$ otherwise. Letting $b_1 \equiv b^{\dagger}_{11} - b^{\dagger}_{10} > 0$ for a in-house app and $b_i \equiv b^{\dagger}_{ix_i} > 0$ for an independent app i, and assuming, without loss of generality,[52] that $\Delta > b_1 - b_2$, the analysis carries through with appropriate modifications:

Proposition (asymmetric ancillary benefits). *Consider a hybrid platform. With asymmetric ancillary benefits* $b^{\dagger}_{ix_i}$*, denoting* $b_1 \equiv b^{\dagger}_{11} - b^{\dagger}_{10}$ *and* $b_2 \equiv b^{\dagger}_{21}$*, with* $\Delta > b_1 - b_2$*:*

(i) The platform has an incentive to engage in self-preferencing vis-à-vis a superior app if and only if $a < b_1$*.*

(ii) Any $a \in [b_1, b_1 + v]$ *yields a fair reward to a superior app and so maximizes social welfare; these levels are strictly lower than the platform's profit-maximizing fee,* $a^* = v + b_1 + G(b_2 + v + \Delta - a^*)/g(b_2 + v + \Delta - a^*)$*.*

[51] As in the model with linear access fees, the Pigouvian rule also applies if self-preferencing can be monitored, because lower values of the access fee would result in socially excessive innovation.

[52] $\Delta + b_2 < b_1$ would mean that the "superior" app creates less total value than the "inferior" app, a case that we noted is uninteresting in the hybrid platform case.

Proof of Proposition. If app 2 does not introduce a superior version, it always makes zero profit in equilibrium: (i) If $b_2 > b_1$, app 1 has a cost-disadvantage vis-à-vis app 2, implying equilibrium prices $p_1^* = p_2^* = \max\{a - b_1, 0\}$, with all consumers buying app 2 (whenever app 2 is viable); but then the platform makes less than the foreclosure profit $v + b_{11}^\dagger$ and so has incentives to favour its own app (the same result applies for $b_2 = b_1$). (ii) If $b_2 < b_1$, app 2 has a cost-disadvantage vis-à-vis app 1 and so cannot make any sale in equilibrium.

If app 2 introduces a superior version, equilibrium app prices are $p_1^* = \max\{a - b_1, 0\}$ and $p_2^* = \min\{p_1^* + \Delta, v + \Delta\}$ (provided app 2 is viable). The third-party app's profit is then $\pi_2^* = \Delta + (b_2 - b_1)$ in the neutrality region $[b_1, b_1 + v]$, and $\pi_2^* = b_2 + v + \Delta - a$ in the squeeze region.

The third-party app's contribution to the ecosystem is now $\Delta + (b_2 - b_1)$. Therefore, picking access fees in the neutrality region creates neither a squeeze nor an incentive for self-preferencing, and yields socially efficient innovation incentives – i.e., app 2 innovates if and only if $\Delta + (b_2 - b_1) \geq \gamma$.

By contrast, the platform gains $v + b_{11}^\dagger$ for all $a \leq b_1 + v$, but a higher profit $a + b_{10}^\dagger$ in the squeeze region if app 2 innovates. Hence, platform's expected profit is $v + b_{11}^\dagger + G(b_2 + v + \Delta - a)[a - (v + b_1)]$ for $a > b_1 + v$. Profit maximization yields $a^* > b_1 + v$ given in the statement. □

The analysis can also be generalized to pure platforms. The novel insight is that a pure platform may engage in non-price foreclosure, even if the access conditions are "non-discriminatory", i.e., if the access fee a and the share $\xi \in (0, 1)$ of ancillary benefits received by the platform are app-independent; the platform then may reduce the attractiveness of an app that is preferred by consumers – i.e., that has introduced the innovation – but offers a low ancillary benefit. To see this, suppose that $a \leq b_{21}^\dagger < b_{11}^\dagger$. If $\delta_2 = 0$, equilibrium app prices are $p_1^* = 0$ and $p_2^* = \Delta$, and so $p_0^* = v$ and, as consumers select app 2, $\pi_0 = v + a + \xi b_{21}^\dagger$. The platform would then make more money by setting δ_2 slightly above Δ, so that $p_1^* = p_2^* = 0$, consumers select app 1 and the platform still charges $p_0^* = v$ and makes higher profit $\pi_0 = v + a + \xi b_{11}^\dagger$.

B.4 Physical devices

Suppose that consumers can access an app store (Apple's App Store, Google Play, Microsoft Store...) only upon purchase of a costly physical device (smartphone, laptop, or game console). Assume that the device brings stand-alone value v_d, the same for all consumers. For example smartphones can be used for "non-gatekeeping purposes" such as taking pictures and making calls; likewise desktops have other usages than supporting services intermediated by a gatekeeper. Let c_d denote the device's production cost (the ancillary benefit is denoted $b^\dagger$). As in the basic model, all prices are set simultaneously and consumers then take their consumption decisions.

If the device is "cheap" relative to its stand-alone value, that is if $v_d \geq c_d$, consumers own IT regardless of whether there is a competitive original equipment manufacturers (OEM) sector or a monopoly, vertically integrated platform, and so the foregoing analysis is literally unchanged:[53] the platform engages in margin squeeze of a superior app by setting $a = a^*$, whereas the socially optimal access fees are the strictly lower levels $a \in [b, b+v]$, with $b \equiv b^\dagger$, at which a superior app provider obtains its fair reward.

Next consider a "costly" device, i.e. $v_d < c_d$.

(a) Suppose first that the platform is *not vertically integrated into device manufacturing* (and cannot subsidize device manufacturers); instead, the device (say, an Android-powered smartphone) is manufactured by a competitive OEM industry. The device is sold at cost, i.e. at price c_d.

The consumer surplus from the apps must exceed $c_d - v_d > 0$. For that reason, the equilibrium cannot be in the squeeze region, as this would imply that a superior app is constrained by, and charges the consumer's willingness to pay for the app. This means that a superior app must receive Δ and that the gross surplus to be divided between the consumer and the platform is $v + b^\dagger$. Letting

$$b \equiv b^\dagger - (c_d - v_d),$$

the platform must set $a \leq b + v$. Then, the market exists if and only if $v + v_d - c_d = v + b - b^\dagger \geq 0$, because the core ZLB prevents the ancillary benefit $b^\dagger$ from being passed through to consumers (as would also be the case of Δ if a superior app were squeezed).

(b) When *the device is produced by a vertically integrated platform* (as is the case for, e.g., videogame platforms) or the platform can subsidize device manufacturers, the core ZLB can be circumvented by subsidizing the device. The platform can then (i) squeeze a superior app by setting $a = a^*$, and (ii) charge price v_d for the device (implying a loss $c_d - v_d$ per device) and $p_0 = 0$ for the app store. Concretely, the platform can bundle device and app store and sell the bundle at v_d. The market then exists provided that the total surplus is positive: $v + b + G(v + \Delta + b^\dagger - a^*)[a^* - (v + b^\dagger)] \geq 0$.

Proposition (devices). *Suppose that the device is produced by a competitive sector. Let $b^\dagger$ denote the ancillary benefit and $b \equiv b^\dagger - \max\{0, c_d - v_d\}$ denote the adjusted ancillary benefit.*

(i) When the device is cheap ($b = b^\dagger$), the baseline analysis is unchanged.

(ii) When the device is costly ($b = b^\dagger - (c_d - v_d)$):

[53]The only difference between these two cases is that with a competitive OEM sector the device is sold at cost and access to the app store is sold at p_0^* characterized in Lemma 1 (the consumers therefore obtain surplus $v_d - c_d$), whereas a monopoly, vertically integrated platform can bundle device and app store at price $v_d + p_0^*$, extracting consumer surplus.

- *When the platform is not vertically integrated into device manufacturing, the market exists if and only if* $v+b-b^{\dagger} \geq 0$. *The platform optimally sets* $a = b+v$ *and* $p_0 = 0$. *It does not squeeze superior apps.*

- *When the platform is vertically integrated into device manufacturing, the market exists if and only if* $v + b + G(v + \Delta + b^{\dagger} - a^*)[a^* - (v + b^{\dagger})] \geq 0$. *The platform bundles device and app store at price* v_d *and squeezes a superior app* ($a^* > v + b^{\dagger}$ *is as in the basic model), and so has an incentive to vertically integrate into device manufacturing.*

- *If* $v + b - b^{\dagger} < 0 < v + b + G(v + \Delta + b^{\dagger} - a^*)[a^* - (v + b^{\dagger})]$, *allowing vertical integration into device manufacturing and (at least "some") squeeze is necessary for the ecosystem's viability.*

An interesting result here is that vertical integration into devices can enable ecosystem viability. The intuition relates to the familiar Tinbergen rule requiring at least as many independent instruments as there are targets: the access fee alone cannot achieve two contradictory goals. To make the platform more attractive to consumers, a must make apps cheap and therefore be low. But that may not suffice to induce the consumers to purchase the costly device. The second instrument is a platform subsidy to such purchases. In turn, a platform integrated into device manufacturing may not want to pay this subsidy unless the apps themselves are put to contribution through a high access fee. This high fee can be offset on the user side through the subsidy instrument.

B.5 Platform viability and entry

Assume that the social welfare function is $U + \omega\Pi$, where Π is total profit (platforms and apps) and $\omega \in (0, 1)$ is the weight on industry profits relative to consumer surplus U.[54] Suppose that there is (sequential) entry into the platform segment, with entry cost J. Suppose further that self-preferencing cannot be monitored, and so the access fee must be no lower than b.

From the results of Section 4.1, it follows that the socially optimal number of platforms is at most two, because extra platforms beyond $N = 2$ do not alter the consumer surplus and variable profit, but add entry costs and so are necessarily suboptimal. Given that, with $N = 3$, each entrant makes $\frac{b}{3}$ under the Pigouvian rule, and higher profits under laissez-faire, for all $J \leq \frac{b}{3}$ there is always too much entry into the platform segment,

[54] Under a social welfare standard ($\omega = 1$), platform competition just entails socially wasteful duplicative entry costs: welfare maximization dictates $N = 1$. On the contrary, under a consumer surplus standard ($\omega = 0$), as a monopolist brings zero net value to consumers, entry by any number $N \geq 2$ of platforms would be optimal (i.e., there is never excessive entry in equilibrium from consumers' standpoint).

which, without monitoring of non-price foreclosure, cannot be prevented by access fee regulation.

Under a monopoly platform, the consumers obtain no surplus ($U = 0$). A second entrant increases consumer surplus by v, at the expense of platform total profit, but also entails a socially wasteful entry cost J. Formally, $U + \omega\Pi = v + \omega(b + \Delta - 2J)$ under duopoly and $U + \omega\Pi = \omega(b + v + \Delta - J)$ under monopoly. Hence, a duopoly is preferred to monopoly if and only if $(1 - \omega)v \geq \omega J$, or $J \leq \frac{1-\omega}{\omega}v$.

Thus, assuming that the access fee is set, by regulation, at the Pigouvian level, for $J \in [\max\{\frac{1-\omega}{\omega}v, \frac{b}{3}\}, \frac{b}{2}]$ there is again too much entry, as two platforms enter but it would be optimal to have one. The region of parameters where excessive entry prevails of course expands when platforms are free to set access fees, as the absence of regulation increases their profits. If, on the contrary, $J \in (\frac{b}{2}, \min\{\frac{1-\omega}{\omega}v, b + v\}]$, then spurring the welfare-maximizing second entry requires setting the access fee above the Pigouvian level. Similarly, if $J \in (b+v, b+v+\Delta]$, there is a potential trade-off between the first platform's viability, which requires a squeeze in the app's profit, and app viability, which calls for staying away from the squeeze region to obtain the proper level of innovation.

Proposition (entry). *Because the core ZLB prevents platform profits from being competed away, socially excessive entry prevails when the entry cost is low and foreclosure cannot be monitored. By contrast, for high entry costs, setting access fees above the Pigouvian level is desirable to spur platform entry, if no other instrument is available (as we saw, $a > b$ introduces distortions).*

These results suggest that, while access fee regulation is an effective instrument to achieve fairness, thereby promoting efficient entry and investment decisions in the app segment, it may not be a jack of all trades, able to take on extra tasks such as ensuring contestability of the core segment.

B.6 Off-path measurement and appeals

Measuring b systematically would imply considerable costs and delays. At best one can, when the access fee is determined by the platform, allow appeals that hopefully will not be frequent if the incentive scheme is designed properly. Let us focus on the more interesting case of a hybrid platform and a superior third-party app (as any inferior third-party app is never foreclosed and makes zero profit for all $a \geq b$ and the platform would never set $a < b$).

Suppose the regulator is equipped with a noisy measure of the ancillary benefit if called upon by a party. More precisely, the platform chooses the access conditions $\{a^k, \delta_2^k\}$ (of which the regulator observes only a), then prices are set, and finally the superior third-party app chooses whether to appeal "against a high access fee". In this appeal procedure, the authority observes a noisy, but unbiased, version $\tilde{b}^k$ of the ancillary benefit, with cdf $R(\tilde{b}^k)$ such that $\int_{\mathbb{R}} \tilde{b}^k \mathrm{d}R(\tilde{b}^k) = b^k$. If $a^k > \tilde{b}^k$, then the access fee is assessed to be unfair, and the defendant (the platform) must pay a fine $\tau(a^k - \tilde{b}^k)$ (with $\tau > 0$) to the plaintiff (the third-party app); and vice versa if $a^k \leq \tilde{b}^k$.

The outcome of the appeal procedure interferes neither with the platform's choice of whether to foreclose the third-party app, nor with access and app prices chosen by the firms before the appeal. This implies that the third-party app will appeal whenever $\int_{\mathbb{R}} \tau(a^k - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k) > 0 \Leftrightarrow a^k > b^k$. So if $\tau \geq 1$, the platform does not gain from inflating the access fee beyond b^k.

This analysis however understates the platform's ability to extort high access fees. Let us therefore "empower" the platform by allowing it to foreclose the third-party app *after* the latter has appealed (silence means assent, so the absence of appeal means that the proposed a applies). Suppose that the platform faces a sequential entry of superior third-party apps in distinct but identical app submarkets and discounts future profits at a rate β. The platform can therefore build a reputation for preying on apps that dare to appeal.

Proposition (appeals). *Give the third-party app a right to appeal against a high access fee chosen by the platform. If the regulator can produce a noisy measure $\tilde{b}^k$ of the ancillary benefit, and impose sufficiently large fines to the platform if it loses the appeal – namely, $\tau(a^k - \tilde{b}^k)$ if $a^k > \tilde{b}^k$, with $\tau \geq \max\{1, \beta/(1-\beta)\}$ – then the Pigouvian rule can be implemented even when the platform can build a reputation for engaging in foreclosure after being challenged by an app.*

Proof of Proposition. The arguments above imply that the third-party app in market k will appeal whenever $\int_{\mathbb{R}} \tau(a^k - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k) > 0 \Leftrightarrow a^k > b^k$.

Moving backwards to the pricing stage, $p_1^k = 0$ and $p_2^k = \Delta^k$ is the worst-case scenario for the third-party app in any platform-pivotality equilibrium. Then, for all $a^k \leq b^k + \Delta^k$ (no access-price foreclosure), the third-party app's profit, absent non-price foreclosure, is at least $\pi_2^k = \Delta^k + b^k - a^k$. Because total profit is at most $v^k + b^k + \Delta^k$, the platform's maximal expected profit from setting any $a^k > b^k$ and therefore being challenged is

$$v^k + a^k - \int_{\mathbb{R}} \tau(a^k - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k),$$

which is decreasing in a^k provided $\tau > 1$. In contrast, the platform makes profit $v^k + b^k$ either by setting $a^k < b^k$ and foreclosing the superior app, or by choosing $a^k = b^k$, while any other (a^k, δ_2^k)-choice yields strictly lower profit. Therefore, the Pigouvian principle can always be implemented by giving the platform a tiny advantage in the appeal procedure – e.g., the appeal benefits the third-party app if and only if $a^k > \tilde{b}^k + \epsilon$ for a small positive ϵ, so that a small squeeze is tolerated and the platform strictly prefers not to foreclose.

The observation that the third-party app appeals for any $a^k > b^k(+\epsilon)$ crucially hinges on the fact that appealing has no impact on its market profit. This would not be the case if the platform had the possibility (and the incentive) to foreclose it *post* appeal. When such a post-appeal foreclosure threat is credible, the third-party app does not appeal whenever $\Delta^k + b^k - a^k \geq \int_{\mathbb{R}} \tau(a^k - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k)$, or equivalently $a^k \leq a^\dagger \equiv b^k + \frac{\Delta^k}{\tau+1}$. If τ is large enough relative to the platform's discount factor β, however, such reputation building strategy can be prevented. To see this, suppose for simplicity that, by foreclosing after $a = a^\dagger$ is appealed in the first market, the platform is able to secure profit $v^k + a^\dagger$ forever after, which implies a discounted extra profit $\frac{\beta\Delta^k}{(1-\beta)(1+\tau)}$ from future markets relative to the profit $v^k + b^k$ it obtains by proposing $a^k = b^k$,[55] at an expected loss $\int_{\mathbb{R}} \tau(a^\dagger - \tilde{b}^k)\mathrm{d}R(\tilde{b}^k) = \tau\frac{\Delta^k}{1+\tau}$ from the appeal. Therefore, setting $\tau \geq \frac{\beta}{1-\beta}$ prevents such reputation-building strategy. □

[55] For $\tau > 1$, this profit in turn exceeds the profit from proposing $a^\dagger$ in the first market and not foreclosing after being challenged, thereby failing to build a reputation for foreclosing future apps.

Chapter III

Externalities of Responsible Investments

This chapter is based on joint work with Alessio Piccolo (Indiana Kelley) and Jan Schneemeier (MSU Broad). All authors contributed equally to this research.

1 Introduction

Companies face increasing pressure from investors to include environmental and social factors in their policies. By March 2024, 4,827 investors managing $128 trillion have signed the United Nations Principles for Responsible Investment (UN PRI), pledging to incorporate corporate social responsibility (CSR) issues into their investment analysis and ownership policies. In 2020, $1 in every $3 under professional management in the United States was allocated to sustainable investments (Edmans, 2023). The primary rationale for this socially responsible investing (SRI) is to change or divest from firms that exert negative externalities, to reduce their harm to society.

In a world where political institutions fail to control externalities, responsible investors can have a positive impact by serving as substitutes for regulation and other forms of government intervention. While there is extensive literature on the limits to government intervention (e.g., Acemoglu, 2003; Besley and Persson, 2023), the inefficiencies that may arise when capital markets try to control firms' externalities are less understood. The objective of this article is to investigate the impact of SRI in a framework where its potential inefficiencies, and the externalities that such inefficiencies may impose on other stakeholders, are carefully articulated and analyzed.

Our starting point is a traditional model of firms' externalities, where firms' private incentives are insufficient to incentivize the socially optimal level of externalities reduction. For the sake of concreteness, we interpret firms' externalities as the pollution they generate through production, and externalities reduction as their efforts to abate pollution. We use the model to study the following question: Suppose some investors care about externalities and can influence firms' policies (e.g., by engaging in governance, or because managers cater to the preferences of their shareholders), but only feel "responsible" for the externalities generated by the firms in their portfolios; how do they invest in equilibrium, and with which effects on the economy?

Sustainable funds are typically evaluated based on the average ESG scores of the stocks in their portfolios (see, e.g., Morningstar's "globe" ratings of funds); consistent with this observation, we assume that responsible investors, who make up a fraction of total investors in the market, suffer a disutility from holding polluting firms.[1] The remaining investors are "non-responsible", in the sense that they only care about financial returns. Both types of investors submit demands for the firms' shares in a financial market, and share prices adjust to equate demand and supply.

Our first main insight is that, when the fraction of investors who are responsible is

[1]Our modeling of responsible preferences is consistent with the traditional explanations of individuals' demand for CSR (i.e., warm-glow and image concerns (Andreoni, 1990; Bénabou and Tirole, 2010) and similar to other papers on SRI (see, e.g., Heinkel et al., 2001; Pástor et al., 2021; Goldstein et al., 2022). For evidence on warm-glow preferences for investors see Heeb et al. (2023) and Bonnefon et al. (2025). Hartzmark and Sussman (2019) document that mutual funds with higher globe ratings attract larger flows.

not too large, responsible investors prefer to adopt investment strategies that lead to a concentration of SRI in a few firms in the economy. The basic idea is easy to explain: by coordinating their portfolio choices to concentrate their investments in only a few firms, responsible investors can create enough price pressure on these firms' shares, so that non-responsible investors prefer to go after other firms. This way, responsible investors can acquire large ownership stakes in the firms they target, and significantly change their abatement policies, at the cost of smaller financial returns on their holdings. The equilibrium strategies of responsible investors resemble some of the typical SRI strategies in practice (e.g., best-in-class investing and ESG exclusion)[2] and are broadly consistent with the concentration of green capital we observe in the data (Figure 1).

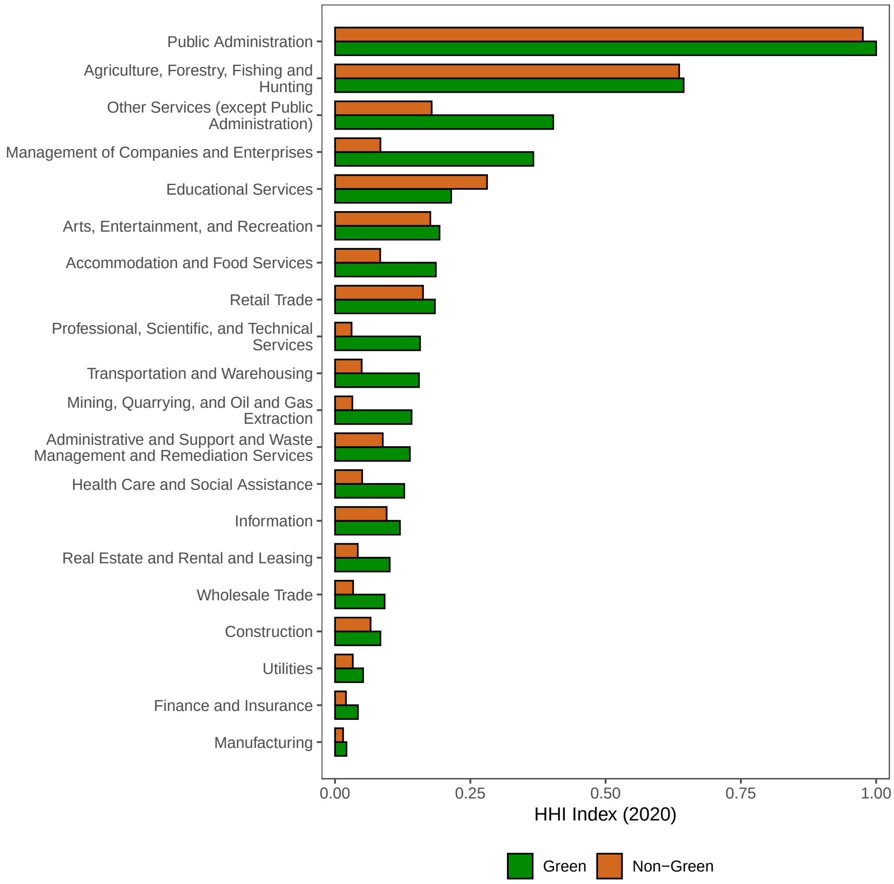


Figure 1: This figure plots HHI indexes of Green and Non-Green Capital at the industry level for NYSE/NASDAQ-listed stocks as of 2020. Industries are defined by NAICS 2-digit codes. NYSE/NASDAQ-listed stocks are from CRSP.

The following step of our analysis is understanding how this concentration of responsible investments affects aggregate abatement, allowing for interdependence in firms' private incentives to abate, first in a general, reduced-form way (similar to Bulow et

[2]Best-in-class investing (or positive screening) refers to the approach of selecting companies that are leaders in their industry or equity sector in terms of ESG criteria. ESG exclusion (or negative screening) refers to the strategy of excluding specific companies or sectors that do not meet certain ESG thresholds. See Starks (2023) for a taxonomy of the typical ESG investing strategies.

al., 1985), and then by microfounding the interdependence through firms' interactions in product markets. The abundance of SRI then spurs the abatement efforts of the firms in responsible investors' portfolios, but it also has an indirect effect on the other, "excluded" firms.

First, suppose that abatement efforts are strategic complements, so that the indirect effect is positive – that is, the excluded firms follow the lead of the targeted firms and also abate more. Complementarity is most likely to arise when there are positive spillovers in developing green technologies (as documented by Aghion et al. (2016) for the auto industry), or when firms' products are complements in consumers' preferences (e.g., electric vehicles and charging stations). In this case, SRI concentration can be the most efficient way to overcome free-riding and coordination issues in adopting green technologies, and aggregate abatement always increases with SRI compared to a world where no investor is responsible.

Next, consider a setting where abatement efforts are strategic substitutes, so the indirect effect is negative – that is, the increased abatement of the firms targeted by responsible investors crowds out those of the excluded. Substitutability is most likely to arise when firms compete for consumers in product markets (e.g., when the traditional "Schumpeterian effect" of competition is at play, as documented by Aghion et al., 2005) or for resources in input markets (e.g., lithium batteries for electric vehicles: see Speirs et al., 2014). From a social welfare perspective, SRI concentration is then less desirable in these environments, to the extreme that the crowding-out can outweigh the direct effect, in which case aggregate abatement would be larger without responsible investors. Even when concentrated SRI increases aggregate abatement, it may still be undesirable: The concentration of SRI helps the "responsible" firms gain product market power, which then allows them to charge higher prices for green products, so that consumption and welfare might still be lower than in a world without SRI.

Perhaps surprisingly, responsible investments concentrate the most in our model when concentration is least desirable – that is, when the crowding-out effect described above is strongest. The reason is that the crowding out makes excluded firms less green and, thus, less attractive to responsible investors, which reinforces their preference for concentration. It is worth emphasizing that, when a majority of investors in the market are responsible, concentration is no longer necessary, so SRI becomes more uniformly distributed across the firms in the economy. Our model thus suggests that, while SRI is likely to help the transition to a greener economy if it becomes the norm in financial markets, it may generate inefficiencies in the interim or if it remains a somewhat limited phenomenon.

A more subtle type of inefficiency of SRI relates to this transition toward a more uniform distribution of responsible investments in the economy. In our model, responsible investors influence firms' abatement policies and, at the same time, select companies based on their expectations about these policies. This feedback loop between SRI and abatement efforts is such that, for intermediate values of the fraction of responsible investors, multiple SRI distributions, with different levels of concentration and aggregate abatement, can be supported in equilibrium. In environments with independence in abatement efforts, we show that responsible investors favor concentration even when a more uniform distribution of SRI would be possible in equilibrium and would yield larger aggregate abatement.

Overall, our results highlight an inherent tension in the role of SRI as a mechanism to control externalities: When their mass is sufficiently small, responsible investors need to concentrate to have a significant impact on firms' policies; however, when they do, their preferences become more detached from aggregate abatement, as they fail to internalize the abatement choices of the excluded firms. So they may end up creating a few "green islands" in a "sea of brown firms," instead of driving a more generalized change toward a greener economy, even when SRI becomes more popular, and driving such a change would be feasible.

Contributions. We make three main contributions to the literature. First, we contribute to the literature on SRI by providing a theoretical analysis of its distributional efficiency. By studying asymmetric equilibria in a framework with multiple firms, we are able to explore the effects of SRI concentration for excluded firms and aggregate externalities. We also examine how SRI concentration spills over to product markets by affecting firms' competitive positions.

While there is extensive literature on the direct effects of SRI on firm externalities (Heinkel et al., 2001; Broccardo et al., 2022; Edmans et al., 2022; Gupta et al., 2022; Oehmke and Opp, 2025), the spillovers to other firms and markets are less understood. The papers closest to ours are those that explore SRI in general equilibrium models with perfect competition (Hakenes and Schliephake, 2021; Landier and Lovo, 2025). Besides exploring asymmetric equilibria and imperfect competition, we consider a general model of interdependence in abatement choices, which allows for both complementarity and substitutability.[3] The analysis of the interaction between SRI and other forms of incentives to limit externalities (e.g., the pricing of CSR issues in product markets) also connects our paper to those exploring the interplay between SRI and regulation (Biais and Landier, 2022; Piatti et al., 2022; Huang and Kopytov, 2023).[4]

Second, we contribute to the vast literature on the objectives of the firm. The traditional view that firms should primarily maximize profits (Friedman, 1970) has been

[3]Landier and Lovo (2025) consider a setting with two production sectors where supply-chain interactions generate a form of complementarity in firms' toxic emissions. Hakenes and Schliephake (2021) abstract from interdependence in abatement efforts. Both papers focus on symmetric equilibria, so they abstract from the implications of SRI concentration.

[4]Other work on SRI focuses on its implications for financial market outcomes, like expected returns (Pástor et al., 2021) and price informativeness (Goldstein et al., 2022).

challenged by several recent papers (Elhauge, 2005; Bénabou and Tirole, 2010; Hart and Zingales, 2017). They argue that when political institutions fail to control externalities (due to commitment problems and political failures (e.g., Acemoglu, 2003; Besley and Persson, 2023), firms should internalize shareholders' social preferences, potentially pursuing social goals at the expense of profits. We explore the efficiency of this market-based approach to controlling firm externalities. Our paper fits into a broader line of research exploring how individual social preferences influence public goods provision (e.g., Bénabou and Tirole, 2006; Besley and Ghatak, 2007; Kaufmann et al., 2024).

Third, we provide a parsimonious model for studying the general equilibrium implications of heterogeneity in investor preferences. While we focus on investors' social attitudes, other dimensions of heterogeneity are relevant for other applications. There is a large body of evidence that investors differ along multiple dimensions, like their investment horizons (Bushee, 1998; Gaspar et al., 2005) governance attitudes (Bubb and Catan, 2022), and social and political ideologies (Bolton et al., 2020). A growing theoretical literature studies how such heterogeneity affects the feedback between trading and governance decisions (Kakhbod et al., 2023; Levit et al., 2024) in single-firm settings. We derive a tractable characterization of (asymmetric) equilibria where ex-ante identical firms differ in their ownership and consequently implement different policies, which allows us to study how investors' heterogeneity affects the distribution of firms' ownership and policies in the economy.

2 A model of Socially Responsible Investments

The model consists of two dates, $t \in \{1, 2\}$, and a finite number of publicly traded firms, $j \in \mathcal{J} \equiv \{1, \ldots, N\}$. At time $t = 1$, investors trade claims to the firms' profits, which determines firm ownership. At time $t = 2$, firms first make costly investments to limit their externalities; then, their profits realize and are distributed to shareholders. All agents in the model are rational, and for simplicity, we assume that there is no discounting.

2.1 Externalities and firm policies

All firms are ex-ante identical. Each firm generates a negative externality λ at a rate $1-\sigma_j$. For concreteness, we refer to λ as the pollution generated by the firm's production process, and to the actions firms take to limit their externalities as abatement efforts.[5] We denote these abatement efforts by $\sigma_j \in [0, 1]$, where σ_j can be interpreted either as the percentage reduction in pollution or as the probability of adopting a clean technology that

[5]More broadly, λ may also capture other externalities, such as the societal cost of a negative corporate culture.

fully abates pollution. The cost of abatement is $\mathcal{C}(\sigma_j)$, which is continuously differentiable, increasing, and convex. Firm j's expected profit Π_j is given by $\Pi(\sigma_j, \vec{\sigma}_{-j}) \equiv \pi_j - \mathcal{C}(\sigma_j)$, where $\pi_j = \pi(\sigma_j, \vec{\sigma}_{-j})$ denotes expected profits gross of abatement costs. We assume that π_j is positive, continuously differentiable in both arguments, and that it is increasing and concave in σ_j. The assumption that π_j increases with σ_j captures firms' *private* incentives to abate, that is, independent of their ownership (more on this shortly), which may reflect pressure from consumers, workers, or government regulation.

We assume that π_j also depends on the abatement efforts of other firms, which are collected in the vector $\vec{\sigma}_{-j}$. This means that j's private incentives to abate can depend on $\vec{\sigma}_{-j}$, which allows for interdependence in abatement efforts. Similar to Bulow et al. (1985), we first capture this interdependence in a general, reduced-form way, through the properties of the cross derivatives $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}}$, for all $j' \neq j$. If $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} > 0$, abatement efforts are strategic complements, and more efforts by another firm strengthen firm j's private incentives to abate. Examples of strategic complementarities include technological spillovers in developing green technologies or firms selling complementary goods. If $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} < 0$, abatement efforts are strategic substitutes, which is the case, for instance, when firms compete for consumers in product markets or for resources in input markets. (iii) Finally, $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} = 0$ describes a benchmark where the abatement efforts are independent across firms. In Section 5, we micro-found this interdependence by explicitly modeling firms' interactions in product markets.

2.2 Ownership market and firm objective

At time $t = 1$, investors' trading determines firms' ownership. Each firm has a fixed supply of shares, normalized to one, traded in a financial market. A unit mass of atomistic investors, indexed by $i \in [0, 1]$, simultaneously submit their demand schedules for the shares of each firm. The market-clearing price p_j adjusts to equate demand and supply.

Investors have heterogeneous preferences. A fraction $1-\chi$, with $\chi \in (0, 1)$, are purely profit-motivated, in that they maximize the expected monetary return from their holdings – that is, the expected value of their claims to firms' profits net of the share prices and a trading cost K_i. The remaining investors also suffer a disutility from holding polluting firms (e.g., Heinkel et al., 2001; Pástor et al., 2021; Goldstein et al., 2022), as they internalize the societal cost of pollution of firms in their portfolios. We refer to the latter as socially responsible investors ($i \in \mathcal{R}$), and to the former as non-responsible investors ($i \in \mathcal{N}$). Formally, for a given vector of anticipated abatement choices $\vec{\sigma} \equiv (\sigma_1, ..., \sigma_N)'$, investor i solves:

$$\max_{\vec{s}_i \geq \vec{0}} \sum_{j \in \mathcal{J}} s_{ij} \left[\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}} \lambda (1 - \sigma_j) \right] - K_i, \tag{1}$$

where $\vec{s}_i \equiv (s_{i1}, ..., s_{1N})'$ collects the number of shares investor i holds in each firm j, and $\mathbb{1}_{i,\mathcal{R}} = 1$ if i is responsible ($i \in \mathcal{R}$) and 0 otherwise ($i \in \mathcal{N}$).[6] We assume that the trading cost of purchasing $\sum_j s_{ij}$ shares is $K_i = \mathcal{K}\left(\sum_j s_{ij}\right)$ and, to obtain closed-form solutions for the trading strategies, we set $\mathcal{K}(x) \equiv \frac{\kappa}{2}x^2$ (similar to Banerjee et al. (2018) and Dávila and Parlatore, 2021).[7] Investors' portfolio decisions determine firm ownership: we let $s_j^{\mathcal{R}} \equiv \int_0^1 s_{ij}\mathbb{1}_{i,\mathcal{R}}\mathrm{d}i$ denote the shares of firm j held by $\mathcal{R}$ investors, and $\vec{s}^{\mathcal{R}} \equiv (s_1^{\mathcal{R}}, \ldots, s_N^{\mathcal{R}})$ the distribution of responsible ownership in the economy.

At the beginning of $t = 2$, firms simultaneously choose their abatement levels. Each firm incorporates its shareholders' preferences into its objective function, so ownership influences abatement efforts. Formally, given its share of responsible owners $s_j^{\mathcal{R}}$, firm j chooses σ_j to solve:

$$\max_{\sigma_j \in [0,1]} \Pi_j - \lambda(1-\sigma_j)s_j^{\mathcal{R}}. \tag{2}$$

The objective in Program (2) is a weighted average of the expected payoff per share to investors, where the weights are equal to the shares held by each shareholder.[8] This specification is typically referred to as *proportional control* assumption (see, e.g., O'Brien and Salop, 1999; López and Vives, 2019), and captures different channels through which shareholders can influence managerial decisions in proportion to their stake in the firm.[9] In our setting, $\mathcal{R}$ investors internalize the cost of the pollution generated by the firms they hold. So, the larger the share of j's responsible ownership, the more the firm internalizes the societal cost of pollution λ when choosing its abatement effort.

To simplify the exposition, we impose the following assumption:

Assumption 1. $\lambda > \mathcal{C}'(1) > \frac{\partial \pi_j}{\partial \sigma_j}(1, \vec{1})$.

The first inequality in Assumption 1 implies that it is *socially* optimal for all firms to fully abate their pollution (i.e., to set $\sigma_j = 1$ for all j), since the societal cost of pollution is always larger than the cost of abatement. The second inequality implies that firms' *private* incentives are, however, insufficient to achieve this outcome: at $\vec{\sigma} = \vec{1}$, the (private) marginal benefit of abatement is lower than the marginal cost. This creates an

[6]Investors are atomistic, so they take aggregate outcomes as given. Adding a disutility for the aggregate externality to Program (1) would thus not affect the equilibrium characterization. More broadly, the equilibrium characterization carries through as long as $\mathcal{R}$ investors care relatively more about the pollution generated by the firms in their portfolio (see Appendix B.4).

[7]The assumption that $\mathcal{K}(\cdot)$ depends on the total size of i's portfolio (rather than its composition) is consistent with this cost reflecting both direct and indirect transaction costs, such as the borrowing or opportunity costs of raising funds for the investor. As shown in Appendix B.5, our results carry through if trading costs instead depend on the individual holdings of each firm.

[8]At the beginning of time $t = 2$, investors have already paid the share price and transaction cost. It follows that a $\mathcal{N}$ investor receives an expected payoff Π_j from holding a share of firm j, while a $\mathcal{R}$ investor receives an expected payoff $\Pi_j - \lambda(1-\sigma_j)$.

[9]Examples include voting (e.g., Levit and Malenko, 2011; Levit et al., 2024), exit, and voice (e.g., Edmans and Manso, 2011). Appendix B.3 analyzes a variation of the model where shareholders vote on whether the firm should internalize the pollution externality when choosing its CSR investment; our main results carry through.

opportunity for $\mathcal{R}$ investors to have a positive impact on the economy, by pushing firms' abatement efforts closer to the first-best.

2.3 Sequence of events and equilibrium definition

The timing of the model unfolds as follows.

t = 1: Investors trade and form their portfolios $\{s_{ij}\}$ for $i \in [0,1]$ and $j \in \mathcal{J}$.

t = 2: Firms choose abatement efforts σ_j; profits are realized and distributed to shareholders.

We use *subgame perfect equilibrium* as the solution concept[10] and restrict our attention to pure-strategy equilibria in which firms with identical ownership structures choose the same abatement levels.[11] An equilibrium consists of a collection $\{\{s_{ij}\}, \vec{\sigma}\}$, where $i \in [0,1]$ and $j \in \mathcal{J}$, that jointly solves Programs (1) and (2) and satisfies sequential rationality. Depending on the parameters, the model may feature two types of equilibria: (i) *symmetric equilibria*, in which all firms have the same proportion of responsible owners – formally, $s_j^{\mathcal{R}} = s_{j'}^{\mathcal{R}}$ for all $j, j' \in \mathcal{J}$ – resulting in uniform abatement levels; and (ii) *asymmetric equilibria*, in which responsible ownership varies across firms – i.e., $s_j^{\mathcal{R}} \neq s_{j'}^{\mathcal{R}}$ for some $j \neq j'$ – leading to heterogeneity in abatement decisions.

3 Equilibrium analysis

3.1 Preliminaries

We begin by characterizing the solutions to Programs (1) and (2) and describing the strategic interactions among investors.

[10]We assume the distribution of SRI $\vec{s}^{\mathcal{R}}$ becomes common knowledge after trading takes place. This means that stage $t = 2$ constitutes a proper subgame, so the standard equilibrium concept is subgame perfect equilibrium. Since firms correctly anticipate $\vec{s}^{\mathcal{R}}$, and each individual investor cannot impact $\vec{s}^{\mathcal{R}}$ by deviating from any given equilibrium, the equilibrium outcomes are the same if firm j only observes $s_j^{\mathcal{R}}$. The equilibrium concept in this case would be Perfect Bayesian equilibrium. In Appendix B.6, we show that our results are robust to a setting where this timing is reversed – i.e., firms first choose their abatement investments to attract investors and maximize their stock prices.

[11]This restriction is without loss of generality when we consider equilibria where there are some firms owned by both type of investors, as then the combination of $\mathcal{R}$ and $\mathcal{N}$ investors' incentive-compatibility conditions ensures that firms with the same $s_j^{\mathcal{R}}$ always choose the same σ_j. Otherwise, there may be equilibria where firms with the same ownership (namely, those with $s_j^{\mathcal{R}} = 0$) choose different abatement levels. However, by the results in Hefti (2017), such equilibria do not exist provided that strategic substitutability is not too strong.

Abatement efforts. Given its responsible ownership $s_j^{\mathcal{R}}$ and conjectures about the other firms' abatement efforts $\vec{\sigma}_{-j}$, firm j's optimal abatement is pinned down by the following first-order condition:

$$\Gamma\left(\sigma_j, \vec{\sigma}_{-j}; s_j^{\mathcal{R}}\right) \equiv \underbrace{\frac{\partial \pi_j}{\partial \sigma_j}(\sigma_j, \vec{\sigma}_{-j}) - \mathcal{C}'(\sigma_j)}_{\textit{Private abatement incentives}} + \underbrace{\lambda s_j^{\mathcal{R}}}_{\textit{SRI pressure}} \gtreqless 0, \tag{3}$$

where the inequality sign is $\leq$ ($\geq$) if $\sigma_j = 0$ ($\sigma_j = 1$), and it is an equality if $\sigma_j \in (0,1)$.

The function $\Gamma\left(\sigma_j, \vec{\sigma}_{-j}; s_j^{\mathcal{R}}\right)$ describes firms' overall abatement incentives, which consist of two components: their private incentives and the internalization of $\mathcal{R}$ owners' preferences. The private incentives depend on how σ_j affects expected profits $\Pi_j = \pi_j - \mathcal{C}(\sigma_j)$. Unless abatement efforts are strategically independent, the marginal impact of σ_j on Π_j also depends on the abatement choices of other firms: when other firms abate more, firm j's incentive to abate decreases when abatement efforts are strategic substitutes and increases when they are strategic complements. In these cases, an equilibrium value of $\vec{\sigma}$ will have to solve the first-order conditions in Eqn. (3) simultaneously for all firms. The second component depends on the share of responsible owners $s_j^{\mathcal{R}}$, with a larger $s_j^{\mathcal{R}}$ inducing firm j to abate more.

Portfolio choices. Investors' portfolio choices determine the equilibrium distribution of responsible ownership $\vec{s}^{\mathcal{R}}$. Investors are atomistic, so each investor i takes the price vector $\vec{p} \equiv (p_1, \ldots, p_N)$ as given when choosing its portfolio. Moreover, i rationally anticipates firms' abatement efforts and expected profits, and how these depend on the distribution of $\mathcal{R}$ investors across firms. The investor's optimal portfolio satisfies:

$$\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}} \lambda (1 - \sigma_j) \leq \kappa \sum_{j \in \mathcal{J}} s_{ij}, \tag{4}$$

where the inequality is strict only if i does not invest in firm j (i.e., if $s_{ij} = 0$).

The marginal cost of shareholding, i.e., the right-hand side of Eqn. (4), is constant across firms. To break ties, we assume that investors incur a small cost of acquiring shares in multiple firms, so each individual investor prefers to hold shares in one firm only.[12]

We can then write investor i's demand for firm j's shares as:

$$s_{ij} = \begin{cases} \max\left\{\frac{1}{\kappa}\left[\Pi_{j^*} - p_{j^*} - \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_{j^*})\right], 0\right\} & \text{for } j^* \in \text{argmax}_j \{\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_j)\} \\ 0 & \text{for } j \neq j^*. \end{cases} \tag{5}$$

[12]Without this tie-breaking assumption, each of the equilibria we characterize coexists with observationally equivalent ones (i.e., featuring the same distribution of SRI) where investors hold diversified portfolios, i.e., divide their optimal demands across firms.

i is willing to invest if the preference-adjusted return, $\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_j)$, is positive for at least one firm; otherwise, i does not invest in any firm. If the preference-adjusted return is positive for more than one firm, i invests in the one where it is the largest, j^*.

Market clearing and share prices. $\mathcal{N}$ types have, all else equal, higher valuations for firms that do not fully abate (i.e., for which $\sigma_j < 1$) and the same valuations as $\mathcal{R}$ types for those that do.[13] So $\mathcal{N}$ investors always hold some shares in equilibrium, while $\mathcal{R}$ investors may not. Let $\alpha_j^{\mathcal{R}}$ ($\alpha_j^{\mathcal{N}}$) denote the fraction of $\mathcal{R}$ ($\mathcal{N}$) investors buying positive shares of firm j, where the quantity purchased is given in Eqn. (5) above. The following market-clearing condition pins down j's equilibrium price:

$$\underbrace{\chi\alpha_j^{\mathcal{R}}\frac{1}{\kappa}\left[\Pi_j - p_j - \lambda(1-\sigma_j)\right]}_{s_j^{\mathcal{R}}} + (1-\chi)\alpha_j^{\mathcal{N}}\frac{1}{\kappa}\left[\Pi_j - p_j\right] = 1. \tag{6}$$

The first summand in Eqn. (6) is j's responsible ownership, $s_j^{\mathcal{R}}$, as it represents $\mathcal{R}$ investors' aggregate demand for j's shares. Solving this expression for the share price p_j yields:

$$p_j = \Pi_j - \underbrace{\frac{\kappa}{\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}}}}_{\textit{Liquidity discount}} - \underbrace{\frac{\chi\alpha_j^{\mathcal{R}}\lambda(1-\sigma_j)}{\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}}}}_{\textit{Pollution discount}}. \tag{7}$$

j's equilibrium share price equals its expected profits, net of two discounts. The first is a standard liquidity discount, which compensates investors for the trading cost. This discount is reduced when there is more demand for j (more investors choose $j^* = j$), so each investor holds smaller stakes and, thus, incurs lower trading costs. The second discount reflects the preferences of the $\mathcal{R}$ investors trading the firm (it compensates them for the disutility of holding a polluting firm), so it increases in both their mass, $\chi\alpha_j^{\mathcal{R}}$, and the pollution externality net of abatement, $\lambda(1-\sigma_j)$. Since Π_j is fully reflected in p_j, the financial return from buying j's shares is the sum of these two discounts.

In equilibrium, abatement efforts and investors' demands must jointly satisfy Eqns. (3) and (5), with the price vector $\vec{p}$ given in Eqn. (7).

Interaction among investors. The interaction between a firm's responsible ownership $s_j^{\mathcal{R}}$ and its abatement efforts σ_j is characterized by a two-way feedback: On the one hand, when σ_j goes up, $\mathcal{R}$ investors' return from holding firm j increases; so more $\mathcal{R}$ investors want to demand firm j's shares, and those that do so have valuations closer to $\mathcal{N}$ investors. Both these effects lead to an increase in $s_j^{\mathcal{R}}$. On the other hand, when $s_j^{\mathcal{R}}$

[13] Appendix B.7 shows that our qualitative results are unchanged a setting where firms' investments bring up a positive externality λ at a rate σ_j, which increases the utility of responsible agents, who then have higher valuations for firms' shares relative to $\mathcal{N}$ investors.

is larger, j internalizes more of the pollution cost and abates more. This feedback loop between a firm's policies and its ownership is similar to the interaction between trading and voting in Levit et al. (2024). In our model, this loop gives rise to two different types of coordination motives in the portfolio choices of $\mathcal{R}$ investors: How aggressively they should trade in a given firm, and which firms they should include in their portfolios.

To understand the first type, hold fixed the mass of investors of each type demanding firm j, $\alpha_j^{\mathcal{R}}$ and $\alpha_j^{\mathcal{N}}$. An individual investor $i \in \mathcal{R}$ trades more aggressively (that is, submits a larger demand schedule s_{ij} for any given price p_j) when it expects the firm to abate more and, thus, its preference-adjusted return to be larger. If i thinks the other $\mathcal{R}$ investors will trade j more aggressively, it anticipates $s_j^{\mathcal{R}}$ will be larger and, thus, j will abate more. So, i wants to trade j more aggressively when it expects other $\mathcal{R}$ investors to do the same. We refer to this first type of coordination motive – operating through the *intensity* of trading – as the *intensive margin*. This type of coordination motive is also present in single-firm settings like Levit et al. (2024).

To understand the second type of coordination motive, suppose investors have not yet decided which firms to include in their portfolios, so they are in the process of determining $\alpha_j^{\mathcal{R}}$ and $\alpha_j^{\mathcal{N}}$, and compare two different firms, j and j'. If relatively more $\mathcal{R}$ investors demand firm j, and the overall demand for this firm is larger, but not too much larger, each type of investors prefers a different firm: $\mathcal{R}$ investors prefer j, since it has more $\mathcal{R}$ ownership and thus abates more; $\mathcal{N}$ investors prefer j', since it has less demand pressure and thus a larger liquidity discount. For the same reason, if the overall demand for j is too much larger, even the $\mathcal{R}$ investors would prefer to demand j'. So, $i \in \mathcal{R}$ is more willing to select firm j when she expects enough, but not too many, other $\mathcal{R}$ investors to do the same. We refer to this second type of coordination motive – operating through the *selection* of firms – as the *extensive margin.* This second type of coordination motive only arises in the asymmetric equilibria of a model with multiple firms, so formalizing it and studying its implications for firm ownership and policies is one of the contributions of our paper.

3.2 Symmetric equilibria

In the next step, we analyze symmetric equilibria, in which all firms have the same proportion of $\mathcal{R}$ owners $s^{\mathcal{R}}$ and, thus, have the same abatement efforts σ^* and share price p^*. There are two types of symmetric equilibria: those where only $\mathcal{N}$ investors hold stocks and those where all firms have the same positive share of $\mathcal{R}$ investors.

For a given σ, the investors' demands and, as a consequence, the degree of responsible ownership in a symmetric equilibrium are uniquely pinned down by the market-clearing conditions:

$$s^{\mathcal{R}}(\sigma) = \max\left\{\frac{\chi}{N}\frac{1}{\kappa}\left[N\kappa - (1-\chi)(1-\sigma)\lambda\right], 0\right\}. \tag{8}$$

The equilibrium value of σ follows from the first-order condition for the abatement efforts in Eqn. (3) evaluated at $\sigma_j = \sigma$ and $s_j^{\mathcal{R}} = s^{\mathcal{R}}(\sigma)$ defined in Eqn. (8) for all firms.

In principle, there can be multiple values of σ that satisfy this first-order condition and, thus, multiple symmetric equilibria σ^*. This is because of two different channels: First, independently from their ownership, firms' private incentives to abate may feature enough strategic complementarity to generate multiple equilibria. Second, the feedback loop between $s^{\mathcal{R}}$ and σ may be sufficiently strong to give rise to multiple equilibria, even when $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} \leq 0$. In this second case, $\mathcal{R}$ investors need to coordinate on how aggressively they trade, that is, on the *intensive margin*, as each $\mathcal{R}$ investor will want to trade more aggressively when it expects the others to do the same and, thus, σ to be larger. Proposition 1 characterizes the symmetric equilibria.

Proposition 1 (symmetric equilibria). *A symmetric equilibrium always exists. There exist thresholds $\widetilde{\chi}$ and χ_0, where $0 \leq \widetilde{\chi} \leq \chi_0 < 1$, for the mass χ of $\mathcal{R}$ investors such that:*

1. *For $\chi \leq \widetilde{\chi}$, all symmetric equilibria feature no SRI (i.e., $s^{\mathcal{R}} = 0$);*
2. *For $\chi \in (\widetilde{\chi}, \chi_0]$, symmetric equilibria with and without SRI may coexist;*
3. *For $\chi > \chi_0$, all symmetric equilibria feature SRI (i.e., $s^{\mathcal{R}} > 0$).*

Firm j's equilibrium abatement effort σ^* satisfies the optimality condition in Eqn. (3) for $\sigma_j = \sigma^*$ and $s_j^{\mathcal{R}} = s^{\mathcal{R}}(\sigma^*)$ defined in Eqn. (8) for all j. If the symmetric equilibrium is unique for all $\chi \in (0,1)$, it features no SRI iff $\chi \leq \chi_0$, and the equilibrium abatement effort σ^* and responsible ownership $s^{\mathcal{R}}$ are continuous and increasing in χ.

When their mass is small, $\mathcal{R}$ investors may be unable to become owners in any symmetric equilibrium. If the pollution externality is relatively large, the valuation gap between responsible and non-responsible investors is too large for $\mathcal{R}$ investors to hold positive shares when there is too many $\mathcal{N}$ types: We have $\widetilde{\chi} > 0$, so $\mathcal{R}$ investors do not hold any shares for all $\chi \in (0, \widetilde{\chi}]$. For intermediate values of χ, symmetric equilibria with and without SRI may coexist, provided that the feedback loop between σ and $s^{\mathcal{R}}$ is sufficiently strong. In this case, the equilibrium with SRI is the one where $\mathcal{R}$ investors coordinate to trade more aggressively and close the valuation gap enough to hold positive shares. Finally, as χ becomes sufficiently large, there are too many $\mathcal{R}$ investors, and equilibria without SRI are no longer possible. Otherwise, the few remaining $\mathcal{N}$ investors would have to hold excessively large stakes to ensure market clearing.

The equilibrium multiplicity that may derive from complementarity in firms' strategic decisions (Bulow et al., 1985) or investors' coordination along the intensive margin (Levit et al., 2024) has already been studied in previous work. To simplify the exposition and focus on the most novel aspects of our paper (the characterization of asymmetric equilibria, and their efficiency implications), we impose the following assumption.

Assumption 2. *We assume that* $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ *is decreasing in* σ *so that the symmetric equilibrium is unique – formally,*

$$\frac{\partial}{\partial \sigma}\left(\frac{\partial \Pi_j}{\partial \sigma_j}(\sigma, \sigma \cdot \vec{1})\right) + \frac{\lambda^2}{4N\kappa} < 0 \quad \forall \sigma \in (0,1).$$

If firms' private abatement incentives decrease steeply enough with the symmetric level of abatement σ (that is, if $\frac{\partial}{\partial \sigma}\left(\frac{\partial \Pi_j}{\partial \sigma_j}(\sigma, \sigma \cdot \vec{1})\right)$ is sufficiently negative), neither of the two channels described above is strong enough to generate equilibrium multiplicity: The function $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ crosses 0 at most once and, if so, from above. So, the first-order condition pins down a unique value of σ^*. This is the case when the complementarity in firms' abatement choices is not too strong and the abatement cost is sufficiently convex, so that large changes in σ are too costly.

3.3 Asymmetric equilibria

In addition to symmetric equilibria, the game may also admit asymmetric equilibria, where firms differ in their share of $\mathcal{R}$ investors and abatement efforts, even though they are ex-ante identical.

To build intuition, consider the case where $\mathcal{N}$ investors hold firms $j \leq n$ while $\mathcal{R}$ investors hold firms $j > n$, which implies $s^{\mathcal{R}}_{j \leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = 1$. Existence of these equilibria requires $\mathcal{R}$ investors to prefer investing in firms $j > n$ to $j \leq n$, and vice versa for $\mathcal{N}$ investors. These two incentive-compatibility conditions can be written as:

$$\begin{cases} \Pi_{j>n} - p_{j>n} - \lambda(1 - \sigma_{j>n}) & > \quad \Pi_{j\leq n} - p_{j\leq n} - \lambda(1 - \sigma_{j\leq n}) \\ \Pi_{j\leq n} - p_{j\leq n} & > \quad \Pi_{j>n} - p_{j>n}, \end{cases} \tag{9}$$

where the share prices are as described in Eqn. (7), and firms' abatement efforts solve the system of best-responses implied by their optimality conditions (Eqn. 3), evaluated at $s^{\mathcal{R}}_j = 1$ for the firms only held by $\mathcal{R}$ investors, and $s^{\mathcal{R}}_j = 0$ for those only held by $\mathcal{N}$ investors.

Expected profits Π_j cancel out in $\Pi_j - p_j$, so the way abatement impacts profits does not directly affect the financial component of returns. Firms $j > n$ have more $\mathcal{R}$ investors than the others, so they abate more; all else equal, $\mathcal{R}$ investors then prefer firms $j > n$ to $j \leq n$. $\mathcal{N}$ investors only compare financial returns, so they prefer firms $j \leq n$ if there is sufficiently more demand pressure and, thus, a smaller liquidity discount in firms $j > n$. This requires the overall mass of $\mathcal{R}$ investors, χ, to be sufficiently large, so that the overall demand for firms $j > n$ is sufficiently larger. χ cannot be too large, however, otherwise there would be too much demand pressure in firms $j > n$ and even $\mathcal{R}$ investors would prefer firms $j \leq n$. Therefore, for the two incentive-compatibility conditions in Eqn. (9) to simultaneously hold and, thus, for these types of equilibria to exist, we need χ to lie in an intermediate range.

In these equilibria, investors need to coordinate their choices of which firms to include in their portfolios, so that $\mathcal{R}$ investors only include $j > n$, and $\mathcal{N}$ investors only include $j \leq n$. This occurs through the extensive margin of coordination described before: $\mathcal{R}$ investors are willing to go after shares in high demand if they know this demand comes from like-minded investors, as they anticipate that these firms will abate more. A similar logic applies to the equilibria in which a subset of firms is held by both types of investors.[14] Proposition 2 formalizes these insights.

Proposition 2 (asymmetric equilibria). *For any pair $(\underline{n}, \overline{n})$ with $\overline{n} \geq \underline{n}$, an asymmetric equilibrium, where $\mathcal{R}$ investors hold shares in firms $j > \underline{n}$ and $\mathcal{N}$ investors hold shares in firms $j \leq \overline{n}$, exists iff the mass χ of $\mathcal{R}$ investors is in an intermediate range, $\underline{\chi} < \chi < \overline{\chi}$. In equilibrium, firm j's abatement choice is given by:*

$$\sigma_j = \begin{cases} \overline{\sigma} & \text{for } j > \overline{n}, \\ \widehat{\sigma} & \text{for } \underline{n} < j \leq \overline{n}, \\ \underline{\sigma} & \text{for } j \leq \underline{n}, \end{cases}$$

where $\overline{\sigma} = 1$ by Assumption 1, and $(\widehat{\sigma}, \underline{\sigma})$, and the share of $\mathcal{R}$ ownership in firms $\underline{n} < j \leq \overline{n}$ jointly solve firms' and investors' optimality conditions in Eqns. (3) *and* (5)*, with $1 > \widehat{\sigma} > \underline{\sigma}$.*

Among all equilibria, responsible ownership is most concentrated (in the Herfindahl-Hirschman Index sense) in the set of "best-in-class" equilibria, where $\overline{n} = \underline{n} = N - 1$.

Since firms are ex ante identical, we can describe any asymmetric equilibrium by a pair of integers $(\underline{n}, \overline{n})$ such that only $\mathcal{N}$ investors hold shares in firms $j \leq \underline{n}$, both $\mathcal{R}$ and $\mathcal{N}$ hold shares in firms $\underline{n} < j \leq \overline{n}$, and only $\mathcal{R}$ investors hold shares in firms $j > \overline{n}$, and a set of abatement efforts, $\underline{\sigma}$, $\widehat{\sigma}$, and $\overline{\sigma}$, one for each group of firms.[15] Within each group, firms have the same proportion of responsible owners and, thus, the same abatement levels. Across groups, firms with more responsible ownership abate more, as they face stronger pressure to cut pollution. Figure 2 illustrates the equilibrium characterization.

[14] In asymmetric equilibria where some firms have mixed ownership (both $\mathcal{R}$ and $\mathcal{N}$ owners), both the extensive and intensive margin of coordination are at play, with the extensive margin shaping the distribution of investors across firms, and the intensive margin how aggressively investors trade in firms with mixed ownership.

[15] The same notation can be used to describe symmetric equilibria, where the equilibrium without SRI has $\underline{n} = \overline{n} = N$, and the one with SRI has $\underline{n} = 0$ and $\overline{n} = N$. As we discussed before, equilibria where only $\mathcal{R}$ investors hold shares ($\underline{n} = \overline{n} = 0$) are not possible. Except for these three pairs, all other values of $(\underline{n}, \overline{n})$ with $0 \leq \underline{n} \leq \overline{n} \leq N$ describe an asymmetric equilibrium.

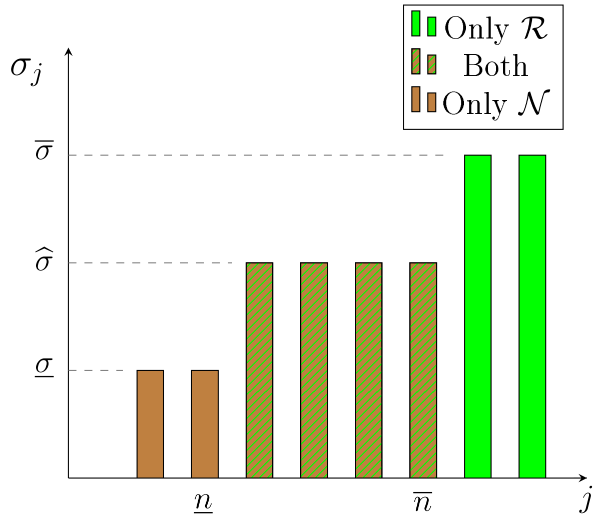


Figure 2: Illustration of asymmetric equilibria.

In any asymmetric equilibrium with $\underline{n} \geq 1$, SRI is concentrated in a subset of firms in the economy. SRI concentration is maximized in the set of *best-in-class* equilibria, where $\mathcal{R}$ investors *target* firm N and *exclude* the remaining $N-1$ firms, while $\mathcal{N}$ investors exclude firm N and invest in the other firms, which is a special case ($\overline{n} = \underline{n} = n = N-1$) of the equilibria we described at the beginning of this section.[16] $\mathcal{R}$ investors' strategy in a best-in-class equilibrium mirrors a common ESG strategy in practice, where SRI funds concentrate their holdings in the most sustainable firm within a sector ("best-in-class"). To simplify the exposition, we focus on this type of equilibria when we describe the economics of SRI concentration in the rest of the paper, but our insights extend to asymmetric equilibria more generally.

4 The economics of SRI concentration

In this section, we compare the two types of equilibria characterized in Section 3. The results we obtain here shed light on the causes and consequences of SRI concentration.

Preliminaries. As we will see shortly, the abatement efforts of firms excluded from responsible investors' portfolios in asymmetric equilibria play a central role in the economics of SRI concentration. As a preliminary to our main results, we compare the abatement levels of these firms in the set of best-in-class equilibria to those in the coexisting symmetric equilibrium, which serves as the counterfactual to SRI concentration.

[16]With substitutability or independence in abatement efforts, the best-in-class equilibrium is unique, since there is a unique value of $\underline{\sigma}$ that satisfies the excluded firms' optimality conditions in Eqn. (3). With complementarity, there might be instead multiple values of $\underline{\sigma}$ that satisfy these conditions and, thus, multiple best-in-class equilibria.

Lemma 1 (effect on excluded firms). *Compare the abatement levels of the excluded firms in any best-in-class equilibrium ($\underline{\sigma}$) to those in the coexisting symmetric equilibrium (σ^*).*

1. *If the symmetric equilibrium does not feature SRI, then $\underline{\sigma} \leq \sigma^*$ if abatement choices are strategic substitutes, $\underline{\sigma} \geq \sigma^*$ if they are strategic complements, and $\underline{\sigma} = \sigma^*$ if they are independent.*

2. *If the symmetric equilibrium features SRI, then $\underline{\sigma} \leq \sigma^*$ under strategic substitutability or independence, while $\underline{\sigma} \gtreqless \sigma^*$ under strategic complementarity.*

Except for the last one, the inequalities above are all strict if $\sigma^* > 0$.

When the symmetric counterfactual does not feature SRI, excluded firms are fully owned by $\mathcal{N}$ investors in both equilibria, so any change in their abatement across the two equilibria is driven by the interdependence in their private incentives, $\frac{\partial \pi_j}{\partial \sigma_j}$ (hence, if abatement decisions are independent, their behavior remains unchanged). If abatement efforts are strategic substitutes, these firms reduce their abatement in response to the higher abatement efforts of the best-in-class firm; conversely, if efforts are strategic complements, the excluded firms abate more in any best-in-class equilibrium. If the symmetric equilibrium also features SRI, all firms face pressure from shareholders to reduce emissions. If this pressure is sufficiently large, excluded firms abate more than in the symmetric equilibrium even under strategic complementarity.

4.1 Why does concentration arise?

Our next step is to understand when and why $\mathcal{R}$ investors may adopt investment strategies that lead to concentration of SRI.

Proposition 3 (preference for concentration). *If two equilibria coexist, each group of investors prefers the one where they hold larger shares, and $\mathcal{R}$ and $\mathcal{N}$ investors never strictly prefer the same equilibrium.*

1. *For $\chi \in (0, \chi_0]$, $\mathcal{R}$ investors do not trade in the symmetric equilibrium; thus, they always prefer a best-in-class equilibrium if it exists. If the pollution externality λ is large ($\lambda \geq \overline{\lambda}$), a best-in-class equilibrium exists for $\chi \in (\frac{1}{N}, \min\{\chi_0, \overline{\chi}\}]$.*

2. *For $\chi \in (\chi_0, 1)$, $\mathcal{R}$ investors hold positive shares in the symmetric equilibrium. With independence or substitutability in abatement efforts, a unique best-in-class equilibrium exists, and $\mathcal{R}$ investors prefer it to the symmetric equilibrium, for $\chi \in [\max\{\frac{1}{N}, \chi_0\}, \chi'')$, with $\chi_0 < \chi'' < \overline{\chi} < 1$.*

Proposition 3 yields our first main implication: When the societal cost of pollution is large, $\mathcal{R}$ investors have a strict preference for concentration whenever their mass is not too large. Intuitively, concentration allows the $\mathcal{R}$ investors to create enough price pressure on the firms they target (e.g., the best-in-class firm), so that the $\mathcal{N}$ types (who have relatively higher valuations for polluting firms) go after other firms, and $\mathcal{R}$ investors can have a significant impact on those they hold.

Below, we describe the specific logic behind these results. An investor's payoff increases with the number of shares it holds in equilibrium, and investors of the same type always get the same payoff in equilibrium. This implies that all $\mathcal{R}$ investors always prefer the same equilibrium among different equilibria, and they prefer a best-in-class to the symmetric equilibrium if they hold larger stakes in the former. In the symmetric equilibrium, $\mathcal{R}$ investors become owners only if χ is above a threshold χ_0. For $\chi \leq \chi_0$, $\mathcal{R}$ investors are then strictly better off in a best-in-class equilibrium whenever it exists. If the pollution externality is large, the valuation gap between $\mathcal{R}$ and $\mathcal{N}$ investors is substantial, so it is hard for $\mathcal{R}$ investors to become owners unless they concentrate and turn the best-in-class firm green. In this case, χ_0 is larger than $\frac{1}{N}$, which is the lower bound for the existence of any best-in-class equilibrium, so any such equilibrium exists and it is thus strictly preferred by $\mathcal{R}$ investors. Indeed, $\underline{\chi} = \frac{1}{N}$ because there is a mass χ of investors in the best-in-class firm, and a mass $\frac{1-\chi}{N-1}$ in the excluded firms; so we must have $\chi > \frac{1}{N}$ to have overall more investors and, thus, more demand pressure in the best-in-class firm. This is because the best-in-class firm ($j = N$) fully abates its emission, so there is no pollution discount in p_N and, thus, $\chi > \frac{1}{N}$ is enough to make $\mathcal{N}$ investors prefer the excluded firms ($j < N$).

The existence threshold $\overline{\chi}$ is instead such that, for all $\chi < \overline{\chi}$, an individual $\mathcal{R}$ investor i prefers to hold the best-in-class firm ($j = N$) to any of the excluded ones ($j \leq N-1$); χ_0 is instead such that, for all $\chi \leq \chi_0$, i prefers not to invest compared to holding any firm in the symmetric equilibrium. The excluded firms have more demand pressure and, under strategic substitutability or independence, abate less in the best-in-class equilibrium than in the symmetric without SRI.[17] Investing in the excluded firms must then be relatively less attractive in the best-in-class equilibrium. It follows that there must be a non-empty interval for χ where i is willing to invest in $j \leq N$ in the symmetric equilibrium, but still prefers firm $j = N$ in a best-in-class equilibrium.

At least under substitutability or independence, the two equilibria must then coexist also for some $\chi > \chi_0$, provided that χ is not too large ($\chi < \overline{\chi}$). For χ close to χ_0, the valuation gap with $\mathcal{N}$ investors is still large, so $\mathcal{R}$ investors hold smaller stakes in the symmetric equilibrium; for χ close to $\overline{\chi}$, the best-in-class firm is too crowded, so $\mathcal{R}$

[17] There is a mass $\frac{1-\chi}{N-1}$ of $\mathcal{N}$ investors in each excluded firm in the best-in-class equilibrium, and a mass $\frac{1-\chi}{N}$ in any firm in the symmetric equilibrium without SRI. For the comparison of the excluded firms' abatement efforts across equilibria, see Lemma 1.

investors hold larger stakes in the symmetric equilibrium.[18] It follows that, even when χ is large enough for $\mathcal{R}$ investors to participate in both types of equilibria, they still prefer the best-in-class equilibrium unless χ is sufficiently large.

We conclude this section with a brief discussion of equilibrium selection. Proposition 3 shows that, when two equilibria coexist, $\mathcal{R}$ and $\mathcal{N}$ investors always disagree on which equilibrium they prefer. The reason is simple. The total number of shares on the market is fixed. If two equilibria A and B coexist, for investor i to prefer A over B, then some other investor i' must hold fewer shares in equilibrium A than in equilibrium B and, thus, prefer B to A. Since investors of the same type always get the same payoff in equilibrium, i' must be of a different type than i. Whenever $\mathcal{R}$ investors strictly prefer one of two different equilibria, $\mathcal{N}$ investors then strictly prefer the other. So, even if $\mathcal{R}$ investors prefer to concentrate, it is unclear which equilibrium investors will coordinate on, since neither equilibrium is payoff-dominant. In Appendix B.2, we show that equilibria with concentration are relatively more "stable," in the traditional sense of being more robust to small perturbations to the equilibrium strategies. We describe how this points to a simple mechanism for equilibrium selection in favor of equilibria with concentration (e.g., some SRI funds making commitments to best-in-class investing).

4.2 When does concentration reduce pollution?

Having established that $\mathcal{R}$ investors prefer to concentrate when a sufficient fraction of investors are non-responsible, our next step is to understand how SRI concentration affects aggregate pollution. As before, the symmetric equilibrium plays the role of counterfactual to SRI concentration.

Proposition 4 (effect on aggregate abatement). *Compare the aggregate abatement ($\sum_j \sigma_j$) in any best-in-class equilibrium and in the coexisting symmetric equilibrium.*

1. *If the symmetric equilibrium does not feature SRI, $\sum_j \sigma_j$ can be smaller in the best-in-class equilibrium only if abatement efforts are strategic substitutes. A sufficient condition for this result to hold is $\Gamma(\frac{1}{N}, \frac{1}{N} \cdot \vec{1}; 0) > 0 \geq \Gamma(0, (1, \vec{0}); 0)$.*

2. *If the symmetric equilibrium features SRI, $\sum_j \sigma_j$ can be smaller in a best-in-class equilibrium in all three cases of strategic interaction. If there is independence in abatement efforts and $\Gamma(0, \vec{0}; 0) \geq 0$, $\sum_j \sigma_j$ is always smaller in the best-in-class equilibrium if $\chi \in (\chi', \overline{\chi})$, with $\chi_0 < \chi' < \overline{\chi}$.*

[18] $s_j^{\mathcal{R}} \to 0$ as χ approaches χ_0. As χ approaches instead $\overline{\chi}$, $\mathcal{R}$ investors become indifferent between investing in the best-in-class or the excluded firms. Since, as we described above, these firms are less attractive in the best-in-class equilibrium, $\mathcal{R}$ investors must hold strictly larger stakes in (and, thus, prefer) the symmetric equilibrium.

Proposition 4 characterizes the two mechanisms through which SRI concentration can reduce aggregate abatement $\sum_j \sigma_j$, in which case SRI concentration is inefficient from a social welfare perspective. Since the best-in-class firm always abates (at least weakly) more than in the symmetric equilibrium, both mechanisms must operate through the indirect effects of concentration on the abatement efforts of the excluded firms, $\sigma_{j\leq N-1}$.

If the symmetric equilibrium does not feature SRI, $\sum_j \sigma_j$ can be smaller in a best-in-class equilibrium only if $\sigma_{j\leq N-1}$ is smaller than in a world where all shareholders and, thus, all firms are purely profit-motivated. This can be true only if abatement efforts are strategic substitutes, as the increased abatement efforts of the best-in-class firm must crowd out those of the excluded. In this case, the inefficiency of concentration is to "tilt the playing field" too much in favor of the best-in-class firm, which causes the excluded firms to drop their abatement efforts compared to a counterfactual where there is no SRI, the playing field is level, and all firms pursue such efforts.

The sufficient condition at the end of Part 1 of Proposition 4 helps us think about economic environments where this inefficiency materializes. Formally, the condition implies that (i) all firms abate more than $\frac{1}{N}$ in the symmetric equilibrium ($\sigma^* > \frac{1}{N}$ and, thus, $\sum_j \sigma_j > 1$) and (ii) the excluded firms give entirely up on abating in a best-in-class equilibrium ($\overline{\sigma} = 1, \underline{\sigma} = 0$, so $\sum_j \sigma_j = 1$). An environment where this is more likely to hold is one where firms compete for consumers whose demand is very sensitive to differences in pollution across firms, so that the excluded firms give up on abating in the best-in-class equilibrium, where they lag behind the best-in-class, but they abate a lot in the symmetric equilibrium, where all firms compete neck-and-neck. This is similar to how competition affects innovation in Aghion et al. (2005).

When the symmetric equilibrium also features SRI, $\sum_j \sigma_j$ is smaller in a best-in-class equilibrium when $\sigma_{j\leq N-1}$ would be much larger if $\mathcal{R}$ investors were to pressure also the excluded firms, like they do in the symmetric equilibrium. In this case, the inefficiency of concentration consists in creating a few "green islands" in a "sea of brown firms," rather than driving a more generalized change toward a greener economy. This type of inefficiency can arise in all three cases of strategic interaction in abatement efforts, including the benchmark without strategic interaction.

The benchmark with independence in abatement efforts is more tractable, so we use it to derive further results. In Part 2 of Proposition 4, we show that, when $\sigma^* > 0$, aggregate abatement is always smaller in a best-in-class equilibrium for values of χ near the upper end of its existence range, $\overline{\chi}$. In a best-in-class equilibrium, $\mathcal{R}$ investor i holds a mass $\frac{1}{\chi}$ of shares of the best-in-class firm, and $\overline{\chi}$ is the smallest value of χ such that i would hold a larger position if it were to deviate and invest in one of the excluded firms. For χ close to $\overline{\chi}$, i must then hold close to $\frac{1}{\chi}$ shares in the symmetric equilibrium, where firms abate more than the excluded firms do in a best-in-class equilibrium, and there is fewer $\mathcal{N}$ investors in each firm. Collectively, $\mathcal{R}$ investors then hold close to $\frac{\chi}{N} \cdot \frac{1}{\chi} = \frac{1}{N}$

shares in each firm in the symmetric equilibrium which, by the concavity of the firms' abatement choice problem, implies $\sigma^* > \frac{1}{N} + \frac{N-1}{N}\underline{\sigma}$ and, thus, $N\sigma^* > 1 + (N-1)\underline{\sigma}$.

To wrap up, the results in Proposition 4 yield our second main insight: Concentration may be necessary for $\mathcal{R}$ investors to participate and, unless there is too much strategic substitutability in abatement choices, socially desirable when their mass χ is small. Even when this is the case, however, concentration persists even as χ becomes larger, and it is no longer neither needed for $\mathcal{R}$ investors to participate nor socially desirable.

4.3 Externalities of concentration

The last step of our analysis is to study whether $\mathcal{R}$ investors' preference for concentration aligns with the broader social objective of reducing aggregate pollution. As a benchmark, we first note that this alignment always holds when we restrict attention to symmetric equilibria. Indeed, if Assumption 2 does not hold, and multiple symmetric equilibria coexist, among these $\mathcal{R}$ investors always prefer the equilibrium with the highest aggregate abatement. This is because, in that equilibrium, firms pollute less, so the valuation gap with $\mathcal{N}$ investors is smaller and, as a result, $\mathcal{R}$ investors hold larger stakes.

Proposition 5 shows that this alignment can break down in asymmetric equilibria.

Proposition 5 (private vs. social value of concentration). *The following results hold:*

1. *$\mathcal{R}$ investors may prefer a best-in-class equilibrium, even when aggregate abatement ($\sum_j \sigma_j$) is lower than in the coexisting symmetric equilibrium. If there is independence in abatement efforts and $\Gamma(0, \vec{0}; 0) \geq 0$, this holds for all $\chi \in (\chi', \chi'')$, with $\chi_0 < \chi' \leq \chi'' < \overline{\chi}$, with strict inequality provided that $\sigma^* < 1$ at $\chi = \frac{1}{N}$.*
2. *When the excluded firms' abatement levels ($\underline{\sigma}$) in a best-in-class equilibrium are higher, aggregate abatement ($\sum_j \sigma_j = 1 + (N-1)\underline{\sigma}$) increases, but the equilibrium exists over a smaller interval of χ. In the limit as excluded firms approach full abatement ($\underline{\sigma} \to 1$), this equilibrium ceases to exist.*

There are two ways in which $\mathcal{R}$ investors' preference for concentration diverges from the social desirability of concentration. First, $\mathcal{R}$ investors may prefer a best-in-class equilibrium even when this leads to lower aggregate abatement than the coexisting symmetric equilibrium. With independence in abatement choices, we are able to show this always happens for some χ within the existence interval of the best-in-class equilibrium. Intuitively, $\mathcal{R}$ investor i is indifferent between the two equilibria at $\chi = \chi''$, where i holds $\frac{1}{\chi}$ shares in both, and prefers the best-in-class equilibrium for $\chi < \chi''$. In Part 2 of Proposition 4, we have shown that, at $\chi = \chi''$, $\sum_j \sigma_j$ is already larger in the symmetric equilibrium. So, there must be some values of χ smaller than χ'' such that $\sum_j \sigma_j$ is larger in the symmetric equilibrium, but i still prefers the best-in-class one.[19] Figure 3 illustrates these results in a numerical example of the model.

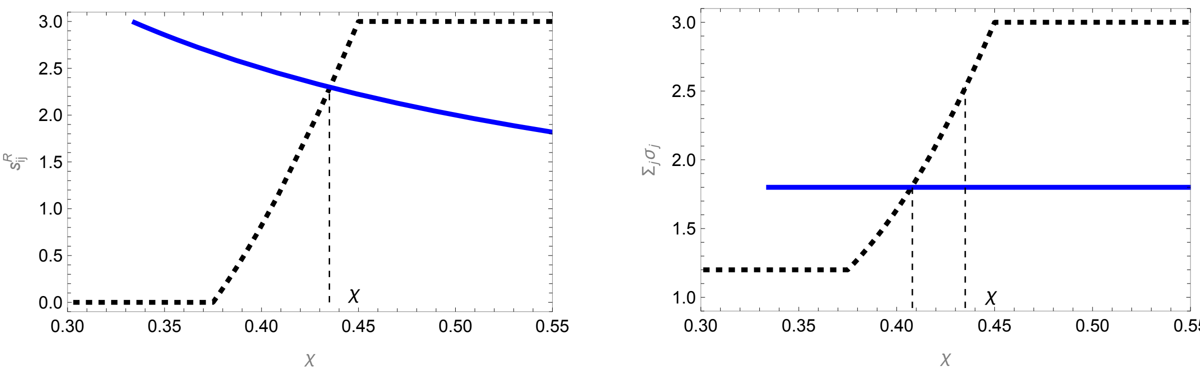


Figure 3: This left (right) panel plots the position of a responsible investor in a given firm $s_{ij}^{\mathcal{R}}$ (aggregate abatement $\sum_j \sigma_j$) as a function of the mass of responsible investors χ. The dotted black line corresponds to the symmetric equilibrium, and the solid blue line to the best-in-class equilibrium. Functional forms and parameters: $\pi_j = \sigma_j \Delta$, $\mathcal{C}(\sigma_j) = b\sigma_j^2$, $N = 3$, $\lambda = 2$, $\kappa = 0.25$, $b = 0.75$, and $\Delta = 0.6$.

Second, from a social welfare perspective, concentration is more desirable when the increased abatement of the best-in-class firm has a larger positive effect on the abatement efforts of the excluded firms. That is, when $\underline{\sigma}$ and, thus, $\sum_j \sigma_j$ are larger. One would then want a best-in-class equilibrium to be more likely in environments where $\underline{\sigma}$ is larger. We show that the exact opposite happens: a best-in-class equilibrium exists over a *smaller* interval of χ when $\underline{\sigma}$ is larger, and it disappears entirely in the limit as excluded firms approach full abatement ($\underline{\sigma} \to 1$). Intuitively, when excluded firms abate more, they become more attractive to $\mathcal{R}$ investors, so investing only in the best-in-class firm becomes harder to sustain in equilibrium.

Corollary 1 shows that the same inverse relationship can arise between the existence of concentration and its relative impact on aggregate abatement (relative to the symmetric counterfactual).

Corollary 1 (existence vs impact). *Consider any pair of profit functions π' and π'' such that (i) $\Gamma(0, \vec{0}; 0)$ is negative only under π', and (ii) $\Gamma(\sigma, (1, \sigma \cdot \vec{1}); 0)$ is weakly greater under π' for all $\sigma \in [0, 1]$. Then there always exists a best-in-class equilibrium such that the following results hold:*

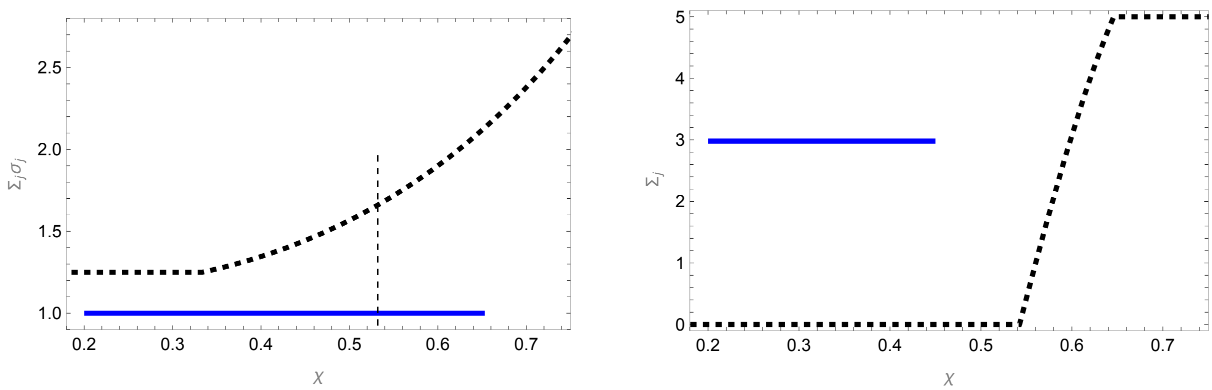
1. *The symmetric equilibrium without SRI features higher abatement under π'', but excluded firms in the best-in-class equilibrium abate more under π'.*

[19]The set $\chi \in (\chi', \chi'')$ is always non-empty when $\lambda \geq \overline{\lambda}$, so that the best-in-class equilibrium and the symmetric equilibrium with SRI coexist for $\chi \in (\chi_0, \overline{\chi})$. If $\lambda < \overline{\lambda}$, the interval is still non-empty provided that $\mathcal{R}$ investors strictly prefer the best-in-class equilibrium at $\chi = 1/N$, which is the case if $\sigma^* < 1$, so that the best-in-class firm still abates more than in the symmetric equilibrium.

2. *The difference in aggregate abatement between the best-in-class equilibrium and the symmetric equilibrium without SRI is larger under π', but the existence interval for the best-in-class equilibrium is larger under π''.*

Condition (i) implies that, in the symmetric equilibrium without SRI, firms choose $\sigma^* = 0$ under π', but $\sigma^* > 0$ under π''. Condition (ii) implies that this ranking flips in the best-in-class equilibrium: excluded firms abate more under π' than under π''. The interpretation we have in mind is that, under π', $\mathcal{R}$ investors can solve a coordination problem in the adoption of green technologies: if they turn the best-in-class firm green, the excluded firms will also abate a lot, due to complementarity in abatement choices, but all firms do not abate if they are left on their own. By contrast, under π'', firms have strong abatement incentives in the symmetric equilibrium, but the excluded firms give up on abating in the best-in-class one, due to substitutability in abatement choices.

Since σ^* is smaller and $\underline{\sigma}$ is larger under π', one would want SRI concentration to be more likely under π', where it has a more positive effect on $\sum_j \sigma_j$ compared to the symmetric counterfactual. However, the opposite occurs: because a higher $\underline{\sigma}$ makes it harder to sustain concentration, the best-in-class equilibrium exists over a smaller range of parameters under π'. Figure 4 illustrates the results in Corollary 1 in a numerical example of the model.



(a) Strategic substitutability.

(b) Strategic complementarity.

Figure 4: This figure plots aggregate abatement $\sum_j \sigma_j$ as a function of the mass of responsible investors χ. The dotted black line corresponds to the symmetric equilibrium, and the solid blue line to the best-in-class equilibrium. For both panels, $\mathcal{C}(\sigma_j) = a\sigma_j + b\sigma_j^2$. Panel (a) features strategic substitutability (π' in Corollary 1) based on the product market model in Section 5.3 with a mass of responsible consumers equal to 0.3 and $u(x) = \alpha x + 0.5\beta x^2$ with $\alpha = 3$ and $\beta = 0.5$; Panel (b) features strategic complementarity (π'' in Corollary 1) with $\pi_j = \sigma_j \Delta + \gamma \sigma_j \left(\mathcal{S}(\vec{\sigma}_{-j}) - 0.5\sigma_j\right)$ where $\mathcal{S}$ is the softmax-weighted average $\frac{\sum_{j' \neq j} \sigma_{j'} \exp(\epsilon^{-1}\sigma_{j'})}{\sum_{j' \neq j} \exp(\epsilon^{-1}\sigma_{j'})}$ with $\epsilon = 0.065$. Parameters: $N = 5, \lambda = 1, \kappa = 0.1, a = 0.39, b = 0.3, \Delta = 0.344$, and $\gamma = 0.68$.

Proposition 5 yields the third main insight of the paper, which is to highlight an inherent tension in the role of SRI as a mechanism to control externalities: When SRI is not too popular, $\mathcal{R}$ investors will favor concentration in equilibrium, but when they do concentrate, their preferences become detached from aggregate abatement. The wedge between $\mathcal{R}$ investors' preference for concentration and its social desirability manifests in two different ways. First, $\mathcal{R}$ investors favor concentration for "too long," even as SRI becomes more popular and less concentrated strategies that would increase aggregate abatement are also possible in equilibrium. Second, concentration is more likely to arise exactly when it is the least desirable, that is, when the abatement efforts of the best-in-class firm *crowd out* those of the excluded firms.

Why does this wedge only arise when $\mathcal{R}$ investors concentrate? In equilibria without concentration, $\mathcal{R}$ investors are equally represented across firms and, as a consequence, firms have the same abatement levels. Individually, $\mathcal{R}$ investors still only care about the pollution generated by the firms in their portfolios but, since all firms choose the same abatement levels, this is a sufficient statistics for aggregate abatement. So, the preferences of $\mathcal{R}$ investors become more aligned with aggregate abatement as they become more uniformly distributed across firms.

5 Downstream effects: Shaping consumption norms

Most climate scientists believe that limiting the risk of a natural disaster requires a green transition – a significant transformation of production and consumption norms. Like portfolio choices, consumption norms reflect both preferences and prices (Besley and Persson, 2023). Another dimension of the overall impact of SRI is thus the way it affects the supply and price of clean products. To investigate these effects, we add to our model a unit mass of households, indexed by $h \in [0, 1]$, to study how SRI, and its concentration, affect consumption.

5.1 Product market model

We assume some households prefer to buy from firms that pollute less, so that, by affecting the distribution of firms' abatement efforts, $\mathcal{R}$ investors indirectly impact households' consumption decisions. The presence of responsible households also implies that, independently of SRI, abatement increases the demand for a firm's products and thereby boosts its revenues, providing a microfoundation for the private abatement incentives assumed in the previous sections.

We use the notation $\theta_j = c$ ($\theta_j = d$) to signify that j uses a clean (dirty) production technology. The dirty technology generates a negative pollution externality λ, while the clean one does not. The random vector $\vec{\theta} \equiv (\theta_1, \ldots, \theta_N)'$ describes each firm's technology. We define σ_j as the probability that firm j uses the clean technology, $\sigma_j \equiv \Pr(\theta_j = c)$. For a given realization of $\vec{\theta}$, the economy is populated by n_c clean and $n_d \equiv N - n_c$

dirty firms, each supplying one product unit, with products being potentially different across firms. The equilibrium prices $\vec{\rho} \equiv (\rho_1, \ldots, \rho_N)'$ equate demand and supply for all products. To simplify the exposition, we assume zero costs of production, so firm j's profit is the market-clearing price for its product net of the abatement cost.

Households have heterogeneous preferences. Similar to investors, a fraction $\chi \in (0,1)$ of households incur a disutility λ from consuming dirty products (those produced using the dirty technology), so we refer to them as responsible consumers ($h \in \mathcal{R}$). The remaining households ($h \in \mathcal{N}$) are, all else equal, indifferent between clean and dirty products. Given firms' technologies $\vec{\theta}$ and product prices $\vec{\rho}$, household h chooses a consumption bundle $\vec{x}_h \equiv (x_{h1}, \ldots, x_{hN})'$ to maximize its utility from consumption $U(\vec{x}_h) : (\mathbb{R}_+)^N \to \mathbb{R}_+$, net of the expenditure and, if h is responsible, the disutility for the consumption of dirty products:[20]

$$\max_{\vec{x}_h \geq \vec{0}} \; U(\vec{x}_h) - \sum_{j \in \mathcal{J}} x_{hj} \left(\rho_j + \lambda \mathbb{1}_{h \in \mathcal{R}} \mathbb{1}_{\theta_j = d} \right). \tag{10}$$

For a given realization of $\vec{\theta}$, the product market equilibrium does not depend on investors' portfolio choices $\{s_{ij}\}$ and firms' abatement efforts $\vec{\sigma}$. It follows that the equilibrium characterization in Section 3 carries through, with the following equilibrium specification for firm j's expected gross profits:

$$\pi_j = \sigma_j \mathbb{E}[\rho_j | \theta_j = c] + (1 - \sigma_j) \mathbb{E}[\rho_j | \theta_j = d], \tag{11}$$

where $\mathbb{E}[\rho_j | \theta_j = \tilde{\theta}]$ denotes the expected market-clearing price ρ_j given j's technology $\tilde{\theta} \in \{c, d\}$.

The expectation in Eqn. (11) is taken over the technology realizations of the other $N - 1$ firms, given their abatement efforts $\vec{\sigma}_{-j}$. All else equal, a clean product will always sell at a (at least weakly) larger price than a dirty one in equilibrium, since $\mathcal{R}$ households suffer a disutility from consuming dirty products, so they have a larger demand for clean products. The marginal revenue from j's abatement efforts $(\frac{\partial \pi_j}{\partial \sigma_j})$ is then the expected price premium for selling a clean product, $\mathbb{E}[\rho_j | \theta_j = c] - \mathbb{E}[\rho_j | \theta_j = d]$. For some specifications of $U(\cdot)$, this premium also depends on other firms' abatement choices, as $\mathcal{R}$ households' willingness to pay for j's clean product depends on the availability of other clean products in the economy.

Since firms' production costs (gross of abatement) are zero, the expected surplus generated in the product market is equal to the aggregate utility from consumption

$$PS = \int_0^1 \mathbb{E}\left[U(\vec{x}_h) \right] dh,$$

[20]Program (10) implicitly assumes that consumers' budget constraints are not binding at the equilibrium consumption, so their choice is equivalent to an unconstrained problem. This assumption simplifies the exposition but does not affect our results.

where the expectation is taken with respect to the random vector $\vec{\theta}$ and households' demand is evaluated at its equilibrium value.[21]

Households and investors are both atomistic, so the characterization of equilibria (and aggregate welfare) does not depend on whether some households are both consumers and investors.[22] Their preferences over equilibria, however, do: In some cases, SRI concentration hurts consumers and, thus, households that are both responsible investors and consumers may have less of a preference for asymmetric equilibria. To simplify the exposition, here we take SRI concentration as given, and focus on its effects on product market surplus. We consider three specifications for $U(\cdot)$, each yielding a different type of interaction in abatement efforts (independence, substitutability, and complementarity), and study the implications of SRI concentration on consumption in each setting. In what follows, $u(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuously differentiable, increasing, and concave function.

Since the utility from consumption is concave, in all three specifications for $U(\cdot)$, PS is maximized when all households have the same consumption levels. This is the case only if there is sufficiently many clean products, as $\mathcal{R}$ households suffer a disutility from consuming dirty products and, thus, have lower demand for these products.

5.2 Benchmark with independent products

First, we consider a benchmark in which consumers view products as independent, that is, the demand for each product does not depend on the characteristics or prices of the others:

$$U(\vec{x}_h) = \sum_{j \in \mathcal{J}} u(x_{hj}). \tag{12}$$

In this setting, each firm can be interpreted as a representative firm of a distinct industry.

Proposition 6 (independence). *Suppose firms sell independent products – i.e., $U(\cdot)$ is defined in Eqn. (12):*

[21] Following Dewatripont and Tirole (2024), PS excludes the disutility incurred by $\mathcal{R}$ households when consuming dirty products. Similarly, excluding $\mathcal{R}$ investors' disutility from purchasing shares in polluting firms, total surplus is defined as

$$S = PS - \sum_{j \in \mathcal{J}} [\lambda(1 - \sigma_j) + \mathcal{C}(\sigma_j)] - \tfrac{\kappa}{2} \int_0^1 (\iota' \vec{s}_i)^2 \, di,$$

where the second component is the sum of pollution and abatement costs, and the last is the aggregate trading cost. S is largest when all firms fully abate their emissions, so that PS is the largest, firms abate to the socially optimal level, and trading costs are the smallest (since the trading cost investor i incurs from holding s_{ij^*} shares is convex in s_{ij^*}, aggregate trading costs are minimized when s_{ij^*} is the same across investors, which requires $\sigma_j = 1$ for all j, so that all investors have the same valuations). We think of κ as being small compared to λ, so we focus on the first two components of S in our analysis.

[22] Both take equilibrium outcomes as given, so, under the assumption that budget constraints are not binding, investment and consumption decisions are two separate optimization problems even for households that are both investors and consumers.

1. *Firm j's product price ρ_j only depends on its production technology $\theta_j \in \{c, d\}$. Therefore, firms' abatement efforts are independent.*

2. *Expected product market surplus PS is an increasing function of $\sum_j \sigma_j$. Hence, compared to the coexisting symmetric equilibrium, PS is smaller in the best-in-class equilibrium if and only if $\sum_j \sigma_j$ is also smaller.*

If firms are active in different markets, each firm's price premium from selling a clean product does not depend on other firms' production technologies. As a result, firms' abatement efforts are independent. As the expected number of clean firms is equal to aggregate abatement, and PS strictly increases with the number of clean firms, SRI concentration can only reduce expected-product market surplus if it reduces aggregate abatement.

5.3 From SRI concentration to product market power

Next, we analyze a setting where firms compete within the same industry, so their products are substitutes. Formally, we specify household h's utility from consumption as:

$$U(\vec{x}_h) = u\left(\sum_{j \in \mathcal{J}} x_{hj}\right). \tag{13}$$

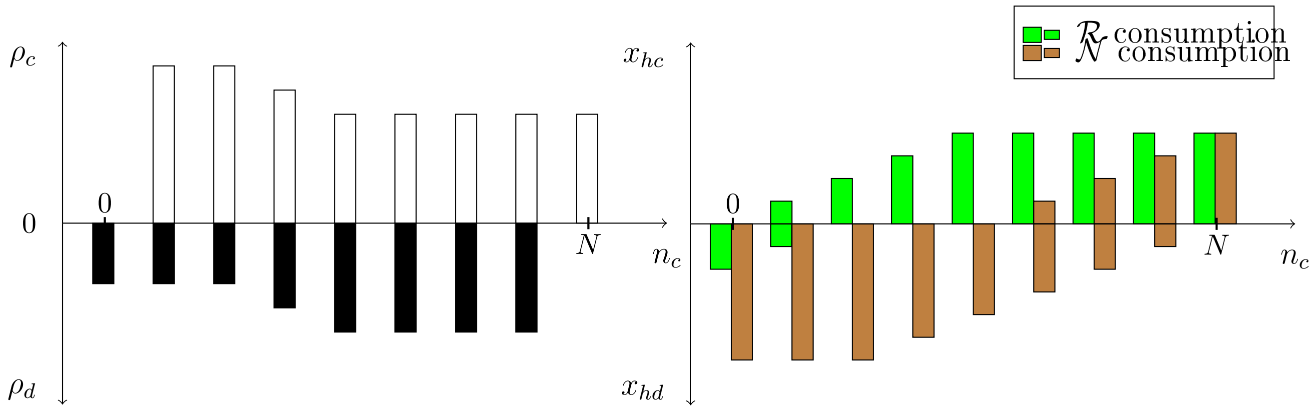
Conditional on the technology, products here are perfect substitutes, as h's utility depends only on its aggregate consumption. There are then only two types of products: clean and dirty, with the price of each type determined by its aggregate supply.

Proposition 7 (substitutability). *Suppose firms sell (perfect) substitutes – i.e., $U(\cdot)$ is defined in Eqn. (13):*

1. *The price premium for a clean product decreases with the number of clean rivals. Therefore, firms' abatement efforts are strategic substitutes.*

2. *Expected product market surplus PS depends on $\vec{\sigma}$. Compared to the coexisting symmetric equilibrium, PS can be smaller in the best-in-class equilibrium, even when $\sum_j \sigma_j$ is higher.*

The equilibrium depends on the number of clean firms n_c. (1) If n_c is small, the supply of clean products is too little, so, if there is any clean product, it sells at a price premium λ. $\mathcal{R}$ consumers are then indifferent between paying the price premium or incurring the disutility λ, so their consumption is inefficiently low relative to $\mathcal{N}$ consumers. (2) If n_c takes intermediate values, there is enough clean firms for $\mathcal{R}$ consumers to buy only clean products, at a price premium lower than λ and decreasing with n_c. $\mathcal{R}$ households though still have lower consumption than $\mathcal{N}$ consumers, as they still pay a price premium, though this inefficiency here is less pronounced. (3) If n_c is sufficiently large, there is more clean products than $\mathcal{R}$ households demand, so there is no longer a price premium for clean products and also some $\mathcal{N}$ households consume these products. All households' consumption is equalized, so product market surplus reaches its maximum level. The equilibrium characterization is illustrated in Figure 5.



(a) Product prices. (b) Consumption choices.

Figure 5: Illustration of product market equilibrium with perfect substitutes goods (see Eqn. (13)). In both panels, n_c is the number of clean firms, with n_c taking values in $\{0, 1, ..., N\}$. In Panel (a), ρ_c and ρ_d describe the equilibrium prices of clean and dirty products. In panel (b), x_{hc} and x_{hd} describe household h's consumption choices of clean and dirty products.

The equilibrium characterization has two important implications. First, the relative profitability (i.e., the price premium) of selling clean products decreases with the number of clean rivals. For any firm j, an increase in a rival's abatement effort shifts the distribution of the number of clean rivals in a first-order stochastic dominance sense, which reduces the expected price premium. This shift reduces the expected extra-profit j makes from adopting the clean technology, which discourages its abatement efforts: When products are substitutes, abatement efforts are thus strategic substitutes.[23] In this setting, SRI concentration then generates a crowding-out of the excluded firms' abatement efforts, as shown in the numerical example in Figure 4, Panel (a).

Second, product market surplus strictly increases with the number of clean firms n_c (up to its maximum level) only if n_c is already sufficiently large. When only a few firms supply clean products, they fully capture the extra utility λ they generate for $\mathcal{R}$ households through the price premium; so the net surplus and consumption level of $\mathcal{R}$ households is the same as if only dirty products were available. Through this channel, even when concentrated SRI increases aggregate abatement relative to the symmetric

[23]Provided clean products are preferred by $\mathcal{R}$ consumers, the condition that the expected extra profit from being clean decreases with the number of clean rivals and, thus, that abatement efforts are strategic substitutes is satisfied in other standard oligopoly models (e.g., Hotelling/Salop models, imperfect Bertrand competition).

equilibrium, it may still reduce expected responsible consumption and product market surplus: The crowding out of abatement efforts by the excluded firms can still reduce the likelihood that n_c is large enough, leading to product-market power for clean firms and higher prices for clean products.

5.4 From SRI concentration to green catalysts

Finally, we consider a setting where firms produce complementary products:

$$U(\vec{x}) = u\left(\min_{j\in\mathcal{J}} x_{hj}\right). \tag{14}$$

Products here are perfect complements, as households derive utility from consuming a bundle of goods in which each firm supplies a necessary component.[24] Since demand only depends on the total cost of the bundle, there is a continuum of market-clearing prices. We look at the unique symmetric equilibrium of the product market, where all products sell at the same price.[25]

Proposition 8 (complementarity). *Suppose firms sell (perfect) complements – i.e., $U(\cdot)$ is defined in Eqn. (14) – and $\mathcal{R}$ consumers never consume dirty products ($\lambda > u'(0)$):*

1. *The price premium for a clean product increases with the number of clean rivals, implying that firms' abatement efforts are strategic complements.*
2. *Expected product market surplus PS depends on $\prod_j \sigma_j$. Compared to the coexisting symmetric equilibrium, PS is always larger in a best-in-class equilibrium if the symmetric one does not feature SRI; if it does, PS can be smaller in a best-in-class equilibrium even when it increases $\sum_j \sigma_j$.*

Since households always purchase the N goods together, $\mathcal{R}$ types here may be willing to buy only if all products are clean, in which case the price for all products is higher due to the increased demand. As a result, firm j's marginal benefit from abatement is higher when the probabilities $\sigma_{j'}$ that the other firms $j' \neq j$ develop the clean technology are larger. Complementarity in consumption thus induces strategic complementarity in firms' abatement efforts.

In this environment, the best-in-class firm acts as a *green catalyst*, pushing up the investments of the excluded firms relative to a coexisting symmetric equilibrium without SRI. However, when the symmetric equilibrium also features SRI, concentration can still

[24] Alternatively, this model applies to a vertical supply chain, where firm j supplies an essential input used by firm $j+1$.

[25] Restricting attention to a symmetric equilibrium price is natural in this setting: $\mathcal{N}$ consumers are indifferent between clean and dirty products, while $\mathcal{R}$ consumers purchase only when all products are clean.

reduce expected product market surplus even when it increases aggregate abatement. This happens when the differentiation it induces in firms' abatement levels lowers the likelihood that *all* firms develop the clean technology, thereby reducing the chances that $\mathcal{R}$ consumers can buy a bundle of clean products.

6 Conclusion

The practice of incorporating environmental and social factors into investment decisions and ownership policies is becoming increasingly popular. This paper has developed a framework to explore how this socially responsible investing (SRI) affects firms' concerns with Corporate Social Responsibility (CSR) issues and analyzed the implications for firm externalities and product markets. We have shown that the typical SRI strategies (e.g., best-in-class investing and ESG exclusion) lead to distributional inefficiencies that can either prevent a transition to a greener economy from materializing or push one where firms have more market power, and consumers are ultimately worse off. We have also demonstrated that these inefficiencies are unlikely to persist if SRI becomes sufficiently popular.

Recognizing the potential inefficiencies of SRI has important implications for regulators, which could be addressed in future research. For example, it would be interesting to study which mechanisms may help direct investors toward a more efficient distribution of green capital. Natural candidates could be redesigning the information reflected in ESG scores, the objectives of SRI funds, and firms' green disclosure requirements. Such interventions may face less political resistance than traditional approaches, like taxes or caps on carbon emissions.

References

Acemoglu, D. (2003). Why not a political Coase theorem? Social conflict, commitment, and politics. *Journal of Comparative Economics, 31*(4), 620–652.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *Quarterly Journal of Economics, 120*(2), 701–728.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy, 124*(1), 1–51.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal, 100*(401), 464–477.

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review, 90*(1), 34–45.

Banerjee, S., Davis, J., & Gondhi, N. (2018). When transparency improves, must prices reflect fundamentals better? *Review of Financial Studies, 31*(6), 2377–2414.

Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review, 83*(4), 1181–1206.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review, 96*(5), 1652–1678.

Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica, 77*(305), 1–19.

Besley, T., & Ghatak, M. (2007). Retailing public goods: The economics of corporate social responsibility. *Journal of Public Economics, 91*(9), 1645–1663.

Besley, T., & Persson, T. (2023). The political economics of green transitions. *Quarterly Journal of Economics, 138*(3), 1863–1906.

Biais, B., & Landier, A. (2022). Emission caps and investment in green technologies. Unpublished manuscript.

Bolton, P., Li, T., Ravina, E., & Rosenthal, H. (2020). Investor ideology. *Journal of Financial Economics, 137*(2), 320–352.

Bonnefon, J.-F., Landier, A., Sastry, P., & Thesmar, D. (2025). The moral preferences of investors: Experimental evidence. *Journal of Financial Economics, 163*, 103955.

Broccardo, E., Hart, O., & Zingales, L. (2022). Exit vs. voice. *Journal of Political Economy, 130*(12).

Bubb, R., & Catan, E. M. (2022). The party structure of mutual funds. *Review of Financial Studies, 35*(6), 2839–2878.

Bulow, J. I., Geanakoplos, J. D., & Klemperer, P. D. (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy, 93*(3), 488–511.

Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review*, 305–333.

Dávila, E., & Parlatore, C. (2021). Trading costs and informational efficiency. *Journal of Finance, 76*(3), 1471–1539.

Dewatripont, M., & Tirole, J. (2024). The morality of markets. *Journal of Political Economy, 132*(8), 2655–2694.

Dikolli, S. S., Frank, M. M., Guo, Z. M., & Lynch, L. J. (2022). Walk the talk: ESG mutual fund voting on shareholder proposals. *Review of Accounting Studies*, 1–33.

Edmans, A. (2023). The end of ESG. *Financial Management, 52*(1), 3–17.

Edmans, A., Levit, D., & Schneemeier, J. (2022). Socially responsible divestment. Unpublished manuscript.

Edmans, A., & Manso, G. (2011). Governance through trading and intervention: A theory of multiple blockholders. *Review of Financial Studies, 24*(7), 2395–2428.

Elhauge, E. (2005). Sacrificing corporate profits in the public interest. *New York University Law Review, 80*, 733–869.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York Times Magazine*.

Gaspar, J.-M., Massa, M., & Matos, P. (2005). Shareholder investment horizons and the market for corporate control. *Journal of Financial Economics, 76*(1), 135–165.

Goldstein, I., Kopytov, A., Shen, L., & Xiang, H. (2022). On ESG investing: Heterogeneous preferences, information, and asset prices. Unpublished manuscript.

Gupta, D., Kopytov, A., & Starmans, J. (2022). The pace of change: Socially responsible investing in private markets. Unpublished manuscript.

Hakenes, H., & Schliephake, E. (2021). Responsible investment and responsible consumption. Unpublished manuscript.

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. *Journal of Law, Finance, and Accounting, 2*(2), 247–274.

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance, 74*(6), 2789–2837.

Heeb, F., Kölbel, J. F., Paetzold, F., & Zeisberger, S. (2023). Do investors care about impact? *Review of Financial Studies, 36*(5), 1737–1787.

Hefti, A. (2017). Equilibria in symmetric games: Theory and applications. *Theoretical Economics, 12*(3), 979–1002.

Heinkel, R., Kraus, A., & Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis, 36*(4), 431–449.

Huang, S., & Kopytov, A. (2023). Sustainable finance under regulation. Unpublished manuscript.

Kakhbod, A., Loginova, U., Malenko, A., & Malenko, N. (2023). Advising the management: A theory of shareholder engagement. *Review of Financial Studies, 36*(4), 1319–1363.

Kaufmann, M., Andre, P., & Kőszegi, B. (2024). Understanding markets with socially responsible consumers. *Quarterly Journal of Economics, 139*(3), 1989–2035.

Kreps, D. M., & Wilson, R. (1982). Sequential equilibria. *Econometrica, 50*(4), 863–894. FOR OA

Landier, A., & Lovo, S. (2025). Socially responsible finance: How to optimize impact. *Review of Financial Studies, 38*(4), 1211–1258.

Levit, D., & Malenko, N. (2011). Nonbinding voting for shareholder proposals. *Journal of Finance, 66*(5), 1579–1614.

Levit, D., Malenko, N., & Maug, E. (2024). Trading and shareholder democracy. *Journal of Finance, 79*(1), 257–304.

López, Á. L., & Vives, X. (2019). Overlapping ownership, R&D spillovers, and antitrust policy. *Journal of Political Economy, 127*(5), 2394–2437.

O'Brien, D. P., & Salop, S. C. (1999). Competitive effects of partial ownership: Financial interest and corporate control. *Antitrust Law Journal, 67*, 559–614.

Oehmke, M., & Opp, M. M. (2025). A theory of socially responsible investment. *Review of Economic Studies, 92*(2), 1193–1225.

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics, 142*(2), 550–571.

Piatti, I., Shapiro, J. D., & Wang, X. (2022). Sustainable investing and public goods provision. Unpublished manuscript.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4*, 25–55. OA only

Speirs, J., Contestabile, M., Houari, Y., & Gross, R. (2014). The future of lithium availability for electric vehicle batteries. *Renewable and Sustainable Energy Reviews, 35*, 183–193.

Starks, L. T. (2023). Presidential address: Sustainable finance and ESG issues—value versus values. *Journal of Finance*.

Appendix A: Proofs

We first introduce the function $\widehat{\chi}$, which serves as a threshold for the fraction χ of $\mathcal{R}$ investors. This expression will be useful in several proofs.

$$\widehat{\chi}(\sigma, n) \equiv \frac{1}{2} + \frac{\sqrt{4\kappa\lambda(N-n)(1-\sigma) + (\lambda(1-\sigma) - N\kappa)^2} - N\kappa}{2\lambda(1-\sigma)} \quad \forall \sigma \in [0,1),$$

and $\widehat{\chi}(1, n) \equiv 1 - \frac{n}{N}$. The threshold $\widehat{\chi}(\cdot)$ is decreasing in σ and n.

Proof of Proposition 1. The derivative of $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ with respect to σ equals

$$\frac{\lambda^2 \chi(1-\chi)}{N\kappa} \mathbb{1}\left\{\sigma > 1 - \frac{N\kappa}{\lambda(1-\chi)}\right\} + \frac{\partial}{\partial \sigma}\left(\frac{\partial \Pi_j}{\partial \sigma_j}(\sigma, \sigma \cdot \vec{1})\right). \tag{15}$$

As $\frac{\partial \Pi_j}{\partial \sigma_j}(1, \vec{1}) < 0$ by Assumption 1, the equilibrium abatement level in any equilibrium without SRI is strictly lower than 1; if $\frac{\partial \Pi_j}{\partial \sigma_j}(0, \vec{0}) \leq 0$, then $\sigma^* = 0$ is an equilibrium, else there exists an interior equilibrium $\sigma^* \in (0,1)$. In both cases, there may be more than one equilibrium without SRI. Let σ_0 denote the smallest abatement policy across equilibria without SRI. Note that

$$\sigma_0 > 1 - \frac{N\kappa}{\lambda(1-\chi)} \iff \chi > 1 - \frac{N\kappa}{\lambda(1-\sigma_0)} = \widehat{\chi}(\sigma_0, N).$$

Denoting $\widehat{\chi}(\sigma_0, N) \equiv \chi_0$, we have that:[26]

1. For $\chi \leq \chi_0$, $\sigma = \sigma_0$ is such that $\Gamma(\sigma_0, \sigma_0 \cdot \vec{1}; s^{\mathcal{R}}(\sigma_0)) \leq 0$ (with equality in an interior solution) with $s^{\mathcal{R}} = 0$, and so the game always admits at least one equilibrium with no SRI. However, for larger values of σ such that $s^{\mathcal{R}} > 0$, $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ can be increasing in σ, and so the first-order condition $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma)) = 0$ game can admit at least one other solution $\sigma^* > \sigma_0$, which corresponds to an equilibrium with SRI. These equilibria, however, cannot exist for χ sufficiently small, as the positive term in the derivative (15) vanishes for $\chi \to 0$, and so the function $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ crosses zero for the last time (if ever) at the largest abatement level across equilibria without SRI. As a result, there exists a threshold $\widetilde{\chi} \in (0, \chi_0]$ such that for $\chi \leq \widetilde{\chi}$ the game only admits equilibria without SRI, whereas for $\chi \in (\widetilde{\chi}, \chi_0]$ equilibria without SRI can coexist with at least one equilibrium with SRI.

2. For $\chi > \chi_0$, if $\sigma_0 > 0$, as $\frac{\partial \Pi_j}{\partial \sigma_j} = 0$ and $s^{\mathcal{R}} > 0$, $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ is positive at $\sigma = \sigma_0$ (hence, a fortiori, also at the larger abatement levels of any other equilibrium without SRI), which implies that the equilibria without SRI do not exist. Then, this

[26] Note that the region of parameters $\chi \in (0, \chi_0]$ is empty if $\chi_0 \leq 0$ – i.e., if $\lambda \leq \frac{N\kappa}{1-\sigma_0}$. In these circumstances, the results in point 2 below apply for all $\chi \in (0,1)$.

expression either equals zero at some $\sigma^* > \sigma_0$, with any such σ^* corresponding to an equilibrium with SRI ($s^{\mathcal{R}} > 0$), or it is positive for all $\sigma \in [0,1]$ (which is always the case for $\chi \to 1$ under Assumption 1) and so $\sigma^* = 1$ is the unique equilibrium, and it features SRI. If instead $\sigma_0 = 0$, i.e. $\Gamma(0, 0 \cdot \vec{1}; 0) < 0$, then for $\chi > \chi_0$ but not too large so that also $\Gamma(0, 0 \cdot \vec{1}; \frac{\chi}{N\kappa}[N\kappa - \lambda(1-\chi)]) < 0$, $\sigma^* = 0$ is the abatement policy in an equilibrium with SRI ($s^{\mathcal{R}} = \frac{\chi}{N\kappa}[N\kappa - \lambda(1-\chi)] > 0$ and so, as above, there are no equilibria without SRI).

These arguments imply that the equilibrium is always unique if $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ is globally decreasing. From (15) it follows that this is always the case if

$$\left| \frac{\partial}{\partial \sigma} \left(\frac{\partial \Pi_j}{\partial \sigma_j}(\sigma, \sigma \cdot \vec{1}) \right) \right| > \frac{\lambda^2 \chi (1-\chi)}{N\kappa} \quad \forall \sigma \in [0,1],$$

this threshold being weakly larger than $\frac{\lambda^2}{4N\kappa}$ for all $\chi \in (0,1)$. Therefore, under Assumption 2, the symmetric equilibrium is unique for all $\chi \in (0,1)$, and it features no SRI (hence, $\sigma^* = \sigma_0$, with σ_0 being the abatement level in the unique equilibrium without SRI) for all $\chi \leq \chi_0$, and SRI (hence, $\sigma^* > \sigma_0$) for all $\chi > \chi_0$.

As $\Gamma(\sigma, \sigma \cdot \vec{1}; s^{\mathcal{R}}(\sigma))$ is continuous and decreasing in χ,[27] by the implicit function theorem, it follows that σ^* is a continuous and increasing function of χ (strictly so in the equilibrium with SRI provided that $\sigma^* \in (0,1)$). To establish that also $s^{\mathcal{R}}$ is increasing in χ, note that σ^* depends on χ only indirectly through $s^{\mathcal{R}}$, and that σ^* is increasing in $s^{\mathcal{R}}$. As a result, the sign of $\frac{\partial s^{\mathcal{R}}}{\partial \chi}$ must be the same as the sign of $\frac{\partial \sigma^*}{\partial \chi}$. □

Proof of Proposition 2. In an asymmetric equilibrium, $\mathcal{R}$ investors must be indifferent across all firms $j > \underline{n}$ and strictly prefer them over firms $j \leq \underline{n}$, whereas $\mathcal{N}$ investors must be indifferent across all firms $j \leq \overline{n}$ and strictly prefer them over firms $j > \overline{n}$. Using the notation in the Proposition for firms' abatement efforts, and an analogous notation for expected profits and share prices, we must have

$$\begin{cases} \overline{\Pi} - \overline{p} - \lambda(1-\overline{\sigma}) = \widehat{\Pi} - \widehat{p} - \lambda(1-\widehat{\sigma}) > \underline{\Pi} - \underline{p} - \lambda(1-\underline{\sigma}) \\ \overline{\Pi} - \overline{p} < \widehat{\Pi} - \widehat{p} = \underline{\Pi} - \underline{p}. \end{cases} \tag{16}$$

[27] Differentiating this function for $\chi \geq \widehat{\chi}(\sigma^*, N)$ with respect to χ gives

$$-\frac{\lambda(N\kappa - \lambda(1-\sigma^*)(1-2\chi))}{N\kappa}.$$

As this derivative is decreasing in χ, a sufficient condition for it to be negative for all $\chi \geq \widehat{\chi}(\sigma^*, N)$ is

$$-\frac{\lambda(N\kappa - \lambda(1-\sigma^*)(1-2\widehat{\chi}(\sigma^*, N)))}{N\kappa} = -\frac{\lambda(\lambda(1-\sigma^*) - N\kappa)}{N\kappa} < 0,$$

which holds for all $N\kappa < \lambda(1-\sigma^*)$. This condition needs to hold as it is equivalent to $\widehat{\chi}(\sigma^*, N) > 0$.

As a result,

$$\overline{\Pi} - \overline{p} - \lambda(1 - \widehat{\sigma}) < \widehat{\Pi} - \widehat{p} - \lambda(1 - \widehat{\sigma}) = \overline{\Pi} - \overline{p} - \lambda(1 - \overline{\sigma}),$$

where the inequality follows from $\mathcal{N}$ investors' preferences and the equality from $\mathcal{R}$ investors' indifference, which implies $\widehat{\sigma} < \overline{\sigma}$; and similarly

$$\underline{\Pi} - \underline{p} - \lambda(1 - \underline{\sigma}) < \widehat{\Pi} - \widehat{p} - \lambda(1 - \widehat{\sigma}) = \underline{\Pi} - \underline{p} - \lambda(1 - \widehat{\sigma}),$$

where the inequality follows from $\mathcal{R}$ investors' preferences and the equality from $\mathcal{N}$ investors' indifference, which implies $\underline{\sigma} < \widehat{\sigma}$.[28]

We next characterize the existence conditions for these equilibria. Let $\overline{\alpha}_j^{\mathcal{R}} > 0$ and $\underline{\alpha}_j^{\mathcal{R}} > 0$ be the fraction of $\mathcal{R}$ investors who buy shares in any firm $j > \overline{n}$ and $j \in (\underline{n}, \overline{n}]$, respectively, with $\sum_{j>\overline{n}} \overline{\alpha}_j^{\mathcal{R}} + \sum_{j \in (\underline{n}, \overline{n}]} \underline{\alpha}_j^{\mathcal{R}} = 1$. Similarly, let $\overline{\alpha}_j^{\mathcal{N}} > 0$ and $\underline{\alpha}_j^{\mathcal{N}} > 0$ be the fraction of $\mathcal{N}$ investors who buy shares in any firm $j \leq \underline{n}$ and $j \in (\underline{n}, \overline{n}]$, respectively, with $\sum_{j \leq \underline{n}} \overline{\alpha}_j^{\mathcal{N}} + \sum_{j \in (\underline{n}, \overline{n}]} \underline{\alpha}_j^{\mathcal{N}} = 1$.

Then, the market-clearing condition for firm $j \in \mathcal{J}$ is given by:

$$1 = \begin{cases} (1-\chi)\overline{\alpha}_j^{\mathcal{N}} \frac{1}{\kappa} (\Pi_j - p_j) & \text{if } j \leq \underline{n} \\ \chi \underline{\alpha}_j^{\mathcal{R}} \frac{1}{\kappa} [\Pi_j - p_j - \lambda(1 - \sigma_j)] + (1-\chi)\underline{\alpha}_j^{\mathcal{N}} \frac{1}{\kappa} (\Pi_j - p_j) & \text{if } j \in (\underline{n}, \overline{n}] \\ \chi \overline{\alpha}_j^{\mathcal{R}} \frac{1}{\kappa} [\Pi_j - p_j - \lambda(1 - \sigma_j)] & \text{if } j > \overline{n}. \end{cases} \tag{17}$$

Note that for $j > \overline{n}$ we require that all firms have the same $\Pi_j - p_j - \lambda(1 - \sigma_j)$ to make $\mathcal{R}$ investors indifferent. It directly follows that $\overline{\alpha}_j^{\mathcal{R}} \equiv \overline{\alpha}^{\mathcal{R}}$ for $j > \overline{n}$. By the same logic, we have $\overline{\alpha}_j^{\mathcal{N}} \equiv \overline{\alpha}^{\mathcal{N}}$ for $j \leq \underline{n}$. Similarly, all $j \in (\underline{n}, \overline{n}]$ must offer the same $\Pi_j - p_j$ and $\Pi_j - p_j - \lambda(1 - \sigma_j)$ so that $\underline{\alpha}_j^{\mathcal{R}} \equiv \underline{\alpha}^{\mathcal{R}}$ and $\underline{\alpha}_j^{\mathcal{N}} \equiv \underline{\alpha}^{\mathcal{N}}$.

The equilibrium abatement policies are obtained as a fixed-point of firms' best-response functions in Eqn. (3), with

$$s_j^{\mathcal{R}} = \begin{cases} 0 & \text{for } j \leq \underline{n} \\ \frac{\chi \underline{\alpha}^{\mathcal{R}} (\kappa - \underline{\alpha}^{\mathcal{N}} \lambda (1-\chi)(1-\widehat{\sigma}))}{\kappa (\chi \underline{\alpha}^{\mathcal{R}} + (1-\chi) \underline{\alpha}^{\mathcal{N}})} \in (0,1) & \text{for } \underline{n} < j \leq \overline{n} \\ 1 & \text{for } j > \overline{n}, \end{cases}$$

For firms $j > \overline{n}$, Assumption 1 implies that $\overline{\sigma} = 1$.

Proof of $\underline{n} = \overline{n} \equiv n$. In this case, we have that $\underline{\alpha}^{\mathcal{R}} = \underline{\alpha}^{\mathcal{N}} = 0$ so that $\overline{\alpha}^{\mathcal{R}} = \frac{1}{N-n}$ and

[28] Note that, while we considered so far the more general case with three "firm types" – i.e., $0 < \underline{n} < \overline{n} < N$ – the inequalities in Eqn. (16) imply that (i) if no firm is held exclusively by $\mathcal{R}$ investors ($\overline{n} = N$), $\underline{\sigma} < \widehat{\sigma}$; (ii) if no firm has "mixed ownership" ($\underline{n} = \overline{n}$), $\underline{\sigma} < \overline{\sigma}$; and (iii) if no firm is held exclusively by $\mathcal{N}$ investors ($\underline{n} = 0$), $\widehat{\sigma} < \overline{\sigma}$. The remainder of the proof only relies on the presence of SRI in at least one firm.

$\overline{\alpha}^{\mathcal{N}} = \frac{1}{n}$. Hence, we have $\sigma_j = \overline{\sigma}$ for $j \leq n$ and $\sigma_j = \underline{\sigma}$ otherwise, with $\overline{\sigma} > \underline{\sigma}$ by the results above. Using these results, we can rewrite Eqn. (7) as:

$$\Pi_j - p_j = \begin{cases} \frac{(N-n)\kappa}{\chi} + \lambda(1-\overline{\sigma}) & \text{if } j > n \\ \frac{n\kappa}{1-\chi} & \text{if } j \leq n. \end{cases}$$

Equilibrium existence requires that $\mathcal{R}$ ($\mathcal{N}$) investors are indifferent between holding shares in any firm $j > n$ ($j \leq n$) and that they prefer this to buying shares in other firms or not trading at all. We denote the expected utility of investor $i \in [0,1]$ by U_i. Plugging in the equilibrium demands given in Eqn. (5) yields $U_i = \frac{\kappa}{2}(s_i^*)^2$, increasing in s_i^*. In this equilibrium, the demands for an $\mathcal{R}$ and $\mathcal{N}$ investor is, thus, equal to $\frac{N-n}{\chi}$ and $\frac{n}{1-\chi}$, respectively. Since $U_i > 0$ for all investors, we just have to ensure that neither type has an incentive to deviate to buy shares in a different firm.

If an individual $\mathcal{R}$ investor deviates and trades in firm $j \leq n$, its optimal demand is equal to $\frac{1}{\kappa}\left(\frac{n\kappa}{1-\chi} - \lambda(1-\underline{\sigma})\right)$. We therefore require that:

$$\frac{N-n}{\chi} > \frac{1}{\kappa}\left(\frac{n\kappa}{1-\chi} - \lambda(1-\underline{\sigma})\right) \Leftrightarrow \chi < \widehat{\chi}(\underline{\sigma}, n).$$

Similarly, if an individual $\mathcal{N}$ investor deviates and trades in firm $j > n$, its optimal demand is equal to $\frac{1}{\kappa}\left(\frac{N-n)\kappa}{\chi} + \lambda(1-\overline{\sigma})\right)$. We therefore require that:

$$\frac{n}{1-\chi} > \frac{1}{\kappa}\left(\frac{(N-n)\kappa}{\chi} + \lambda(1-\overline{\sigma})\right) \Leftrightarrow \chi > \widehat{\chi}(\overline{\sigma}, n).$$

Summing up, these equilibria exist if and only if $\widehat{\chi}(\overline{\sigma}, n) < \chi < \widehat{\chi}(\underline{\sigma}, n)$. Therefore, in these equilibria, as $(\overline{\sigma}, \underline{\sigma})$ do not depend on χ, we have that $\underline{\chi} \equiv \widehat{\chi}(\overline{\sigma}, n)$ and $\overline{\chi} \equiv \widehat{\chi}(\underline{\sigma}, n)$, with $\underline{\chi} < \overline{\chi}$ because $\widehat{\chi}(\cdot)$ is decreasing in σ and $\overline{\sigma} > \underline{\sigma}$.

Last, we clarify the sense in which the best-in-class equilibria are those with the highest concentration of responsible capital. We formalize this notion using the Herfindahl-Hirschman Index (HHI), a standard measure of concentration.

The HHI of responsible capital is defined as the sum of squared portfolio shares across firms:

$$\text{HHI}_{\mathcal{R}} = \sum_{j=1}^{N} z_j^2,$$

where z_j denotes the share of total responsible investment allocated to firm j. If $\mathcal{R}$ investors invest in some firms, then the index ranges from $1/N$ (equal-weighted allocation across all firms) to 1 (full allocation to a single firm). Hence, the HHI is *maximized* when all responsible investors allocate their capital to a *single firm*, reaching $\text{HHI}_{\mathcal{R}} = 1$.

Two equilibria achieve this maximum value:

1. An equilibrium in which responsible investors hold *only firm* N and non-responsible investors hold some shares in *all firms*, including firm N; and

2. An equilibrium in which responsible investors hold only firm N, and non-responsible investors hold only firms 1 through $N-1$, with *no overlap* in firm ownership.

We refer to the second allocation as the *best-in-class equilibrium*, and it is the one we focus on in the main text. This equilibrium not only yields the highest possible $\text{HHI}_{\mathcal{R}}$, but also maximizes the concentration of non-responsible capital ($\text{HHI}_{\mathcal{N}}$) among the remaining firms, given the complete segmentation of ownership across investor types.

Thus, among all equilibria with concentrated responsible investment, the best-in-class equilibrium features the highest possible concentration of responsible capital.

Proof of $\underline{n} < \overline{n}$. In this case, firms $j \in (\underline{n}, \overline{n}]$ are held by both $\mathcal{N}$ and $\mathcal{R}$ investors, which, by the results above, will set $\sigma_j = \widehat{\sigma} \in (\underline{\sigma}, \overline{\sigma})$. Using the results above, we can rewrite Eqn. (7) as:

$$\Pi_j - p_j = \begin{cases} \frac{\kappa}{(1-\chi)\overline{\alpha}^{\mathcal{N}}} & \text{if } j \le \underline{n} \\ \frac{\kappa + \lambda(1-\widehat{\sigma})\chi\underline{\alpha}^{\mathcal{R}}}{\chi\underline{\alpha}^{\mathcal{R}} + (1-\chi)\underline{\alpha}^{\mathcal{N}}} & \text{if } j \in (\underline{n}, \overline{n}] \\ \frac{\kappa}{\chi\overline{\alpha}^{\mathcal{R}}} + \lambda(1-\overline{\sigma}) & \text{if } j > \overline{n}, \end{cases} \tag{18}$$

with $(N-\overline{n})\overline{\alpha}^{\mathcal{R}} + (\overline{n}-\underline{n})\underline{\alpha}^{\mathcal{R}} = 1$ and that $\underline{n}\overline{\alpha}^{\mathcal{N}} + (\overline{n}-\underline{n})\underline{\alpha}^{\mathcal{N}} = 1$. Next, we use the indifference conditions for $\mathcal{R}$ and $\mathcal{N}$ investors to solve for the optimal $\underline{\alpha}^{\theta}$ and $\overline{\alpha}^{\theta}$ with $\theta \in \{\mathcal{R}, \mathcal{N}\}$.

1. Suppose that $\underline{n} > 0$ and $\overline{n} < N$ so that we have three distinct firm types (with different σ_j). $\mathcal{R}$ investors must be indifferent between holding firms $j > \overline{n}$ and $j \in (\underline{n}, \overline{n}]$:[29]
$$\frac{\kappa}{\chi\overline{\alpha}^{\mathcal{R}}} = \frac{\kappa - \lambda(1-\widehat{\sigma})(1-\chi)\underline{\alpha}^{\mathcal{N}}}{\chi\underline{\alpha}^{\mathcal{R}} + (1-\chi)\underline{\alpha}^{\mathcal{N}}}.$$
Similarly, $\mathcal{N}$ investors must be indifferent between holding firms $j > \overline{n}$ and $j \in (\underline{n}, \overline{n}]$:
$$\frac{\kappa}{(1-\chi)\overline{\alpha}^{\mathcal{N}}} = \frac{\kappa + \lambda(1-\widehat{\sigma})\chi\underline{\alpha}^{\mathcal{R}}}{\chi\underline{\alpha}^{\mathcal{R}} + (1-\chi)\underline{\alpha}^{\mathcal{N}}}.$$
Next, we sum over the N market-clearing conditions in Eqn. (17) and the indifference conditions described above to write:
$$N\kappa = \Pi_j - p_j - \lambda(1-\widehat{\sigma})\chi = \frac{\kappa + \lambda(1-\widehat{\sigma})\chi\underline{\alpha}^{\mathcal{R}}}{\chi\underline{\alpha}^{\mathcal{R}} + (1-\chi)\underline{\alpha}^{\mathcal{N}}} - \lambda(1-\widehat{\sigma})\chi \quad \forall j \in (\underline{n}, \overline{n}].$$

[29] As above, $\mathcal{R}$ investors always make positive profits when buying shares in firms $j > \overline{n}$. Therefore, the following indifference condition implies that they are willing to participate in the financial market. The same argument applies to $\mathcal{N}$ investors.

We can then solve for $\overline{\alpha}^{\mathcal{R}}$ and $\overline{\alpha}^{\mathcal{N}}$:

$$\overline{\alpha}^{\mathcal{R}} = \frac{\kappa}{\chi(N\kappa - \lambda(1-\widehat{\sigma})(1-\chi))},$$

$$\overline{\alpha}^{\mathcal{N}} = \frac{\kappa}{(1-\chi)(N\kappa + \lambda(1-\widehat{\sigma})\chi)}.$$

$\underline{\alpha}^{\mathcal{R}}$ and $\underline{\alpha}^{\mathcal{N}}$ follow from $(N-\overline{n})\overline{\alpha}^{\mathcal{R}} + (\overline{n}-\underline{n})\underline{\alpha}^{\mathcal{R}} = 1$ and $\underline{n}\overline{\alpha}^{\mathcal{N}} + (\overline{n}-\underline{n})\underline{\alpha}^{\mathcal{N}} = 1$. Equilibrium existence requires $(N-\overline{n})\overline{\alpha}^{\mathcal{R}} \in (0,1)$ and $\underline{n}\overline{\alpha}^{\mathcal{N}} \in (0,1)$, which is equivalent to $\widehat{\chi}(\widehat{\sigma},\overline{n}) < \chi < \widehat{\chi}(\widehat{\sigma},\underline{n})$.

2. Suppose that $\underline{n} > 0$ and $\overline{n} = N$ so that firms $j \leq \underline{n}$ are only held by $\mathcal{N}$ investors and firms $j \in (\underline{n}, N]$ are held by both types. Therefore, we have $\overline{\alpha}^{\mathcal{R}} = 0$ and $\underline{\alpha}^{\mathcal{R}} = \frac{1}{N-\underline{n}}$. Summing over the market clearing conditions and using the indifference of $\mathcal{N}$ investors yields:

$$N\kappa = \Pi_j - p_j - \lambda(1-\widehat{\sigma})\chi = \frac{\kappa + \lambda(1-\widehat{\sigma})\chi\frac{1}{N-\underline{n}}}{\chi\frac{1}{N-\underline{n}} + (1-\chi)\underline{\alpha}^{\mathcal{N}}} - \lambda(1-\widehat{\sigma})\chi \quad \forall j \leq \overline{n}.$$

Solving this equation for $\underline{\alpha}^{\mathcal{N}}$ leads to:

$$\underline{\alpha}^{\mathcal{N}} = \frac{\kappa(1 - N\frac{\chi}{N-\underline{n}}) + \lambda(1-\widehat{\sigma})\frac{\chi}{N-\underline{n}}(1-\chi)}{(1-\chi)\left[N\kappa + \lambda(1-\widehat{\sigma})\chi\right]}.$$

Equilibrium existence requires that $(N-\underline{n})\underline{\alpha}^{\mathcal{N}} \in (0,1)$ or $\chi < \widehat{\chi}(\widehat{\sigma},\underline{n})$. Finally, $\mathcal{R}$ investors are willing to hold firms $j > \underline{n}$ if and only if:

$$\Pi_j - p_j - \lambda(1-\widehat{\sigma}) > 0 \Leftrightarrow N\kappa - \lambda(1-\widehat{\sigma})(1-\chi) > 0 \Leftrightarrow \chi > \widehat{\chi}(\widehat{\sigma}, N).$$

3. Suppose that $\underline{n} = 0$ and $\overline{n} < N$ so that firms $j > \overline{n}$ are solely held by $\mathcal{R}$ investors and all other firms are held by both types of investors. Therefore, we have $\overline{\alpha}^{\mathcal{N}} = 0$ and $\underline{\alpha}^{\mathcal{N}} = \frac{1}{\overline{n}}$. Summing over the market-clearing conditions and using the indifference of $\mathcal{R}$ investors yields:

$$N\kappa = \Pi_j - p_j - \lambda(1-\widehat{\sigma})\chi = \frac{\kappa + \lambda(1-\widehat{\sigma})\chi\underline{\alpha}^{\mathcal{R}}}{\chi\underline{\alpha}^{\mathcal{R}} + \frac{1-\chi}{\overline{n}}} - \lambda(1-\widehat{\sigma})\chi \quad \forall j > \underline{n}.$$

Solving this equation for $\underline{\alpha}^{\mathcal{R}}$ leads to:

$$\underline{\alpha}^{\mathcal{R}} = \frac{\kappa(\overline{n} - N(1-\chi)) - \lambda\chi(1-\chi)(1-\widehat{\sigma})}{\overline{n}\chi(N\kappa - \lambda(1-\chi)(1-\widehat{\sigma}))}.$$

Equilibrium existence requires that $\overline{n}\underline{\alpha}^{\mathcal{R}} \in (0,1)$, which is equivalent to $\chi > \widehat{\chi}(\widehat{\sigma},\overline{n})$.

The expressions derived above also imply that $\mathcal{N}$ investors make positive trading profits and do not have an incentive to deviate from this equilibrium.

To wrap up, in all equilibria with $\underline{n} < \overline{n}$, where $\widehat{\sigma}$ depends on χ through $s_j^{\mathcal{R}}$, $\underline{\chi}$ is the solution to $\chi = \widehat{\chi}(\widehat{\sigma}, \overline{n})$, and $\overline{\chi}$ is the solution to $\chi = \widehat{\chi}(\widehat{\sigma}, \underline{n})$, with $\underline{\chi} < \overline{\chi}$ because $\widehat{\chi}(\cdot)$ is decreasing in n. □

Proof of Lemma 1. The proof proceeds in two parts, corresponding to Part 1 and Part 2 of the Proposition.

Proof of Part 1. For $\chi \leq \chi_0$, $s_j^{\mathcal{R}} = 0$ for all j in the symmetric equilibrium and for all $j \leq N-1$ in the best-in-class equilibrium. Hence, any such σ_j is obtained from $\frac{\partial \pi_j}{\partial \sigma_j}(\sigma_j, \vec{\sigma}_{-j}) - \mathcal{C}'(\sigma_j) \leq 0$, with equality in an interior solution ($\sigma_j > 0$). Hence, $\underline{\sigma}$ is different from $\sigma^* = \sigma_0$ if and only if $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} \neq 0$ for some $j' \neq j$ (hence, $\underline{\sigma} = \sigma_0$ under strategic independence) and $\vec{\sigma}_{-j} \neq \vec{\sigma}_0$. Since, by Proposition 2, $\underline{\sigma} < \overline{\sigma} = 1$, we have that, if abatement choices are strategic substitutes (complements), $\frac{\partial \pi_j}{\partial \sigma_j}(\sigma, (\sigma \cdot \vec{1}, 1)) < (>)\frac{\partial \pi_j}{\partial \sigma_j}(\sigma, \sigma \cdot \vec{1})$ for all $\sigma < 1$, and so (by the second-order condition of Program (2)) we must have $\underline{\sigma} \leq (\geq) \sigma_0$, with strict inequality if $\sigma_0 > 0$.

Proof of Part 2. For $\chi > \chi_0$, the symmetric equilibrium features $\sigma^* \geq \sigma_0$, with strict inequality in an interior solution. Therefore, from the previous results it follows that $\underline{\sigma} < \sigma^*$ if abatement choices are strategic substitutes or independent, whereas the comparison between $\underline{\sigma}$ and σ^* is in general ambiguous under strategic complementarity as both values are larger than σ_0. □

Proof of Proposition 3. As seen above, the expected utility of investor $i \in [0,1]$ is $U_i = \frac{\kappa}{2}(s_i^*)^2$, increasing in the number of shares purchased, s_i^*. Note that all investors of a certain type must have the same U_i, and therefore s_i^*, in a specific equilibrium. Hence, we let s^θ denote the equilibrium demand of an individual θ-investor with $\theta \in \{\mathcal{R}, \mathcal{N}\}$. As a result, the total demand across all firms for $\mathcal{R}$ ($\mathcal{N}$) investors is equal to $\chi s^{\mathcal{R}}$ ($(1-\chi)s^{\mathcal{N}}$). Imposing market clearing, i.e., equating total demand across all firms with the total supply N, leads to $\chi s^{\mathcal{R}} + (1-\chi)s^{\mathcal{N}} = N$. Let $s^{\theta'}$ and $s^{\theta''}$ denote the equilibrium demand of a θ-type investor in two different equilibria. Since $\chi s^{\mathcal{R}'} + (1-\chi)s^{\mathcal{N}'} = \chi s^{\mathcal{R}''} + (1-\chi)s^{\mathcal{N}''}$, we must have that $s^{\mathcal{R}'} > s^{\mathcal{R}''}$ if and only if $s^{\mathcal{N}'} < s^{\mathcal{N}''}$. Therefore, $\mathcal{R}$ and $\mathcal{N}$ investors never strictly prefer the same equilibrium.

Proof of Part 1. As $\overline{\sigma} = 1$, the best-in-class equilibrium exists for $\widehat{\chi}(1,1) = \frac{1}{N} < \chi < \widehat{\chi}(\underline{\sigma}, 1)$. We have that $\frac{1}{N} < \chi_0$ if and only if

$$\lambda > \overline{\lambda} \equiv \frac{\kappa N^2}{(N-1)(1-\sigma_0)}.$$

The threshold $\overline{\lambda}$ is well defined because $\sigma_0 < 1$ does not depend on λ.

Proof of Part 2. As $\chi_0 = \widehat{\chi}(\sigma_0, N)$ and $\widehat{\chi}(\cdot)$ is decreasing in σ and N, a sufficient condition for $\widehat{\chi}(\underline{\sigma}, 1) > \chi_0$ is that $\sigma_0 \geq \underline{\sigma}$, which always holds if firms' abatement policies are strategic substitutes or independent, as shown in Lemma 1, Part 1.

$\mathcal{R}$ investors collectively buy $\chi s^{\mathcal{R}} = 1$ shares in the best-in-class equilibrium (as $s^{\mathcal{R}}_{j\leq N-1} = 0$ and $s^{\mathcal{R}}_N = 1$), whereas in the symmetric equilibrium they buy no shares for $\chi \leq \chi_0$ and a positive amount of shares, increasing in χ, for $\chi > \chi_0$. Therefore, they strictly prefer the best-in-class equilibrium (whenever it exists) for $\chi < \chi_0$. We prove below that this result holds for all $\chi < \chi''$, with $\chi'' \in (\chi_0, \overline{\chi})$.

Recall that each $\mathcal{R}$ investor's preference-adjusted return equals $\frac{\kappa}{\chi}$, which decreases in χ, in the best-in-class equilibrium, and $N\kappa - \lambda(1-\chi)(1-\sigma^*)$, increasing in χ and σ^*, in the symmetric equilibrium. As $\sigma^* > \underline{\sigma}$ whenever firms' abatement policies are strategic substitutes or independent (see Lemma 1, Part 2), a sufficient condition for $\mathcal{R}$ investors to prefer the symmetric equilibrium is $\frac{\kappa}{\chi} < N\kappa - \lambda(1-\chi)(1-\underline{\sigma})$, which simplifies as

$$\frac{\kappa(1-N\chi) + \lambda\chi(1-\chi)(1-\underline{\sigma})}{\chi} < 0.$$

As this expression equals zero at $\chi = \overline{\chi}$, we can conclude that $\mathcal{R}$ investors prefer the symmetric equilibrium if and only if $\chi > \chi''$ for some $\chi'' \in (\chi_0, \overline{\chi})$. Finally, note that if $\chi_0 < \frac{1}{N}$, so that interval in which the two equilibria coexist coincides with $(\frac{1}{N}, \overline{\chi})$, at $\chi = \frac{1}{N}$ $\mathcal{R}$ investors always prefer the best-in-class equilibrium, strictly so whenever $\sigma^* < 1$ – i.e., $\chi'' \geq \frac{1}{N}$, with strict inequality if $\sigma^* < 1$ at $\chi = \frac{1}{N}$. □

Proof of Proposition 4. We prove Part 1 and Part 2 separately.

Proof of Part 1. By the results in Lemma 1, Part 1, for all $\chi \leq \chi_0$ we have that $\underline{\sigma} \geq \sigma_0$ under strategic complementarity or independence, implying that $\overline{\sigma} + (N-1)\underline{\sigma} > N\sigma_0$. By contrast, as under strategic substitutability $\underline{\sigma} < \sigma_0$ in an interior solution, one might have that $\overline{\sigma} + (N-1)\underline{\sigma} < N\sigma_0$. This possibility result is shown in Figure 4.

To derive a sufficient condition for this result to hold, recall that Assumption 1 guarantees that in the best-in-class equilibrium, the best-in-class firm ($j = N$ with $s^{\mathcal{R}}_j = 1$) chooses $\overline{\sigma} = 1$. Eqn. (3) then implies that the excluded firms ($j < N$ with $s^{\mathcal{R}}_j = 0$) choose $\underline{\sigma} = 0$ if and only if $\Gamma(0, (1, \vec{0}); 0) \leq 0$. Under this condition, aggregate abatement equals $\overline{\sigma} + (N-1)\underline{\sigma} = 1$ in the best-in-class equilibrium. This value is smaller than in the equilibrium without SRI ($s^{\mathcal{R}}_j = 0$ for all j) if and only if $\sigma_0 > \frac{1}{N}$. By Eqn. (3), this is the case if and only if $\Gamma(\frac{1}{N}, \frac{1}{N} \cdot \vec{1}; 0) > 0$. These two conditions ensure that aggregate abatement is lower in the best-in-class equilibrium than in the symmetric equilibrium without SRI.

Proof of Part 2. For $\chi > \chi_0$, $\sigma^* > \sigma_0$ is increasing in χ, whereas $(\underline{\sigma}, \overline{\sigma})$ are independent of χ. Therefore, even if firms' abatement efforts are independent or strategic complements, it is possible that aggregate abatement is larger in the symmetric equilibrium than in the best-in-class equilibrium for χ sufficiently large – i.e., $\chi \in (\chi', \overline{\chi})$, with $\chi' > \chi_0$. Note that, if $\chi_0 < \frac{1}{N}$, so that interval in which the two equilibria coexist coincides with $(\frac{1}{N}, \overline{\chi})$, it is possible that aggregate abatement is larger in the symmetric equilibrium already at $\chi = \frac{1}{N}$ – i.e., that $\chi' = \frac{1}{N}$.

We show that the result $\chi' < \overline{\chi}$ is always true under strategic independence. In order for $\overline{\sigma} + (N-1)\underline{\sigma} < N\sigma^*$, it must be that

$$\left[\frac{\partial \Pi_j}{\partial \sigma_j}(\sigma) + \lambda s^{\mathcal{R}}(\sigma)\right]\Bigg|_{\sigma=\frac{1}{N}\overline{\sigma}+\frac{N-1}{N}\underline{\sigma}} > 0, \tag{19}$$

given that π_j does not depend on $\vec{\sigma}_{-j}$ under strategic independence, where $s^{\mathcal{R}}(\sigma)$ given in Eqn. 8 is increasing in σ. At $\chi = \overline{\chi}$, we have that $s^{\mathcal{R}}\left(\frac{1}{N}\overline{\sigma} + \frac{N-1}{N}\underline{\sigma}\right) - \frac{1}{N}$ equals

$$\frac{(\overline{\sigma}-\underline{\sigma})\left(N\sqrt{\kappa^2N^2+\lambda^2(1-\underline{\sigma})^2-2\kappa\lambda(N-2)(1-\underline{\sigma})}-\kappa N^2+\lambda(N-2)(1-\underline{\sigma})\right)}{2\lambda N^2(1-\underline{\sigma})^2} > 0,$$

and so $s^{\mathcal{R}}\left(\frac{1}{N}\overline{\sigma} + \frac{N-1}{N}\underline{\sigma}\right) > \frac{1}{N}$. Therefore, as the left-hand side of (19) is increasing in $s^{\mathcal{R}}(\cdot)$, a sufficient condition for it to hold in a left-neighborhood of $\chi = \overline{\chi}$ is

$$F(N) \equiv \frac{\partial \Pi_j}{\partial \sigma_j}\left(\frac{1}{N}\overline{\sigma} + \frac{N-1}{N}\underline{\sigma}\right) + \frac{\lambda}{N} \geq 0.$$

Assumption 1 guarantees that $F(1) > 0$; moreover, in an interior solution $(\frac{\partial \Pi_j}{\partial \sigma_j}(\underline{\sigma}) = 0)$, $F(N) \to 0$ as $N \to \infty$. As with strategic independence $\overline{\sigma} = 1$ and $\underline{\sigma} = \sigma_0 \in (0,1)$ do not depend on N, we have that

$$F'(N) = -\frac{1}{N^2}\left[\lambda + (\overline{\sigma}-\underline{\sigma})\frac{\partial^2 \Pi}{\partial \sigma_j^2}\left(\frac{1}{N}\overline{\sigma} + \frac{N-1}{N}\underline{\sigma}\right)\right],$$

and

$$F''(N) = \frac{1}{N}\left[\frac{(\overline{\sigma}-\underline{\sigma})^2}{N^3}\frac{\partial^3 \Pi}{\partial \sigma_j^3}\left(\frac{1}{N}\overline{\sigma} + \frac{N-1}{N}\underline{\sigma}\right) - 2F'(N)\right].$$

Therefore,

- If $\frac{\partial^3 \Pi}{\partial \sigma_j^3} \leq 0$, any stationary point of $F(N)$ – i.e., any N^* such that $F'(N^*) = 0$ – is a maximum point. As $F(1) > 0$ and $\lim_{N\to\infty} F(N) = 0$, it follows that either $F(N)$ is inverted-U-shaped, or it is globally decreasing. In either case, $F(N) > 0$ for all N.

- If $\frac{\partial^3 \Pi}{\partial \sigma_j^3} > 0$, any stationary point of $F(N)$ is a minimum point. As $F(1) > 0$ and $\lim_{N\to\infty} F(N) = 0$, it follows that $F(N)$ is is globally decreasing and positive.

The inequality $F(N) \geq 0$ is thus always satisfied. As in equilibrium $s^{\mathcal{R}}$ is increasing in

χ, it then follows that there exists a threshold $\chi' \in (\chi_0, \overline{\chi})$ such that $\sum_j \sigma_j$ is larger in the symmetric equilibrium than in the best-in-class equilibrium if and only if $\chi > \chi'$. $\square$

Proof of Proposition 5. We establish the results in Part 1 and 2 separately.

Proof of Part 1. We know from Proposition 3 that the best-in-class equilibrium can coexist with the symmetric equilibrium without SRI ($\underline{\chi} < \chi_0$) and, in this case, it is always preferred by $\mathcal{R}$ investors, and from Proposition 4 that it can feature lower aggregate abatement under strategic substitutability. The former result holds by continuity in a right-neighborhood of χ_0; the latter result holds a fortiori for $\chi > \chi_0$.

With independence in abatement efforts, we have proved that, relative to the symmetric equilibrium, the best-in-class equilibrium is preferred by $\mathcal{R}$ investors for $\chi < \chi'' \in (\chi_0, \overline{\chi})$ and reduces aggregate abatement for $\chi > \chi' \in (\chi_0, \overline{\chi})$. As $\mathcal{R}$ investors are indifferent if and only if they collectively hold the same shares in the two equilibria, we have that $s^{\mathcal{R}} = \frac{1}{N}$ at $\chi = \chi''$. In the proof of Proposition 4, Part 2, we have shown that if $s^{\mathcal{R}} = \frac{1}{N}$, the symmetric equilibrium features higher aggregate abatement, and so $\chi'' > \chi'$ whenever $\chi'' > \frac{1}{N}$. Recall that $\chi'' = \frac{1}{N}$ if and only if $\sigma^* = 1$ at $\chi = \frac{1}{N}$, which implies that aggregate abatement is always larger in the symmetric relative to the best-in-class equilibrium – i.e., $\chi'' = \chi' = \frac{1}{N}$.

Proof of Part 2. As shown in Proposition 2, the best-in-class equilibrium exists for $\widehat{\chi}(1,1) = \frac{1}{N} < \chi < \widehat{\chi}(\underline{\sigma}, 1)$ (as $\overline{\sigma} = 1$ under Assumption 1). As $\widehat{\chi}(\underline{\sigma}, 1)$ is decreasing in $\underline{\sigma}$, the measure of the interval of χ in which such equilibrium exists shrinks as $\underline{\sigma}$ increases and equals zero for $\underline{\sigma} = 1$. $\square$

Proof of Proposition 6. With independent products, the first-order condition of consumer h's utility maximization gives

$$u'(x_{hj}) = \rho_j + \lambda \mathbb{1}_{h \in \mathcal{R}} \mathbb{1}_{\theta_j = d},$$

yielding demand $x_{hj} = u'^{-1}(\rho_j + \lambda \mathbb{1}_{h \in \mathcal{R}} \mathbb{1}_{\theta_j = d})$ for each firm j.[30] The market-clearing condition for each firm j's product is then

$$\chi u'^{-1}(\rho_j + \lambda \mathbb{1}_{\theta_j = d}) + (1 - \chi) u'^{-1}(\rho_j) = 1.$$

This condition implies that each clean product ($\theta_j = c$) is sold at price $\rho_j = u'(1)$, whereas the market-clearing price of each dirty product is $\rho_j \in (u'(1) - \lambda, u'(1))$, obtained by the above condition for $\mathbb{1}_{\theta_j = d} = 1$. As these prices do not depend on the number of clean firms in the market, firm j's expected profit does not depend on rivals' abatement levels.

[30]This treatment assumes that responsible consumers have positive demand also for dirty products. However, the same results hold if $\lambda > u'(0)$, so that $\mathcal{R}$ consumers *boycott* dirty products.

Define $PS(\theta_j)$ as the total surplus generated by firm j's product:

$$S(\theta_j) = \chi u[u'^{-1}(\rho_j + \lambda \mathbb{1}_{\theta_j=d})] + (1-\chi)u[u'^{-1}(\rho_j)].$$

The concavity of $u(\cdot)$ implies that this value is higher when consumption is equalized across consumers, i.e., $PS(c) = u(1) > S(d)$. As $\sigma_j = \Pr[\theta_j = c]$, we have that $\mathbb{E}[S(\theta_j)] = S(d) + [S(c) - S(d)]\sigma_j$. Therefore, $PS = \sum_{j\in\mathcal{J}} \mathbb{E}[S(\theta_j)] = S(d) + [S(c) - S(d)] \sum_{j\in\mathcal{J}} \sigma_j$ is an increasing function of $\sum_{j\in\mathcal{J}} \sigma_j$. □

Proof of Proposition 7. As products are perfect substitutes, a consumer h only buys from a firm $j^* \in \text{argmin}_{j\in\mathcal{J}}(\rho_j + \lambda \mathbb{1}_{h\in\mathcal{R}} \mathbb{1}_{\theta_j=d})$, and its demand is $x_{hj^*} = u'^{-1}(\rho_j + \lambda \mathbb{1}_{h\in\mathcal{R}} \mathbb{1}_{\theta_j^*=d})$. Consider equilibria where the products of firms operating with the same technology have the same demand and market-clearing price. Denoting by $\rho_c(n_c)$ ($\rho_d(n_c)$) the market-clearing price for each clean (dirty) firm as function of the number $n_c \in \{1, ..., N-1\}$ of clean firms,[31] the equilibrium takes one of the following forms:

1. *$\mathcal{R}$ consumers buy both products and $\mathcal{N}$ consumers buy dirty products.* For $\mathcal{R}$ consumers to be indifferent among all products, it must be that $\rho_d + \lambda = \rho_c$ (which implies that $\rho_c > \rho_d$, and so $\mathcal{N}$ consumers prefer dirty products).

   Denoting by $\rho \equiv \rho_c = \rho_d + \lambda$, and letting $\alpha^{\mathcal{R}}$ be the fraction of $\mathcal{R}$ consumers that buy any of the clean products, the market-clearing condition for clean products is

$$\alpha^{\mathcal{R}} \chi u'^{-1}(\rho) = n_c, \tag{20}$$

   and for dirty products is

$$(1-\alpha^{\mathcal{R}})\chi u'^{-1}(\rho) + (1-\chi)u'^{-1}(\rho - \lambda) = N - n_c. \tag{21}$$

   Any such equilibrium is a pair $(\rho, \alpha^{\mathcal{R}})$, with $\alpha^{\mathcal{R}} \in (0,1)$, that jointly solve Eqns. (20)-(21).

   Summing up conditions (20) and (21) yields

$$\chi u'^{-1}(\rho) + (1-\chi)u'^{-1}(\rho - \lambda) = N,$$

   which admits a unique solution ρ, independent of n_c and $\alpha^{\mathcal{R}}$. Because from Eqn. (20) we have $\rho = u'\left(\frac{n_c}{\alpha^{\mathcal{R}}\chi}\right)$, this implies that the ratio $\frac{n_c}{\alpha^{\mathcal{R}}}$ is a constant. Denoting this constant ratio by y, substituting (20) into (21), after simple manipulations,

[31] For $n_c = 0$ ($n_c = N$), the analysis is identical to case 1 (case 3) below for $\alpha^{\mathcal{R}} = 0$ ($\alpha^{\mathcal{N}} = 1$). Case 1 never arises if $\mathcal{R}$ consumers boycott dirty products ($\lambda > u'(0)$), but the results of Proposition 7 still go through.

yields

$$u'\left(\frac{y}{\chi}\right) - \lambda = u'\left(\frac{N-y}{1-\chi}\right), \tag{22}$$

which, solved for y, pins down $\alpha^{\mathcal{R}}$. The left-hand side (right-hand side) of this equation is decreasing (increasing) in y, hence it is increasing (decreasing) in $\alpha^{\mathcal{R}}$. As it must be that $\alpha^{\mathcal{R}} < 1$ for this equilibrium to exist, given that $y = n_c$ for $\alpha^{\mathcal{R}} = 1$, the considered equilibrium exists, and it is unique, if and only if

$$\lambda < u'\left(\frac{n_c}{\chi}\right) - u'\left(\frac{N-n_c}{1-\chi}\right),$$

this upper bound being decreasing in n_c and positive for $n_c < N\chi$. Hence, this equilibrium exists if and only if $n_c < \underline{n}_c$, where $\underline{n}_c \in [0, N\chi)$.

Therefore, for any $n_c < \underline{n}_c$, using Eqn. (22) product market surplus is given by

$$PS(n_c) = \chi u\left(\frac{y}{\chi}\right) + (1-\chi)u\left(\frac{N-y}{1-\chi}\right), \tag{23}$$

and it is constant varying n_c.

2. *$\mathcal{R}$ consumers buy clean products and $\mathcal{N}$ consumers buy dirty products.* For $\mathcal{R}$ ($\mathcal{N}$) consumers to prefer clean (dirty) products, it must be that $\rho_d + \lambda > \rho_c > \rho_d$.

   The market-clearing conditions for clean and dirty products immediately yield the respective unique equilibrium prices,

$$\chi u'^{-1}(\rho_c) = n_c \implies \rho_c = u'\left(\frac{n_c}{\chi}\right),$$

   and

$$(1-\chi)u'^{-1}(\rho_d) = N - n_c \implies \rho_d = u'\left(\frac{N-n_c}{1-\chi}\right).$$

   Therefore, the conditions $\rho_d + \lambda > \rho_c > \rho_d$ boil down to

$$n_c < N\chi \quad \text{and} \quad \lambda > u'\left(\frac{n_c}{\chi}\right) - u'\left(\frac{N-n_c}{1-\chi}\right).$$

   Hence, this equilibrium exists if and only if $n_c \in [\underline{n}_c, \overline{n}_c]$, with $\overline{n}_c$ being the highest integer smaller than $N\chi$.

   Therefore, for any $n_c \in [\underline{n}_c, \overline{n}_c]$, product market surplus equals

$$PS(n_c) = \chi u\left(\frac{n_c}{\chi}\right) + (1-\chi)u\left(\frac{N-n_c}{1-\chi}\right),$$

   which, since $\frac{n_c}{\chi} < \frac{N-n_c}{1-\chi}$ and $u(\cdot)$ is concave, is strictly increasing in n_c and larger than for any $n_c < \underline{c}$.

3. *$\mathcal{R}$ consumers buy clean products and $\mathcal{N}$ consumers buy both products.* For $\mathcal{N}$ consumers to be indifferent among all products, it must be that $\rho_c = \rho_d$, which implies that $\mathcal{R}$ consumers prefer clean products.

   Denoting by ρ the common market-clearing price, and by $\alpha^{\mathcal{N}}$ the fraction of $\mathcal{N}$ consumers who buy clean products, the market-clearing condition for clean products is

   $$\chi u'^{-1}(\rho) + \alpha^{\mathcal{N}}(1-\chi)u'^{-1}(\rho) = n_c, \tag{24}$$

   and for dirty products is

   $$(1-\alpha^{\mathcal{N}})(1-\chi)u'^{-1}(\rho) = N - n_c. \tag{25}$$

   Any such equilibrium is a pair $(\rho, \alpha^{\mathcal{N}})$, with $\alpha^{\mathcal{N}} \in (0,1)$, that jointly solve Eqns. (24)-(25).

   Summing up conditions (24) and (25) yields the unique equilibrium price $\rho = u'(N)$, which, substituted into (25), after simple manipulations, gives

   $$\alpha^{\mathcal{N}} = 1 - \frac{N - n_c}{(1-\chi)N}.$$

   Therefore, this equilibrium exists and it is unique if and only if $\alpha^{\mathcal{N}} \in (0,1)$, which boils down to $n_c > N\chi$ – i.e., $n_c > \overline{n}_c$.

   Therefore, for any $n_c > \overline{n}_c$, product market surplus is $PS(n_c) = u(N)$, which is independent of n_c and larger than the one for $n_c \in [\underline{n}_c, \overline{n}_c]$ because $u(\cdot)$ is concave.

To wrap up, denoting by $\rho_c(n_c)$ ($\rho_d(n_c)$) the equilibrium price of clean (dirty) products as function of the number of clean firms in the market, the foregoing analysis implies that $\rho_c(n_c)$ ($\rho_d(n_c)$) is decreasing (increasing) in n_c, strictly so for $n_c \in [\underline{n}_c, \overline{n}_c]$, and $\rho_c(n_c) \geq \rho_d(n_c)$, with strict inequality for all $n \leq \overline{n}_c$. Then, denoting by $n_{c,-j}$ the number of firm j's clean rivals, we can write j's expected profit as

$$\pi_j = \sum_{\tilde{n}_{c,-j}=0}^{N-1} \Pr[n_{c,-j} = \tilde{n}_{c,-j}] \left\{\rho_d(\tilde{n}_{c,-j}) + \sigma_j[\rho_c(\tilde{n}_{c,-j}+1) - \rho_d(\tilde{n}_{c,-j})]\right\},$$

from which we have that $\frac{\partial \pi_j}{\partial \sigma_j} = \sum_{\tilde{n}_{c,-j}=0}^{N-1} \Pr[n_{c,-j} = \tilde{n}_{c,-j}][\rho_c(\tilde{n}_{c,-j}+1) - \rho_d(\tilde{n}_{c,-j})]$. This function is decreasing in any rival's abatement effort because (1) as argued above, the difference $\rho_c(\tilde{n}_{c,-j}+1) - \rho_d(\tilde{n}_{c,-j})$ is decreasing in $\tilde{n}_{c,-j}$; and (2) the probability distribution of $n_{c,-j} \in \{0, ..., N-1\}$ is Poisson-Binomial with success probabilities $\vec{\sigma}_{-j}$ and an increase in any rival's abatement effort shifts the distribution of n_c in a first-order stochastic dominance sense, making higher realizations of $n_{c,-j}$ more likely. Therefore, we can conclude that $\frac{\partial^2 \pi_j}{\partial \sigma_j \partial \sigma_{j'}} \leq 0$, with strict inequality provided that $\sigma_{j'} < 1$ for at least $\overline{n}_c$ firm j's rivals.

Finally, expected product market surplus is $PS = \sum_{\tilde{n}_c=0}^{N} \Pr[n_c = \tilde{n}_c] S(\tilde{n}_c))$, where the probability distribution of $n_c \in \{0, ..., N\}$ is Poisson-Binomial with success probabilities σ_j. By the previous results, $PS(\tilde{n}_c)$ is increasing in $\tilde{n}_c$ (strictly so for $\tilde{n}_c \in [\underline{n}_c, \overline{n}_c]$), hence the same first-order stochastic dominance argument put forward above implies that PS is increasing in each σ_j. Suppose that the best-in-class equilibrium exists in a neighborhood of χ_0. Then, as strategic substitutability implies that $\underline{\sigma} < \sigma_0 < \overline{\sigma}$ (recall Lemma 1), whether SRI concentration increases product market surplus is in general ambiguous. □

Proof of Proposition 8. When firms' products are perfect complements, consumer h optimally chooses to consume $x_{hj} \equiv x_h$ for all $j \in \mathcal{J}$, where x_h solves

$$\max_{x_h \geq 0} \; u(x_h) - x_h \sum_{j \in \mathcal{J}} \left(\rho_j + \lambda \mathbb{1}_{h \in \mathcal{R}} \mathbb{1}_{\theta_j = d} \right).$$

The optimal x_h is then obtained from

$$u'(x_h) \leq \sum_{j} \left(\rho_j + \lambda \mathbb{1}_{h \in \mathcal{R}} \mathbb{1}_{\theta_j = d} \right),$$

with equality in an interior solution ($x_h > 0$). Assuming $\lambda > u'(0)$ guarantees that $\mathcal{R}$ consumers do not buy any product unless all firms sell clean products. Then, we have to distinguish two cases:

- If $0 \leq n_c \leq N - 1$, only $\mathcal{N}$ consumers buy firms' (clean or dirty) products, so that the market-clearing condition for each firm j is $(1-\chi) u'^{-1} \left(\sum_{j \in \mathcal{J}} \rho_j \right) = 1$, yielding the symmetric equilibrium price

$$\rho(n_c) = \frac{1}{N} u' \left(\frac{1}{1-\chi} \right),$$

independent of n_c.

- If $n_c = N$, all consumers buy the same quantity $x_h = u'^{-1} \left(\sum_{j \in \mathcal{J}} \rho_j \right)$ from each firm j, and so the market-clearing conditions yield the symmetric equilibrium price

$$\rho(N) = \frac{1}{N} u' (1).$$

Therefore, firm j's expected profit equals

$$\pi_j = \frac{1}{N}\left[u'\left(\frac{1}{1-\chi}\right) + \left(u'(1) - u'\left(\frac{1}{1-\chi}\right)\right)\sigma_j \prod_{j'\neq j}\sigma_{j'}\right],$$

so that

$$\frac{\partial^2 \pi_j}{\partial\sigma_j \partial\sigma_{j'}} = \frac{1}{N}\left(u'(1) - u'\left(\frac{1}{1-\chi}\right)\right)\prod_{j''\neq j,j'}\sigma_{j''} \geq 0,$$

with strict inequality provided that $\sigma_j > 0$ for all j. Firms' abatement policies are then strategic complements.

Finally, expected product market surplus equals

$$PS = (1-\chi)u\left(\frac{1}{1-\chi}\right) + \left(u(1) - u\left(\frac{1}{1-\chi}\right)\right)\prod_{j\in\mathcal{J}}\sigma_j,$$

and so is an increasing function of $\prod_{j\in\mathcal{J}}\sigma_j$. Therefore:

- For $\chi \leq \chi_0$, the symmetric equilibrium features no SRI ($\sigma_j = \sigma_0$) and in best-in-class equilibrium $\overline{\sigma} > \underline{\sigma} > \sigma_0$ by Lemma 1. Therefore, $\prod_{j\in\mathcal{J}}\sigma_j$, and so expected product market surplus, is larger in best-in-class equilibrium than in the symmetric equilibrium.[32]

- For $\chi > \chi_0$, the symmetric equilibrium features SRI ($\sigma_j = \sigma^* > \sigma_0$), and SRI concentration can reduce $\prod_{j\in\mathcal{J}}\sigma_j$. Indeed, if $\sum_{j\in\mathcal{J}}\sigma_j$ is lower in the best-in-class equilibrium – i.e., if $\overline{\sigma}+(N-1)\underline{\sigma} < N\sigma^*$ – then a fortiori SRI concentration reduces $\prod_{j\in\mathcal{J}}\sigma_j$ (and so PS) – i.e., $\overline{\sigma}\underline{\sigma}^{N-1} < (\sigma^*)^N$. This result follows from the AM-GM Inequality $\frac{\overline{\sigma}+(N-1)\underline{\sigma}}{N} > \sqrt[N]{\overline{\sigma}\underline{\sigma}^{N-1}}$. However, SRI concentration can reduce PS even when it increases aggregate abatement: e.g., as $\overline{\sigma} = 1$, for $N = 2$ this happens if $\max\{2\sigma^* - 1, 0\} < \underline{\sigma} < (\sigma^*)^2$. □

Appendix B: Supplementary material

B.1 Data appendix

We obtain mutual funds' stock holding information from Thomson/Refinitiv S12 (S12 hereafter). Since S12 does not have an indicator for ESG funds, following Dikolli et al. (2022), we rely on Morningstar's classification to identify ESG funds in S12 data (https://www.morningstar.com/esg-screener). We classify a fund as an ESG fund if Morningstar states that the fund's management identifies the fund as sustainability-focused in public filings ("Sustainable Investment by Prospectus"). Otherwise, we classify

[32] The same results hold considering other equilibria with concentrated SRI.

the fund as a non-ESG fund.

We match S12 data and Morningstar data using fund tickers. Since tickers can be reused, we manually compare fund names in S12 and Morningstar and verify the match if it can be inferred from the fund names that the same sponsor manages two funds. Throughout this process, we identify 82 ESG funds in S12 data. We define $GreenCapital_{it}$ as the aggregated value of stock holdings in firm i at year t by ESG funds. Similarly, $Non - GreenCapital_{it}$ is defined as the aggregated stock holdings in firm i at year t by non-ESG funds.

We obtain Refinitiv's ESG Combined Scores for US firms listed in NYES and NASDAQ during 2002-2020. The list of NYSE/NASDAQ-listed stocks and stock prices is obtained from CRSP.

To construct the HHI indexes for green (non-green) capital at the industry level, we first compute the share of green capital in a firm for a given year over the aggregated green (non-green) capital in the industry that the firm belongs to in the same year. The HHI indexes for green (non-green) capital are then obtained by summing the squared shares of green (non-green) capital in the industry. Similarly, we compute HHI indexes for ESG combined scores at the industry level are by summing the squared shares of ESG combined scores in a firm for a given year over the aggregated ESG combined scores in the industry that the firm belongs to in the same year. These HHI indexes allow us to measure the concentration of green and non-green capital and ESG combined scores within each industry.

B.2 Stability of concentration

We have shown in Proposition 3 that, when two equilibria coexist, $\mathcal{R}$ and $\mathcal{N}$ investors always disagree on which equilibrium they prefer. So, even if $\mathcal{R}$ investors prefer to concentrate, it is not clear how they choose which equilibrium is played, since neither equilibria is payoff-dominant. The typical stability-based equilibrium refinements (e.g., trembling-hand perfect equilibria (Selten, 1975) and sequential equilibria (Kreps and Wilson, 1982) do not immediately apply, since the investors' action space is continuous in our model.[33] A simple and intuitive way to still explore stability in our model (in the traditional sense of robustness to small perturbations to the equilibrium strategies) is to consider a setting where, when two equilibria coexist, a small set of investors commit to playing the strategies they play in their preferred equilibrium.

Proposition B.1 (stability of concentration). *Suppose the best-in-class equilibrium and the symmetric equilibrium coexist when the mass of* $\mathcal{R}$ *investors* χ *is in a neighborhood*

[33]Investors form conjectures about the profits and abatement choices of each firm and, based on these conjectures, submit a demand schedule, in the space of continuous, positive functions, for each share. The original formulations of trembling-hand perfect equilibria (Selten, 1975) and sequential equilibria (Kreps and Wilson, 1982) both require a finite action space, so they do not apply to our setting.

of the threshold χ_0, which is such that the symmetric equilibrium features SRI iff $\chi > \chi_0$. Define a perturbation of the original game as one where a small mass of investors ($\widehat{\alpha} \to 0$) of a given type always plays the strategies they play in their preferred equilibrium (among these two). The best-in-class equilibrium is always robust to such perturbations (that is, it continues to exist in the perturbed game), while the symmetric equilibrium is not.

Within a subset of the values of χ for which $\mathcal{R}$ investors prefer the best-in-class equilibrium (that is, for χ around the threshold χ_0), the symmetric equilibrium is less "stable": even a small subset of $\mathcal{R}$ investors committing to the strategies they play in the best-in-class equilibrium is sufficient to break the symmetric equilibrium, while a small subset of $\mathcal{N}$ investors committing to play like in the symmetric equilibrium does not suffice to break the best-in-class equilibrium.[34] This suggests a simple mechanism for equilibrium selection in favor of equilibria with concentration (e.g., some SRI funds making commitments to best-in-class investing).

The mechanism that makes the best-in-class equilibrium relatively more stable is similar to the traditional tâtonnement dynamics: When a fraction of $\mathcal{R}$ investors commits to only buying shares of firm j (rather than splitting among the N firms), the other firms become more attractive to $\mathcal{N}$ investors, since there is relatively less demand pressure for their shares. As some $\mathcal{N}$ investors move toward the other firms, j's ownership becomes more responsible and, thus, σ_j goes up. This attracts more $\mathcal{R}$ investors to hold only j, triggering a set of mutually best-responses that diverge from the symmetric equilibrium. This process is not at work in the opposite case: When a fraction of $\mathcal{N}$ investors commits to buying all firms (rather than only the excluded ones), the excluded firms become more attractive to the remaining $\mathcal{N}$ investors, since there is relatively less demand pressure in these firms. So, neither the remaining $\mathcal{N}$ investors in the excluded firms, nor the $\mathcal{R}$ investors in the best-in-class firm,[35] want to move away from the best-in-class equilibrium strategies.

B.3 Shareholder voting

In this section, we consider a variation of the model where shareholders influence firms' abatement efforts through voting. The modeling of the voting stage is similar to Levit et al. (2024).

[34] Of course, as the fraction of commitment types approaches 1, almost all $\mathcal{N}$ investors play the strategies they play in the symmetric equilibrium, at which point $\mathcal{R}$ investors' best-responses also converge to the symmetric strategies.

[35] The commitment types do not affect the abatement of the best-in-class firm, since $\lambda > \mathcal{C}'(1)$ means that a small set of $\mathcal{N}$ investors cannot move the firm away from full abatement, but they do increase price pressure in the firm. In the best-in-class equilibrium, however, $\mathcal{R}$ investors have a strict preference for the best-in-class firm, so a small increase in price pressure is not enough to change their portfolio choices. In the symmetric equilibrium, all investors are instead indifferent across firms, which makes the equilibrium inherently less robust to perturbations.

At time $t = 2$, given the distribution of SRI across firms, the shareholders of each firm vote on an ESG proposal that defines the firm's objective when choosing its abatement efforts. If the proposal is rejected, firm j chooses σ_j to maximize its expected profits Π_j; if the proposal is accepted, it chooses σ_j to maximize $\Pi_j - \lambda(1 - \sigma_j)$, internalizing the firm's pollution externality $\lambda(1 - \sigma_j)$. For simplicity, we assume each share has one vote, and focus on subgame perfect equilibria in undominated strategies of the voting game.[36] The proposal is accepted if at least a fraction $\tau \in (0, 1)$ of all shares are cast in favor (τ may be different than $\frac{1}{2}$ to reflect super-majority voting).

Equilibrium analysis. To simplify the exposition, we focus on the analysis of symmetric equilibria and asymmetric equilibria with full separation (i.e., $\mathcal{N}$ investors hold firms $j \leq n$ and $\mathcal{R}$ investors hold firms $j > n$). Defining the threshold $\widehat{\chi}(\sigma, n)$ as in the main model, the following holds:

Proposition B.2 (Shareholder voting). *Suppose shareholders of each firm vote on an ESG proposal that defines whether the firm should internalize the pollution externality when choosing its abatement efforts. The following two types of equilibria may exist:*

- *(Symmetric equilibria.) A symmetric equilibrium always exists. It features no SRI if $\chi \leq \widehat{\chi}(\sigma_0, N)$, where, as in the main model, $\sigma_0 = \arg\max_{\sigma_j} \Pi_j(\sigma_j, \sigma_0 \cdot \vec{1})$, and SRI otherwise. In particular, defining $\widetilde{\chi}(\sigma) \equiv \frac{1}{2} + \frac{\sqrt{(\lambda(1-\sigma)-\kappa N)^2 + 4\kappa\lambda N(1-\sigma)\tau} - \kappa N}{2\lambda(1-\sigma)}$, for all $\chi \in [\widehat{\chi}(\sigma_0, N), \widetilde{\chi}(\sigma_0))$ there is a symmetric equilibrium with SRI where firms still choose σ_0, and for all $\chi \geq \widetilde{\chi}(\sigma^*)$, with $\widetilde{\chi}(\sigma^*) \in (\widehat{\chi}(\sigma_0, N), \widetilde{\chi}(\sigma_0))$, there is another symmetric equilibrium with SRI where firms choose $\sigma^* = \arg\max_{\sigma_j} \Pi_j(\sigma_j, \sigma^* \cdot \vec{1}) - \lambda(1 - \sigma_j) > \sigma_0$.*

- *(Asymmetric equilibria.) Asymmetric equilibria with full separation where $\mathcal{R}$ investors target firms $j > n$, and $\sigma_j = \underline{\sigma}$ ($\sigma_j = \overline{\sigma}$) for $j \leq n$ ($j > n$), exist if and only if $\widehat{\chi}(\overline{\sigma}, n) < \chi < \widehat{\chi}(\underline{\sigma}, n)$, where $(\underline{\sigma}, \overline{\sigma})$ are the same as in the main model.*

There exist parameter configurations such that the only equilibria with SRI are asymmetric and, in these equilibria, aggregate abatement is lower than in the benchmark without SRI.

The asymmetric equilibria with full separation exist for the same parameter space and feature the same abatement efforts as in the baseline model. This is because investors correctly anticipate that, for all $\tau \in (0, 1)$, the ESG proposal will be accepted in firms $j > n$ (where all shareholders are $\mathcal{R}$ types and so vote in favor) and rejected in firms $j \leq n$ (where all shareholders are $\mathcal{N}$ types and so vote against). This implies that firms'

[36] This restriction is common in the literature on voting games. The usual assumption in this literature is that dispersed shareholders vote as if they were pivotal (see, e.g., Baron and Ferejohn, 1989, and Austen-Smith and Banks, 1996).

policies and profits in this candidate equilibrium are as in the main model, which makes the investors' demands unchanged.

Similarly, there always exists a symmetric equilibrium (with or without SRI). As in the main model, the symmetric equilibrium features no SRI if the mass of responsible investors is sufficiently small. In this case, we have shown (see Figure 4) that, under strategic substitutability, aggregate CSR investments can be lower in the best-in-class equilibrium.

Unlike in the main model, the game always admits two symmetric equilibria with SRI when χ takes intermediate values (i.e., for $\widetilde{\chi}(\sigma^*) \leq \chi < \widetilde{\chi}(\sigma_0)$). This follows from a coordination problem among $\mathcal{R}$ investors: if each of them expects like-minded investors to buy a large number of shares in each firm so that $s^{\mathcal{R}*} > \tau$ and the proposal is expected to be accepted, the investor has incentives to buy more shares. Else, $\mathcal{R}$ investors still buy positive shares in all firms (because $\chi > \widehat{\chi}(\sigma_0, N)$ implies that the share price would be too low if only $\mathcal{N}$ investors trade in the market), but not enough to have an impact on firms' abatement efforts. Thus, in this version of the model, the presence of SRI may not affect aggregate abatement when $\mathcal{R}$ investors do not concentrate in equilibrium.

B.4 Broad vs narrow mandate

In this section, we consider a variation of the model where $\mathcal{R}$ investors suffer a disutility from the pollution generated by *all* firms, irrespective of their holdings. Formally, given $K_i = \frac{\kappa}{2}(\sum_{j\in\mathcal{J}} s_{ij})^2$, the investors' portfolio problem becomes the following:

$$\max_{s_{ij}\geq 0} \sum_{j\in\mathcal{J}} s_{ij}\left(\Pi_j - p_j\right) - \sum_{j\in\mathcal{J}} \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_j) - K_i. \tag{26}$$

Similar to the main model, given its ownership, firm j's optimal abatement policy solves:

$$\max_{\sigma_j\in[0,1]} \Pi_j - s_j^{\mathcal{R}} \sum_{k\in\mathcal{J}} \lambda(1-\sigma_k). \tag{27}$$

The objective in Program (27) is a weighted average of the investors' expected payoff, where the weights are proportional to the shares held by each shareholder.

The way we model the broad investment mandate calls for two remarks. First, the pollution disutility in Program (26) is independent of $\mathcal{R}$ investors' portfolio choices, so it does not affect their optimal demands. $\mathcal{R}$ and $\mathcal{N}$ investors then have the same demand functions in this version of the model. It follows that the equilibrium characterization is the same if *all* investors suffer from aggregate pollution, but firms only internalize the disutility of $\mathcal{R}$ types when choosing their abatement efforts (consistent with these investors engaging more to influence firms' green policies).

Second, j's objective in Program (27) includes the abatement levels of the other firms. However, firms move simultaneously in our model, so j takes these levels as given when choosing σ_j. Therefore, each individual firm is still unable to influence the others' abatement. It follows that, even if $\mathcal{R}$ investors care about aggregate pollution, they are not able to push firms to internalize the *indirect* effects of their investments on other firms (i.e., the crowding-in and crowding-out effects).

Equilibrium analysis. $\mathcal{R}$ investors' pollution disutility is independent of their portfolio choices. Moreover, since investors are atomistic, they take firms' abatement efforts as given. Hence, the pollution disutility does not distort $\mathcal{R}$ investors' demand compared to $\mathcal{N}$ types. Therefore, each investor's demand is now

$$s_{ij} = \begin{cases} \max\left\{\frac{1}{\kappa}\left[\Pi_{j^*} - p_{j^*}\right], 0\right\} & \text{for } j^* \in \text{argmax}_j\{\Pi_j - p_j\} \\ 0 & \text{for } j \neq j^*. \end{cases} \tag{28}$$

Since all investors have the same demand, equilibria without SRI do not exist in this framework.

Proposition B.3 (**broad mandate**). *When $\mathcal{R}$ investors have a broad mandate, the game admits a unique symmetric equilibrium and a continuum of asymmetric equilibria. In some equilibria, $\mathcal{R}$ investors invest only in a subset of firms, and aggregate abatement efforts are lower than in the symmetric equilibrium.*

The game features a prevalence of asymmetric equilibria also under the broad mandate assumption: Since investors can be allocated across firms in any arbitrary way compatible with the market-clearing conditions, the game admits infinitely many equilibria, of which *only* one is symmetric. For given ownership structure, firms' investments are the same as in the main model. It follows that the asymmetric equilibria feature dispersion in abatement efforts, where firms with more SRI invest more. Similar to the main model, SRI concentration can then reduce aggregate abatement efforts relative to the benchmark without SRI and, a fortiori, to the symmetric equilibrium, which features SRI. Yet, unlike in the main analysis, here these asymmetric equilibria with low abatement levels can be Pareto dominated by the symmetric equilibrium, because $\mathcal{R}$ investors do care about aggregate abatement.

Finally, it is worth noting that, since the disutility for aggregate pollution does not affect the investors' demand, adding this disutility to their payoff in our main model would not affect the equilibrium characterization. Formally, let $\xi > 0$ denote the intensity of the warm-glow component in $\mathcal{R}$ investors' utility (in our main model, we set $\xi = 1$ for simplicity), and suppose investor i either maximizes

$$\sum_{j \in \mathcal{J}} s_{ij}\left(\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}}\xi\lambda(1-\sigma_j)\right) - \sum_{j \in \mathcal{J}} \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_j) - K_i$$

(i.e., only $\mathcal{R}$ types suffer from aggregate pollution), or

$$\sum_{j\in\mathcal{J}} s_{ij}\left(\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}}\xi\lambda(1-\sigma_j)\right) - \sum_{j\in\mathcal{J}}\lambda(1-\sigma_j) - K_i$$

(i.e., both $\mathcal{R}$ and $\mathcal{N}$ types suffer from aggregate pollution).

In either case, the investor's demand is the same as in the main model, with $\xi\lambda$ in place of λ, and so the equilibrium characterization would also be the same.[37]

B.5 Alternative trading costs

In this section, we consider an alternative specification of the trading cost, which depends on the individual holdings of each firm:

$$\mathcal{K}(\vec{s}_{ij}) \equiv \kappa \sum_{j\in\mathcal{J}} s_{ij}^2. \tag{29}$$

Notice that, under the assumption that firms' profits are normally distributed with mean Π_j and exogenous variance $\frac{\kappa}{r}$, with $r > 0$, this specification is equivalent to a portfolio-choice problem where investors have CARA preferences with risk aversion coefficient r.

Equilibrium Analysis. For a given distribution of SRI, the optimal CSR investments are the same as in the main model. In the ownership market stage, taking the first-order conditions of investors' problem, we now have

$$s_{ij} = \max\left\{\frac{1}{\kappa}\left[\Pi_j - p_j - \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma_j)\right], 0\right\} \quad \forall j \in \mathcal{J}.$$

It follows that, in equilibrium, each $\mathcal{R}$ investor buys shares in any firm j for which $\Pi_j - p_j - \lambda(1-\sigma_j) > 0$ and each $\mathcal{N}$ investor buys shares in all firms. This is because (a) when $\mathcal{R}$ investors have positive demand, $\mathcal{N}$ investors must also have positive demand, and (b) in order for all markets to clear, $\mathcal{N}$ investors must have positive demand also for firms in which $\mathcal{R}$ investors have zero demand – i.e., similar to our baseline model, in equilibrium $\Pi_j > p_j$ for all $j \in \mathcal{J}$.

The marginal cost of acquiring a small position in a firm is zero, independent of the investor's holdings in other firms. Hence, $\mathcal{N}$ types invest in *all* firms in equilibrium. $\mathcal{R}$ investors, however, might still choose to exclude a subset of firms if the externalities they generate are too high.

[37] In the firm's problem, we would have $(1+\xi)\lambda$ in place of λ, so the equilibrium abatement policies would be slightly different compared to the main model. However, the features of the equilibrium described in Propositions 1 and 2 would continue to hold.

Proposition B.4 (per-firm trading costs). *Consider the trading cost in Eqn. (29). Then, all the equilibria are such that $\mathcal{N}$ investors demand positive shares in all firms and $\mathcal{R}$ investors demand positive shares in $n \in \{0, ..., N\}$ firms. Any such equilibrium exists if and only if*

$$1 - \frac{\kappa}{\lambda(1 - \overline{\sigma}(n))} < \chi \leq 1 - \frac{\kappa}{\lambda(1 - \underline{\sigma}(n))}, \tag{30}$$

where $\overline{\sigma}(n) > \underline{\sigma}(n)$ are the equilibrium abatement levels derived from Eqn. (3).

The equilibrium existence conditions are less tractable in this version of the model, because equilibrium values $(\overline{\sigma}(n), \underline{\sigma}(n))$ depend on χ (through $\mathcal{R}$ investors' aggregate demands in each firm), so the endpoints of the interval for χ defined by the existence conditions in Eqn. (30) are themselves a function of χ. This is not the case for the equilibria with $s^{\mathcal{R}}_{j \leq n} = 0$ and $s^{\mathcal{R}}_{j > n} = 1$ we obtained in the main model. While the analysis is accordingly more complicated in this setting, the main results carry through under this alternative specification of the trading cost.

First, asymmetric equilibria, featuring concentrated SRI, also arise in this version of the model. Similar to the main model, $\mathcal{R}$ investors' impact on abatement policies creates a strategic complementarity in their portfolio choices along the *extensive margin*: by concentrating in a subset of firms $j \geq N - n$, $\mathcal{R}$ investors have more impact on their abatement choices, which reduces the valuation gap with $\mathcal{N}$ investors and make these firms more attractive to $\mathcal{R}$ types. Hence, concentration may help $\mathcal{R}$ investors participate in the financial market. Since firms $j < N - n$ have instead a negative return net of the pollution disutility, $\mathcal{R}$ investors still choose to exclude these firms from their portfolios, even though the marginal cost of acquiring positions in additional firms is zero for them. Second, when firms' abatement choices are strategic substitutes, the crowding-out effect of concentrated equilibria also arises in this model extension, so that concentrated SRI may reduce aggregate abatement relative to the equilibrium without SRI.

B.6 Reverse timing

In our main model, firms choose their abatement levels after the investors trade. Here, we explore a setting where this timing is reversed – i.e., firms first choose their abatement investments to attract investors and maximize their stock prices. This version of the model applies best to settings where firms can make credible commitments to pollution abatement and managers care about their firms' stock prices.

The timing of the game is now as follows. At time $t = 1$, each firm j commits to σ_j to maximize its expected share price p_j. At the beginning of time $t = 2$, given the firms' abatement choices $\vec{\sigma}$, investors trade the firms' shares at their market clearing prices. Finally, profits realize and are distributed to shareholders.

Equilibrium Analysis. Since investors take $\vec{\sigma}$ as given, their optimal demands at $t = 2$ are the same as in our main model (see Eqn. (5)).[38] At $t = 1$, each firm j anticipates the equilibrium share price p_j that results from these demands, and solves

$$\max_{\sigma_j} \; p_j.$$

Let us focus for simplicity on the equilibria with $s^{\mathcal{R}}_{j\leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = 1$, for $n \in \{1, ..., N-1\}$. In these equilibria, market-clearing conditions give $p_{j\leq n} = \Pi_j - \frac{\kappa n}{1-\chi}$ and $p_{j>n} = \Pi_j - \frac{\kappa(N-n)}{\chi} - \lambda(1-\sigma_j)$, and the equilibrium abatement choices are as in the main model ($\sigma_{j\leq n} = \underline{\sigma} < \overline{\sigma} = \sigma_{j>n}$).

Since the investors' demands and market-clearing conditions are the same as in the main model, the conditions in Proposition 2 are still necessary for existence. However, here they are no longer sufficient: while $(\underline{\sigma}, \overline{\sigma})$ are *local* mutual-best-response, that is, holding fixed the investors' portfolio choices, j may have profitable *global* deviations that trigger different portfolio choices by investors. Therefore, equilibrium existence here also requires that all feasible global deviations are unprofitable. Namely, a firm $j \leq n$ ($j > n$) may deviate by attracting also, or only, $\mathcal{R}$ ($\mathcal{N}$) investors, or by attracting a larger mass of $\mathcal{N}$ ($\mathcal{R}$) investors.

When an individual investor is indifferent between the shares of a deviating firm j and those of some other firms, we assume that the investor breaks the indifference against j. This rules out the possibility that a deviating firm attracts only a fraction of the investors of another firm, limiting the number of possible global deviations to consider.[39] In particular, we can show that none of the global deviations are feasible starting from the best-in-class equilibrium ($n = N-1$), which implies the following.

Proposition B.5 (reverse timing). *If each firm j commits to σ_j to maximize its share price, the best-in-class equilibrium exists for $\chi \in (\underline{\chi}, \overline{\chi})$, where $\underline{\chi}$ and $\overline{\chi}$ are as in Proposition 2, and it may feature lower aggregate abatement relative to the coexisting symmetric equilibrium.*

Although firms are ex-ante identical, they may select different abatement policies in equilibrium: some firms abate more to attract $\mathcal{R}$ investors, while others abate less and cater to $\mathcal{N}$ investors. This differentiation makes it harder for each firm to deviate and

[38] The fact that here $\vec{\sigma}$ is observed, whereas in the main model it is correctly anticipated, plays no role on the equilibrium path (off-path events are immaterial to the analysis, given the continuum of investors). Moreover, in both models, investors take as given the market-clearing prices and correctly anticipate firms' expected profits given $\vec{\sigma}$.

[39] It is worth emphasizing that, even when a given deviation is possible, it is not necessarily profitable: holding Π_j fixed, attracting more (or a different type of) investors increases the share price of a deviating firm j. However, the deviation requires choosing a σ_j that is not a best-response to the other firms' investments, which may decrease Π_j and, as a result, p_j. It follows that the possibility result in Proposition B.5 is likely to hold even without this tie-breaking assumption, though the analysis would be more cumbersome.

attract both types of investors, since each equilibrium policy is tailored to the preferences of a specific investor type. As, for any given distribution of SRI, abatement policies are set as in the main model, the implications of SRI are the same under this alternative timing of the game.

B.7 Provision of public goods

In the main model, each firm j generates a negative externality λ at a rate $1-\sigma_j$. In such a setting, responsible agents have lower valuations than non-responsible ones for brown firms' shares and products. This implies that in equilibrium $\mathcal{N}$ investors can crowd out $\mathcal{R}$ investors, but not the other way around.

Here we show that our qualitative results are robust under the opposite assumption. Namely, suppose the status quo entails no externality, but firms' investments bring up a positive externality λ at a rate σ_j, which increases the utility of responsible agents. Therefore, in this setting, $\mathcal{R}$ investors have higher valuations for firms' shares than $\mathcal{N}$ investors, and so responsible investors can crowd out non-responsible ones from the market.

Denoting again by $\Pi_j = \Pi(\sigma_j, \vec{\sigma}_{-j})$ firm j's expected profit, which satisfies the same assumptions as in the base model, for a given vector of firms' investments $\vec{\sigma}$, investor i now solves:

$$\max_{s_{ij}\geq 0} \sum_j s_{ij}\Pi_j - p_j + \mathbb{1}_{i,\mathcal{R}}\lambda\sigma_j - K_i,$$

and, under the proportional control assumption, each firm j chooses its investment level solving:

$$\max_{\sigma_j\in[0,1]} \Pi_j + \lambda\sigma_j s_j^{\mathcal{R}}.$$

Equilibrium Analysis. Taking the first-order condition at the investment stage, it is straightforward to see that firm j's profit is still maximized for σ_j solving Eqn. (3).

Moving backward to the investment stage, we obtain the following shares' demand:

$$s_{ij} = \begin{cases} \max\{\frac{1}{\kappa}\left[\Pi_{j^*} - p_{j^*} + \mathbb{1}_{i,\mathcal{R}}\lambda\sigma_{j^*}\right], 0\} & \text{for } j^* = \text{argmax}_j\{\Pi_j - p_j + \mathbb{1}_{i,\mathcal{R}}\lambda\sigma_j\} \\ 0 & \text{for } j \neq j^*. \end{cases}$$

The market clearing conditions determine the equilibrium share prices

$$p_j = \Pi_j - \underbrace{\frac{\kappa}{\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}}}}_{\text{liquidity discount}} + \underbrace{\frac{\lambda\sigma_j\chi\alpha_j^{\mathcal{R}}}{\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}}}}_{\text{green premium}}, \tag{31}$$

where $\sum_j \alpha_j^{\mathcal{R}} = 1$ in all equilibria and $\sum_j \alpha_j^{\mathcal{N}} \in \{0,1\}$, as $\alpha_j^{\mathcal{N}} = 0$ for all j in an equilibrium where $\mathcal{N}$ investors are crowded out from the market.

The equilibrium characterization mirrors the one shown in the main model. To simplify the exposition, we focus on the analysis of symmetric equilibria and asymmetric equilibria where $s^{\mathcal{R}}_{j\leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = 1$, for $n \in \{1, ..., N-1\}$. Defining

$$\widehat{\chi}_R(\sigma, n) \equiv \frac{1}{2} - \frac{\sqrt{(N\kappa + \lambda\sigma)^2 - 4\kappa\lambda n\sigma} - N\kappa}{2\lambda\sigma}, \tag{32}$$

where the function $\widehat{\chi}_R(\cdot)$ is decreasing in σ, increasing in n, and such that $\widehat{\chi}_R(\cdot) \in (0, \frac{n}{N})$, the following results hold:

Proposition B.6 (green premium). *If each firm j's investments bring up a positive externality λ at a rate σ_j, internalized by $\mathcal{R}$ investors, the game admits the following two types of equilibria:*

- *(Symmetric equilibria.) An symmetric equilibrium always exists and, under Assumption 2, it is unique. It features exclusion of $\mathcal{N}$ investors for $\chi \geq \widehat{\chi}_R(\sigma_R, N)$, where σ_R is firms' equilibrium investment when $s^{\mathcal{R}}_j = 1$ for all $j \in \mathcal{J}$; and coexistence of $\mathcal{R}$ and $\mathcal{N}$ investors ($s^{\mathcal{R}*} \in (0,1)$ and so $\sigma^* < \sigma_R$ for all $j \in \mathcal{J}$) otherwise.*
- *(Asymmetric equilibria.) Equilibria where $s^{\mathcal{R}}_{j\leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = 1$ exist for $\widehat{\chi}_R(\overline{\sigma}, N-n) < \chi < \widehat{\chi}_R(\underline{\sigma}, N-n)$, where the investment levels $\sigma_{j\leq n} \equiv \underline{\sigma} < \overline{\sigma} \equiv \sigma_{j>n}$ are the same defined in Proposition 2. These equilibria may feature lower aggregate investments relative to the coexisting symmetric equilibrium.*

When $\mathcal{R}$ investors are too few to crowd out $\mathcal{N}$ investors from all firms, they can have a larger impact on the investments of firms in their portfolios by targeting a subset of firms. As a result, SRI concentration may still arise in equilibrium even when $\mathcal{R}$ investors have higher valuations compared to $\mathcal{N}$ investors.

In particular, as in the main model there exist equilibria where $s^{\mathcal{R}}_{j\leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = 1$ and firms choose investments levels $\sigma_{j\leq n} \equiv \underline{\sigma} < \overline{\sigma} \equiv \sigma_{j>n}$. The results of Proposition 4 then imply that, also in this version of the model, the best-in-class equilibrium may reduce aggregate investments relative to a world without SRI, hence a fortiori relative to the coexisting symmetric equilibrium with SRI (where $s^{\mathcal{R}*} > 0$ and so firms invest more).

B.8 Proofs for Appendix B

Proof of Proposition B.1. From Proposition 3, we know that in a neighborhood of χ_0: *(i)* the best-in-class equilibrium and the symmetric equilibrium coexist if $\lambda \geq \overline{\lambda}$ and firms' abatement policies are strategic substitutes or independent, and *(ii)* $\mathcal{R}$ investors strictly prefer the best-in-class equilibrium whereas $\mathcal{N}$ investors strictly prefer the symmetric equilibrium.

Recall that, in the symmetric equilibrium, each ($\mathcal{R}$ and $\mathcal{N}$) investor i selects each firm $j \in \{1, ..., N\}$ with probability $\frac{1}{N}$ – equivalently, each firm is selected by a fraction

$\frac{1}{N}$ of each type of investors – and submits a demand schedule $s_i^\theta = \max\{\Pi^* - p_j - \mathbb{1}_{i,\mathcal{R}}\lambda(1-\sigma^*), 0\}$ in the selected firm. By contrast, in the best-in-class equilibrium, each $\mathcal{R}$ investor selects (with probability 1) firm $j = N$ and submits a demand schedule $s_i^{\mathcal{R}} = \max\{\overline{\Pi} - p_N - \lambda(1-\overline{\sigma}), 0\}$, whereas each $\mathcal{N}$ investor selects each firm $j < N$ with probability $\frac{1}{N-1}$ and submits a demand schedule $s_i^{\mathcal{N}} = \max\{\underline{\Pi} - p_{j<N} - \lambda(1-\underline{\sigma}), 0\}$ in the selected firm.

Best-in-class equilibrium. As $\chi_0 \in (\underline{\chi}, \overline{\chi})$, in a neighborhood of χ_0, in the best-in-class equilibrium $\overline{\Pi} - \overline{p} - \lambda(1-\overline{\sigma}) > \max\{\underline{\Pi} - \underline{p} - \lambda(1-\underline{\sigma}), 0\}$ and $\overline{\Pi} - \overline{p} < \underline{\Pi} - \underline{p}$. Starting from the best-in-class equilibrium, consider a perturbed game where a mass $\widehat{\alpha}$ of $\mathcal{N}$ investors play the strategies they play in the symmetric equilibrium – i.e., a fraction $\frac{1}{N}$ of them submit a demand schedule $s_i^{\mathcal{N}} = \max\{\Pi^* - p_N, 0\}$ in firm $j = N$. Then, even though these $\mathcal{N}$ investors are able to purchase positive shares in firm N, since each σ_j (hence, also Π_j) is a continuous function of $s_N^{\mathcal{R}}$, for $\widehat{\alpha}$ sufficiently small it is still the case that $\Pi_N - p_N - \lambda(1-\sigma_N) > \max\{\Pi_{j<N} - p_{j<N} - \lambda(1-\sigma_{j<N}), 0\}$ and $\Pi_N - p_N < \Pi_{j<N} - p_{j<N}$. Therefore, all $\mathcal{R}$ investors still select firm $j = N$ and buy shares $s_i^{\mathcal{R}} = \max\{\Pi_N - p_N - \lambda(1-\sigma_N), 0\} = \Pi_N - p_N - \lambda(1-\sigma_N)$, whereas all non-committed $\mathcal{N}$ investors still select each firm $j < N$ with equal probability $\frac{1}{N-1}$ and buy shares $s_i^{\mathcal{N}} = \max\{\Pi_{j<N} - p_{j<N}, 0\} = \Pi_{j<N} - p_{j<N}$, these strategies converging to the equilibrium ones as $\widehat{\alpha} \to 0$, because $s_N^{\mathcal{R}} \to 1$ and abatement policies and profits are continuous in $s_N^{\mathcal{R}}$.

Moreover, Assumption 1 implies that, for $s_N^{\mathcal{R}}$ sufficiently large, $\sigma_N = 1$ and, since $s_{j<N}^{\mathcal{R}} = 0$ also in the perturbed game, $\sigma_{j<N} = \underline{\sigma}$ (and so also $\Pi_N = \overline{\Pi}$ and $\Pi_{j<N} = \underline{\Pi}$). As a consequence, there exists $\alpha' > 0$ such that, for all $\widehat{\alpha} \in (0, \alpha']$, the perturbed game admits an equilibrium with the same abatement policies (and firms' profits) of the best-in-class equilibrium.

Symmetric equilibrium. Starting from the symmetric equilibrium, consider a perturbation of the original game where a mass $\widehat{\alpha}$ of $\mathcal{R}$ investors play the strategies they play in the best-in-class equilibrium – i.e., they all submit a demand schedule $s_i^{\mathcal{R}} = \max\{\overline{\Pi} - p_N - \lambda(1-\overline{\sigma}), 0\}$ in firm $j = N$. In order for the symmetric equilibrium to be robust to such perturbation, the probability with which each non-committed $\mathcal{R}$ and $\mathcal{N}$ investor selects each firm should smoothly converge to $\frac{1}{N}$ as $\widehat{\alpha} \to 0$. For this to hold, all these investors should be indifferent among all firms in the perturbed game.

(a) Equilibrium with SRI. In order for the perturbed game to admit an symmetric equilibrium which features SRI, it must be that $\sigma_{j<N} = \sigma_N \equiv \sigma$ in order for both types of investors' indifference conditions to hold, which implies that $\Pi_{j<N} = \Pi_N \equiv \Pi$; then, the indifference conditions imply that $p_{j<N} = p_{j=N} \equiv p$.

Let $\alpha^{\mathcal{R}}$ (resp. $\alpha^{\mathcal{N}}$) denote the probability with which each non-committed $\mathcal{R}$ (resp. $\mathcal{N}$) investor selects firm $j = N$. Then, the following market-clearing conditions for firms $j < N$ and $j = N$ must hold, together with the condition $s_{j<N}^{\mathcal{R}} = s_N^{\mathcal{R}}$, to ensure that firms

choose the same abatement policy:

$$\begin{cases} (1-\widehat{\alpha})\frac{1-\alpha^{\mathcal{R}}}{N-1}\chi\frac{1}{\kappa}\max\{\Pi-p-\lambda(1-\sigma),0\}+\frac{1-\alpha^{\mathcal{N}}}{N-1}(1-\chi)\frac{1}{\kappa}\max\{\Pi-p,0\}=1, \\ \widehat{\alpha}\chi\frac{1}{\kappa}\max\{\overline{\Pi}-p-\lambda(1-\overline{\sigma}),0\}+(1-\widehat{\alpha})\alpha^{\mathcal{R}}\chi\frac{1}{\kappa}\max\{\Pi-p-\lambda(1-\sigma),0\}+\alpha^{\mathcal{N}}(1-\chi)\frac{1}{\kappa}\max\{\Pi-p,0\}=1, \\ \alpha^{\mathcal{N}}(1-\chi)\frac{1}{\kappa}\max\{\Pi-p,0\}=\frac{1-\alpha^{\mathcal{N}}}{N-1}(1-\chi)\frac{1}{\kappa}\max\{\Pi-p,0\}. \end{cases}$$

If firms' abatement policies are independent it is always the case that $\overline{\Pi}-\lambda(1-\overline{\sigma}) > \Pi_N-\lambda(1-\sigma_N)$. This is because $\sigma_N = \text{argmax}_\sigma\, \Pi_N - s_N^{\mathcal{R}}\lambda(1-\sigma)$ and so the preference-adjusted gross return $\Pi_N-\lambda(1-\sigma_N)$ of $\mathcal{R}$ investors is maximized when $s_N^{\mathcal{R}}=1$. Under substitutability, this result a fortiori holds because $s_N^{\mathcal{R}}=1$ dampens rivals' abatement levels, further increasing $\mathcal{R}$ investors' preference-adjusted gross return in firm $j=N$. Therefore, whenever non-committed $\mathcal{R}$ investors buy positive shares, committed $\mathcal{R}$ investors a fortiori do so. Of course, also $\mathcal{N}$ investors have higher demand and so they also end up buying positive shares in this candidate equilibrium.

Then, supposing that all demands are strictly positive, solving the system for $(\alpha^{\mathcal{R}}, \alpha^{\mathcal{N}}, p)$ yields the candidate equilibrium values

$$\alpha^{\mathcal{R}} = \frac{1}{(1-\widehat{\alpha})N}\left[1-\widehat{\alpha}N+\frac{\widehat{\alpha}(N-1)[\Pi-\overline{\Pi}-\lambda(\overline{\sigma}-\sigma)]}{\lambda[(1-(1-\widehat{\alpha})\sigma-\widehat{\alpha}\overline{\sigma})\chi-(1-\sigma)]+\widehat{\alpha}\chi(\Pi-\overline{\Pi})+N\kappa}\right],$$

$\alpha^{\mathcal{N}}=\frac{1}{N}$, and $p=\Pi+\widehat{\alpha}\chi(\overline{\Pi}-\Pi)-N\kappa-\lambda\chi[1-\sigma-\widehat{\alpha}(\overline{\sigma}-\sigma)]$.

For $\widehat{\alpha}\to 0$, this solution is not admissible in a neighborhood of χ_0. Indeed, as $\widehat{\alpha}\to 0$, firms' abatement policies should converge to σ_0 at $\chi=\chi_0$ (as $s_j^{\mathcal{R}}=0$ for all j in the symmetric equilibrium of the original game) and then we would have

$$\alpha^{\mathcal{R}}|_{\chi\to\chi_0} = 1+\frac{\kappa(N-1)}{\lambda(1-\sigma_0)-N\kappa} > 1,$$

where the inequality follows because $\lambda(1-\sigma_0)-N\kappa>0$ is equivalent to $\chi_0<1$. Moreover, as $\alpha^{\mathcal{R}}|_{\chi\to\chi_0}$ is bounded away from one, this result implies that there exists a threshold $\alpha''>0$ such that the considered candidate equilibrium does not exist in the perturbed game for all $\widehat{\alpha}\in(0,\alpha'']$.

(b) Equilibrium without SRI. Alternatively, it is possible that non-committed $\mathcal{R}$ investors end up buying zero shares in all firms.[40] In this case, the symmetric equilibrium is not robust to the perturbation if committed $\mathcal{R}$ investors are able to buy positive shares in firm $j=N$. Indeed, in this case, $s_N^{\mathcal{R}}>0=s_{j<N}^{\mathcal{R}}$, and so $\sigma_N>\sigma_{j<N}$. As $\Pi_N-p_N=\Pi_{j<N}-p_{j<N}$ by $\mathcal{N}$ investors' indifference condition,[41] this implies that non-committed $\mathcal{R}$

[40] If they buy positive shares only in a subset of firms, they strictly prefer these firms to the others and so the probability of selecting each firm cannot smoothly converge to $\frac{1}{N}$ as $\widehat{\alpha}\to 0$, implying that the symmetric equilibrium is not robust to the perturbation.

[41] If $\mathcal{N}$ investors were to prefer firms $j<N$ to firm N, $s_N^{\mathcal{R}}=1$ and so non-committed $\mathcal{R}$ investors would also enter firm N.

investors strictly prefer firm N to firms $j < N$, hence their strategies are discontinuous as $\widehat{\alpha} \to 0$.

In the candidate equilibrium under consideration, only the committed $\mathcal{R}$ investors potentially buy shares, and so the market-clearing conditions for firms $j < N$ and $j = N$ and the indifference condition of $\mathcal{N}$ investors are, respectively,

$$\begin{cases} \frac{1-\alpha^{\mathcal{N}}}{N-1}(1-\chi)\frac{1}{\kappa}\max\{\Pi_{j<N} - p_{j<N}, 0\} = 1, \\ \widehat{\alpha}\chi\frac{1}{\kappa}\max\{\overline{\Pi} - p_N - \lambda(1-\overline{\sigma}), 0\} + \alpha^{\mathcal{N}}(1-\chi)\frac{1}{\kappa}\max\{\Pi_N - p_N, 0\} = 1, \\ \Pi_{j<N} - p_{j<N} = \Pi_N - p_N. \end{cases}$$

The previous argument implies that, to prove that the symmetric equilibrium is not robust to the perturbation, we only need to show that there exists a solution to the above system such that $\overline{\Pi} - p_N - \lambda(1-\overline{\sigma}) > 0$ – i.e., the perturbed game admits an equilibrium with (at least) committed SRI, rather than a symmetric equilibrium without SRI. Supposing that this condition holds, the solution to the above system is such that

$$\overline{\Pi} - p_N - \lambda(1-\overline{\sigma}) = \frac{N\kappa - (1-\chi)(\lambda(1-\overline{\sigma}) + \Pi_N - \overline{\Pi})}{1-(1-\widehat{\alpha})\chi},$$

which, at $\chi = \chi_0$, for $\widehat{\alpha} \to 0$ (so that $\Pi_N \to \Pi_0$) equals $\overline{\Pi} - \lambda(1-\overline{\sigma}) - [\Pi_0 - \lambda(1-\sigma_0)] > 0$ (by the arguments above, this inequality always holds under independence or substitutability), so that committed $\mathcal{R}$ investors cannot be crowded out. Moreover, as this value is bounded away from zero, this result implies that there exists a threshold $\alpha''' \in (0,1]$ such that the perturbed game does not admit an equilibrium without SRI for all $\widehat{\alpha} \in (0, \widehat{\alpha}''']$.

Taken together, the above results show that, in a neighborhood of χ_0, (1) as $\widehat{\alpha} \to 0$, non-committed investors' strategies are continuous in the best-in-class equilibrium but they are not in the symmetric equilibrium, and (2) for all $\widehat{\alpha} \in (0, \overline{\alpha}]$, with $\overline{\alpha} \equiv \min\{\alpha', \alpha'', \alpha'''\}$, the perturbed game admits an equilibrium where $(\sigma_{j<N} = \underline{\sigma}, \sigma_N = 1)$ but not a symmetric equilibrium and $\sigma_j \equiv \sigma$ for all j.[42] □

Proof of Proposition B.2. Since investors correctly anticipate $\vec{\sigma}$, their optimal demands at $t = 1$ are the same as in our main model (Eqn. 5). We next characterize the firm choice of σ_j. In the asymmetric equilibrium with full separation, each firm $j > n$ maximizes $\Pi_j - \lambda(1-\sigma_j)$, as $s^{\mathcal{R}}_{j>n} = 1$ implies that the proposal is accepted for all $\tau \in (0,1)$, and firm $j \leq n$ maximizes Π_j, as $s^{\mathcal{R}}_{j\leq n} = 0$ implies that the proposal is always rejected. Since this is identical to our main model, these two types of firms choose the

[42] Indeed, in order for all firms to choose the same abatement policy, $s^{\mathcal{R}}_j$ should be the same for all j. As committed $\mathcal{R}$ investors buy positive shares in firm N for all $\widehat{\alpha} \in (0, \alpha''']$, this is only possible if the candidate equilibrium with SRI described in part (a) above exists, which is not true for all $\widehat{\alpha} \in (0, \alpha'']$.

same abatement efforts $(\underline{\sigma},\overline{\sigma})$ defined in Proposition 2. Moreover, since each individual investor is atomistic and does not change the voting outcome, we obtain the same existence condition $\widehat{\chi}(\overline{\sigma},n)<\chi<\widehat{\chi}(\underline{\sigma},n)$ as in the main model.

In the symmetric equilibrium, we have to differentiate three cases at $t=2$: (i) $s^{\mathcal{R}}=0$ and $\sigma^*=\sigma_0$, (ii) $s^{\mathcal{R}}<\tau$ and so again $\sigma^*=\sigma_0$, and (iii) $s^{\mathcal{R}}\geq\tau$ and so $\sigma^*=\arg\max_{\sigma_j}\Pi_j(\sigma_j,\sigma^*\cdot\vec{1})-\lambda(1-\sigma_j)>\sigma_0$. The first case is identical to our main model so that the symmetric equilibrium without SRI exists for all $\chi<\widehat{\chi}(\sigma_0,N)$. As, in any symmetric equilibrium with SRI, $s^{\mathcal{R}*}$ is given by Eqn. (8), case (ii) requires that:

$$\tau>\chi-\frac{\lambda(1-\chi)\chi}{N\kappa}(1-\sigma_0)\Leftrightarrow\chi<\widetilde{\chi}(\sigma_0),$$

where $\widetilde{\chi}(\sigma_0)>\widehat{\chi}(\sigma_0,N)$.[43] Case (iii) requires that:

$$\tau\leq\chi-\frac{\lambda(1-\chi)\chi}{N\kappa}(1-\sigma^*)\Leftrightarrow\chi\geq\widetilde{\chi}(\sigma^*),$$

where σ^* solves the first-order condition $\mathcal{C}'(\sigma^*)\geq\frac{\partial\pi_j}{\partial\sigma_j}+\lambda$, and $\widetilde{\chi}(\sigma^*)\in(\widehat{\chi}(\sigma_0,N),\widetilde{\chi}(\sigma_0))$.[44] Thus, for $\chi\in[\widetilde{\chi}(\sigma^*),\widetilde{\chi}(\sigma_0))$ the game admits two symmetric equilibria, both of them featuring SRI.

Next, we show that, for some parameter configurations, the only equilibria with SRI are asymmetric, and aggregate abatement is lower than in the benchmark without SRI. For simplicity, we consider the same functional form for π_j and parameters as in Figure 4. Note that the existence threshold $\widehat{\chi}(\sigma,n)$, and the equilibrium CSR investments σ_0, $\overline{\sigma}$, and $\underline{\sigma}$ are the same as in the main model. Therefore, as shown in Figure 4, in a range of the parameters, the best-in-class equilibrium coexists with the symmetric equilibrium with no SRI, and aggregate abatement is lower in the former equilibrium. □

Proof of Proposition B.3. Given the shares' demand in Eqn. (28), for the market-clearing conditions to hold for all firms, it must be that Π_j-p_j is positive and constant across firms – i.e., in any equilibrium, all investors must be indifferent between all firms'

[43]Indeed,

$$\widetilde{\chi}(\sigma_0)-\widehat{\chi}(\sigma_0,N)=\frac{\sqrt{(\lambda(1-\sigma)-N\kappa)^2+4\kappa\lambda N(1-\sigma)\tau}-(\lambda(1-\sigma)-N\kappa)}{2\lambda(1-\sigma)}>0.$$

[44]The result $\widetilde{\chi}(\sigma^*)<\widetilde{\chi}(\sigma_0)$ follows from $\sigma^*>\sigma_0$ and $\widetilde{\chi}(\cdot)$ being decreasing in σ:

$$\frac{\partial\widetilde{\chi}(\cdot)}{\partial\sigma}\propto N\kappa-\lambda(1-2\tau)(1-\sigma)-\sqrt{(\lambda(1-\sigma)-\kappa N)^2+4\kappa\lambda N(1-\sigma)\tau}<0;$$

whereas $\widetilde{\chi}(\sigma^*)>\widehat{\chi}(\sigma_0,N)$ because, for all σ, $\widetilde{\chi}(\sigma)|_{\tau=0}=\widehat{\chi}(\sigma_0,N)$ and $\widetilde{\chi}(\cdot)$ is increasing in τ.

shares. The market-clearing conditions then write as

$$\left[\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}}\right]\frac{1}{\kappa}(\Pi_j - p_j) = 1 \quad \forall j \in \mathcal{J}.$$

Summing across firms, and using the fact that $\Pi_j - p_j$ does not depend on j, yields $\frac{1}{\kappa}(\Pi_j - p_j) = N$ and so

$$\chi\alpha_j^{\mathcal{R}} + (1-\chi)\alpha_j^{\mathcal{N}} = \frac{1}{N}. \tag{33}$$

Any pair $(\alpha_j^{\mathcal{R}}, \alpha_j^{\mathcal{N}}) \in [0,1]^2$ that satisfies Eqn. (33), provided $\sum_{j\in\mathcal{J}}\alpha_j^{\mathcal{R}} = \sum_{j\in\mathcal{J}}\alpha_j^{\mathcal{N}} = 1$, yields an equilibrium of the game. Therefore, the game admits a continuum of equilibria.

In any equilibrium, firms' abatement efforts are determined as in the base model. Indeed, under the proportional control assumption, firm j solves

$$\max_{\sigma_j\in[0,1]} \Pi_j - s_j^{\mathcal{R}}\sum_{k\in\mathcal{J}}\lambda(1-\sigma_k),$$

which yields the same policies of the base model, where, given the market-clearing conditions, $s_j^{\mathcal{R}} = \chi N\alpha_j^{\mathcal{R}}$.

Therefore, as in the main model, the game admits:

- A unique symmetric equilibrium, which here is obtained for $\alpha_j^{\mathcal{R}} = \alpha_j^{\mathcal{N}} = \frac{1}{N}$ for all j, where all firms have the same ownership and choose the same abatement efforts (and accordingly make the same expected profits and have the same shares' prices);
- Multiple (in this case, infinitely many) asymmetric equilibria, in which firms more targeted by $\mathcal{R}$ investors abate more compared to firms that feature less prevalence of responsible capital (and, accordingly, firms make different profits and have different shares' prices).

In particular, there may be equilibria where $\mathcal{R}$ investors concentrate in a subset of firms. As in the base model, consider $\alpha_{j\leq n}^{\mathcal{R}} = 0$ and $\alpha_{j>n}^{\mathcal{R}} = \frac{1}{N-n}$. Then, the market clearing conditions imply that $\alpha_{j\leq n}^{\mathcal{N}} = \frac{1}{N(1-\chi)}$ and $\alpha_{j>N}^{\mathcal{N}} = \frac{N(1-\chi)-n}{(N-n)N(1-\chi)}$. Note that for $\chi = 1 - \frac{n}{N}$, $\alpha_{j>n}^{\mathcal{N}} = 0$, so that we obtain the equilibria with full separation characterized in the main model. The result in Proposition 4 then implies that aggregate abatement may be lower in this equilibrium than in the scenario with no SRI ($\alpha_j^{\mathcal{R}} = s_j^{\mathcal{R}} = 0\,\forall j$). In turn, aggregate abatement is lower in the no-SRI benchmark than in the symmetric equilibrium with SRI (where $\alpha_j^{\mathcal{R}} = \frac{1}{N}$ and so $s_j^{\mathcal{R}} > 0$ for all j), which here exists for all values of the parameters.[45] □

[45] Note that, by continuity of investors' and firms' equilibrium strategies, the possibility result that concentrated SRI may reduce aggregate abatement relative to no-SRI and, a fortiori, to the symmetric equilibrium with SRI, extends to a non-zero measure set of the parameters.

Proof of Proposition B.4. As we have argued that $\mathcal{N}$ investors always demand positive shares in all firms, and $\mathcal{R}$ investors are symmetric, it follows that any candidate equilibrium is characterized by the subset of firms targeted by $\mathcal{R}$ investors. Without loss of generality, let these firms be $j > n$.

For any given n, the corresponding equilibrium exists when $\mathcal{R}$ investors have a positive net payoff from buying shares in firms $j > n$ only. That is, when

$$\Pi_{j>n} - p_{j>n} - \lambda(1 - \sigma_{j>n}) = \kappa - (1-\chi)\lambda(1-\sigma_{j>n}) > 0 \Leftrightarrow \chi > 1 - \frac{\kappa}{\lambda(1-\sigma_{j>n})}$$

and

$$\Pi_{j\leq n} - p_{j\leq n} - \lambda(1 - \sigma_{j\leq n}) = \frac{\kappa}{1-\chi} - \lambda(1-\sigma_{j\leq n}) \leq 0 \Leftrightarrow \chi \leq 1 - \frac{\kappa}{\lambda(1-\sigma_{j\leq n})},$$

where the equalities follow from the market-clearing conditions

$$\frac{\chi}{\kappa}\left[\Pi_{j>n} - p_{j>n} - \lambda(1-\sigma_{j>n})\right] + \frac{1-\chi}{\kappa}\left[\Pi_{j>n} - p_{j>n}\right] = 1 \quad \text{and} \quad \frac{1-\chi}{\kappa}\left[\Pi_{j\leq n} - p_{j\leq n}\right] = 1,$$

and the equilibrium abatement investments $\sigma_{j\leq n} \equiv \underline{\sigma}(n) < \overline{\sigma}(n) \equiv \sigma_{j>n}$ are obtained from Eqn. (3), with $s^{\mathcal{R}}_{j\leq n} = 0$ and $s^{\mathcal{R}}_{j>n} = \chi\left(1 - \frac{(1-\chi)\lambda(1-\overline{\sigma}(n))}{\kappa}\right)$. The equilibrium existence conditions then can be written as in Eqn. (30). □

Proof of Proposition B.5. We consider the best-in-class equilibrium, which (as argued in the text) features the same abatement levels $(\underline{\sigma}, \overline{\sigma})$ characterized in Proposition 2. A necessary condition for this equilibrium to exist is $\widehat{\chi}(\overline{\sigma}, 1) \equiv \underline{\chi} < \chi < \overline{\chi} \equiv \widehat{\chi}(\underline{\sigma}, 1)$. In what follows, we show that, provided that all market-clearing conditions always hold (on- and off-path)[46] and that a deviating firm is unable to target a specific fraction of $\mathcal{N}$ or $\mathcal{R}$ investors, global deviations cannot be implemented, and so this condition is also sufficient for existence of the best-in-class equilibrium. We use the superscript d to denote *deviation* outcomes.

First, consider deviations by the targeted firm $j = N$, which should be held only by $\mathcal{R}$ investors in equilibrium. Suppose it tries to attract *only* $\mathcal{N}$ investors. Then, its market-clearing price would be

$$p^d_N = \Pi^d_N - \frac{\kappa}{1-\chi}.$$

Following the deviation, market clearing implies that $\mathcal{R}$ investors equally distribute their demand among the other firms, so we have $p^d_{j<N} = \Pi^d_{j<N} - \frac{(N-1)\kappa}{\chi} - \lambda(1-\underline{\sigma})$. $\mathcal{N}$ investors

[46] Note that the market-clearing price of any firm j in Eqn. (7) would be unbounded below if there was no demand for j's shares (i.e., $p_j \to -\infty$ if $\alpha^\theta_j \to 0$ for all $\theta \in \{\mathcal{N}, \mathcal{R}\}$), which would lead any investor to buy j's shares. Thus, market-clearing conditions rule out deviations where a single firms tries to attract all investors.

then prefer holding firm N to any firm $j < N$ if $\Pi_N^d - p_N^d > \Pi_{j<N}^d - p_{j<N}^d$, which is equivalent to $\frac{\kappa}{1-\chi} > \frac{(N-1)\kappa}{\chi} + \lambda(1-\underline{\sigma})$. This condition holds for

$$\chi > \frac{1}{2} + \frac{\sqrt{\lambda^2(1-\underline{\sigma})^2 + \kappa\left(\kappa N^2 + 2\lambda(N-2)(1-\underline{\sigma})\right)} - N\kappa}{2\lambda(1-\underline{\sigma})}.$$

However, as this threshold is larger than $\overline{\chi}$, this deviation is not feasible in the region of parameters where the best-in-class equilibrium can exist.

Second, consider deviations by an excluded firm, say (without loss of generality) firm 1, which should be held by $\mathcal{N}$ investors only in equilibrium. Suppose firm 1 tries to attract *only* $\mathcal{R}$ investors. Then, its market-clearing price is

$$p_1^d = \Pi_1^d - \frac{\kappa}{\chi} - \lambda(1-\sigma_1^d).$$

Following the deviation, the market-clearing conditions imply that $\mathcal{N}$ investors allocate their demand equally among the other firms, and so $\Pi_N^d - p_N^d = \Pi_{j\in[2,N-1]}^d - p_{j\in[2,N-1]}^d = \frac{(N-1)\kappa}{1-\chi}$. $\mathcal{R}$ investors prefer firm 1 to firm N if $\Pi_1^d - p_1^d - \lambda(1-\sigma_1^d) > \Pi_N^d - p_N^d - \lambda(1-\overline{\sigma})$, which, given the market clearing prices, is equivalent to $\frac{\kappa}{\chi} > \frac{(N-1)\kappa}{1-\chi} - \lambda(1-\overline{\sigma})$. This inequality boils down to $\chi < \underline{\chi}$. Therefore, this global deviation is not feasible whenever the best-in-class equilibrium can exist.

Finally, suppose that an excluded firm, say $j = 1$, tries to attract *all* the $\mathcal{N}$ investors (instead of just a share $\frac{1}{N-1}$ of them). Then, by the market-clearing conditions it follows that $\mathcal{R}$ investors equally distribute their demand among the other firms, and firm 1's market-clearing price is

$$p_1^d = \Pi_1^d - \frac{\kappa}{1-\chi}.$$

Following the deviation, $\mathcal{N}$ investors prefer firm 1 to any firm $j \in [2, N-1]$ if $\Pi_1^d - p_1^d = \frac{\kappa}{1-\chi} > \Pi_{j\in[2,N-1]}^d - p_{j\in[2,N-1]}^d = \frac{(N-1)\kappa}{\chi} + \lambda(1-\underline{\sigma})$, which is equivalent to

$$\chi > \frac{1}{2} + \frac{\sqrt{\lambda^2(1-\underline{\sigma})^2 + \kappa\left(\kappa N^2 + 2\lambda(N-2)(1-\underline{\sigma})\right)} - \kappa N}{2\lambda(1-\underline{\sigma})}.$$

However, as this threshold is larger than $\overline{\chi}$, also this deviation is not feasible whenever the best-in-class equilibrium can exist.

As we have ruled out all possible global deviations, we can conclude that, also under the alternative timing considered in this section, the best-in-class equilibrium exists for all $\underline{\chi} < \chi < \overline{\chi}$. From the main analysis, we know that this equilibrium may feature lower aggregate abatement relative to the equilibrium without SRI. □

Proof of Proposition B.6. Assume Assumption 2 and consider, first, an equilibrium where $\mathcal{N}$ investors are crowded out. This equilibrium exists if and only if only $\mathcal{R}$ investors are willing to trade shares in all firms. That is, letting σ_R, Π_R, and p_R denote firms' investments, expected profits, and share prices in this candidate equilibrium, it must be

$$\Pi_R - p_R \leq 0 < \Pi_R - p_R + \lambda\sigma_R.$$

Substituting the market clearing price from Eqn. (31) for $\alpha_j^{\mathcal{R}} = \frac{1}{N}$ and $\alpha_j^{\mathcal{N}} = 0$, these condition boil down to $\chi \geq \frac{N\kappa}{\lambda\sigma_R} = \widehat{\chi}_R(\sigma_R, N)$, where σ_R is obtained as the unique solution of Eqn. (3) for $s_j^{\mathcal{R}} = 1$. Then, proceeding as in the proof of Proposition 1, it can be shown that, for all $\chi < \widehat{\chi}_R(\sigma_R, N)$, there is a unique symmetric equilibrium where, for all $j \in \mathcal{J}$, $\sigma_j = \sigma^* < \sigma_R$ is obtained from Eqn. (3) for $s_j^{\mathcal{R}} = s^{\mathcal{R}*} = \frac{\chi}{N\kappa}\left[N\kappa + \lambda(1-\chi)\sigma^*\right] \in (0,1)$.

Consider asymmetric equilibria where $s_{j\leq n}^{\mathcal{R}} = 0$ and $s_{j>n}^{\mathcal{R}} = 1$, for $n \in \{1, ..., N-1\}$. As argued in the text, these equilibria feature the same firms' investments as the main model, i.e., $\sigma_{j\leq n} \equiv \underline{\sigma} < \overline{\sigma} \equiv \sigma_{j>n}$ characterized in Proposition 2. Therefore, provided these equilibria exist, they may yield lower aggregate investments than a benchmark with no SRI (where $\sigma_j = \sigma_0\ \forall j \in \mathcal{J}$) and hence, a fortiori, relative to the coexisting symmetric equilibrium, which always features SRI (and so $\sigma^* > \sigma_0$).

As seen in the main model, these equilibria exist if $\mathcal{N}$ (resp. $\mathcal{R}$) investors prefer holding any firm $j \leq n$ (resp. $j > n$) to any other firm and to buying no shares at all. Similar to the proof of Proposition 2, using the market clearing conditions, these conditions boil down to

$$\frac{(N-n)\kappa}{\chi} > \frac{n\kappa}{1-\chi} + \lambda\underline{\sigma},$$

and

$$\frac{n\kappa}{1-\chi} > \frac{(N-n)\kappa}{\chi} + \lambda\overline{\sigma}.$$

Putting these conditions together finally yields $\widehat{\chi}_R(\overline{\sigma}, N-n) < \chi < \widehat{\chi}_R(\underline{\sigma}, N-n)$, where the threshold $\widehat{\chi}_R(\cdot)$ is defined in Eqn. (32). $\square$