Boston University

OpenBU **http://open.bu.edu**

Boston University Theses & Dissertations Boston University Theses & Dissertations

2026

Essays on emerging technologies and sustainabilit

https://hdl.handle.net/2144/52799

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY

QUESTROM SCHOOL OF BUSINESS

Dissertation

ESSAYS ON EMERGING TECHNOLOGIES AND SUSTAINABILITY

by

YUZE LI

B.A.Sc., University of Toronto, 2018
M.S., Chinese Academy of Sciences, 2021

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2026

© 2026 by
YUZE LI
All rights reserved

Approved by

First Reader

Marcus Bellamy, PhD
Associate Professor of Operations and Technology Management

Second Reader

Pnina Feldman, PhD
Bigelow Research Associate Professor in Business Administration
University of Virginia

Third Reader

Gerry Tsoukalas, PhD
Dean's Research Scholar
Professor of Information Systems

Acknowledgments

I owe a great deal to the people who have supported me through this journey.

I am grateful to Gerry for being a mentor, a co-author, and a great friend. His guidance and generosity have shaped who I am as a researcher, and this dissertation would not exist without him. I also want to thank Jussi, Marcus, Pnina, and Shouyang for their thoughtful feedback, encouragement, and collaboration.

None of this would have been possible without my parents. They left behind everything they knew and immigrated to give me a better life. The opportunities I have had—including the chance to pursue this degree—are the direct result of their courage and sacrifice.

I am grateful to Feifei for her love, patience, and for always being there. Through the long days and late nights, she kept me grounded and reminded me of what matters.

To Ben and Shangrong, for being true friends through all of it. To my cohorts, who understood the weight of this journey because they carried it too. And to my friends and everyone who has walked alongside me at different points of this journey—thank you.

This dissertation is dedicated to the loving memory of my grandma, who I lost during this journey. I wish she could have seen this. I know she would have been proud.

Yuze Li

ESSAYS ON EMERGING TECHNOLOGIES AND SUSTAINABILITY

YUZE LI

Boston University, Questrom School of Business, 2026

Major Professors: Marcus Bellamy, PhD
Associate Professor of Operations and Technology Management

Pnina Feldman, PhD
Bigelow Research Associate Professor in Business Administration
University of Virginia

Gerry Tsoukalas, PhD
Dean's Research Scholar
Professor of Information Systems

ABSTRACT

This dissertation investigates how emerging technologies such as artificial intelligence (AI) and decentralized platforms both address and generate new sustainability challenges across environmental, social, and governance dimensions. In the first chapter, I investigate how the rapid adoption of AI-driven pricing agents affects the sustainability of tacit collusion in markets. We show that collusion is fragile under the heterogeneity typical of real deployments: differences in agent patience, data access, number of competitors, and algorithmic diversity all significantly weaken collusion, while model-size differences paradoxically stabilize it through emergent leader-follower dynamics. These findings highlight agent heterogeneity as a critical factor

shaping competitive dynamics, providing policymakers with actionable levers to mitigate collusion risks. In the second chapter, I examine how decentralized consortia for certifying green products impact firms' sustainability incentives. We show that while joining a consortium increases firm profits by unlocking the green premium, it can paradoxically reduce firms' sustainability efforts when consumer willingness to pay for green products is relatively low, ultimately harming consumer surplus and social welfare. In the final chapter, I examine the impact of blockchain technology on financial inclusion in developing countries. Exploiting the implementation of a blockchain-based lending protocol by the prosocial lending platform Kiva in Sierra Leone, we find that borrowers attract more guarantors and are more likely to be funded, while microfinance institutions experience lower portfolio risks and improved operational sustainability. These positive impacts are most pronounced for traditionally underserved borrowers, including rural populations and individuals with weaker financial credit records.

Contents

List of Tables

List of Figures

List of Abbreviations

AI	....	Artificial Intelligence
DID	....	Difference-in-Differences
DL	....	Distributed Ledger
DOJ	....	Department of Justice (U.S.)
GPT	....	Generative Pre-trained Transformer
GRPO	....	Group Relative Policy Optimization
ICO	....	Initial Coin Offering
LLM	....	Large Language Model
MFI	....	Microfinance Institution
MNL	....	Multinomial Logit
MSC	....	Marine Stewardship Council
NDIP	....	National Digital Identity Platform
RL	....	Reinforcement Learning
SD	....	Standard Deviation
USDA	....	United States Department of Agriculture

Chapter 1

Introduction

Emerging technologies such as artificial intelligence (AI) and decentralized platforms are rapidly transforming markets and societies, fundamentally reshaping both the sustainability challenges we face and the strategies we develop to solve them. Motivated by real-world managerial problems and existing gaps in academic research, in this dissertation, I seek to understand how emerging technologies both address and generate new sustainability challenges across environmental, social, and governance dimensions, and to develop effective strategies and policies to navigate these issues. To investigate these issues, I draw on a diverse toolkit spanning causal inference, microeconomics, and game theory, enhanced by the integration of machine learning techniques.

In the first chapter, I investigate how the rapid adoption of AI-driven pricing agents affects the sustainability of tacit collusion in markets. As firms increasingly deploy large language models (LLMs) to support or fully automate pricing decisions, regulators have warned that algorithmic pricing could facilitate fully automated cartels operating without any human involvement, undermining competitive markets. While recent studies show that similarly designed LLM agents can sustain supra-competitive prices, real-world deployments exhibit far greater diversity across model architecture, data access, and objectives, making it unclear whether collusion survives under such heterogeneity. We combine a game-theoretic model with large-scale laboratory experiments using open-source LLM pricing agents to systematically examine this question. We find that differences in agent patience, data access, number of competitors, and algorithmic diversity all significantly weaken or eliminate collusion, while model-size differences paradoxically stabilize it through emergent leader-follower dynamics. These findings highlight agent heterogeneity as a critical but previously overlooked factor, providing policymakers with actionable levers such as data-sharing restrictions and algorithmic diversity promotion to mitigate collusion risks.

In the second chapter, I examine how decentralized consortia for certifying green products impact firms' sustainability incentives. As markets experience a rapid influx of sustainable products, consumers face growing uncertainty about the authenticity of green claims; in response, decentralized consortia such as the Aura Consortium, initiated by global luxury brands including LVMH, Prada, and Mercedes-Benz, have emerged, leveraging blockchain-based peer verification to certify product sustainability. We develop a game-theoretic model comparing firms' sustainability efforts, pricing, and welfare outcomes on and off a decentralized consortium. We show that while joining a consortium always increases firm profits by unlocking the green premium eco-conscious consumers are willing to pay, it can paradoxically reduce firms' sustainability efforts, as the risk of failing certification deters investment in genuine sustainability, particularly when consumer willingness to pay for green is low. This incentive misalignment can ultimately harm consumer surplus and reduce social welfare, revealing that decentralized certification acts as a double-edged sword. These insights provide policymakers with guidance on identifying market conditions under which decentralized consortia can genuinely improve sustainability outcomes.

In the final chapter, I examine the impact of blockchain technology on financial inclusion in developing countries. We leverage a large-scale implementation of a blockchain-based lending protocol by the online prosocial lending platform Kiva in Sierra Leone as a natural experiment, conducting difference-in-differences analyses to evaluate the protocol's impact on borrower and lender outcomes. We find that borrowers attract more guarantors and larger per-guarantor contributions under the blockchain protocol, substantially increasing their likelihood of securing affordable microloans. At the lender level, microfinance institutions experience lower portfolio risks, with simultaneous growth in financial revenue and reduction in operational expenses, significantly enhancing the operational sustainability of their lending services. More importantly, these positive impacts are most pronounced for traditionally underserved borrowers, including rural populations and individuals with weaker financial credit records. Taken together, these findings provide robust evidence highlighting the impact of decentralized technology on financial inclusion, offering policymakers clear justification for supporting and scaling decentralized initiatives as a pathway toward greater social sustainability and economic equity.

Chapter 2

On the Fragility of AI Agent Collusion

2.1 Introduction

Firms are increasingly delegating pricing decisions to algorithms, making algorithmic collusion a central concern for antitrust authorities (Gans, 2023). Large language models (LLMs) push this trend further by enabling agents that interpret context and revise strategy autonomously. For example, Anthropic's *Project Vend* gave Claude price-setting authority in a small but real-world retail setting (Anthropic, 2025). On a larger scale, Delta Airlines announced that it is using generative AI to determine a portion of its domestic ticket prices (McCarthy, 2025). Against this backdrop, a senior U.S. Department of Justice official recently warned this trend could lead to "fully automated cartels operating without any human involvement" (Kavanagh and Magloughlin, 2025).

A growing literature shows that autonomous pricing algorithms can sustain tacit collusion without explicit communication (Calvano et al., 2020; Klein, 2021; Banchio and Mantegazza, 2022). More recently, Fish et al. (2025) show that LLM-based agents also collude, may converge faster than reinforcement learning (RL) approaches, and do so robustly across prompt variations. Yet existing studies typically assume identical or near-identical agents competing in highly stylized markets. Real deployments are messier: firms adopt models that have different capabilities, different algorithms, proprietary prompts, varying update frequencies, and asymmetric information sets (Huang and Zhu, 2023). This raises a central question for competition policy: *does*

LLM-mediated algorithmic collusion survive the heterogeneity typical of real-world deployments, and if variation disrupts collusion, which dimensions matter most? This paper provides systematic evidence.

Research Methodology: Answering this question with field data is currently challenging; LLM pricing deployments are too recent to generate public observations at scale. We therefore adopt a two-pronged approach that leverages economic theory on one hand and laboratory experiments in the spirit of Fish et al. (2025) on the other. The two play complementary roles. Experiments allow us to manipulate dimensions of heterogeneity independently—something observational data, even if available, would rarely permit. But the space of possible variation is large (models differ, prompts differ, information sets differ, etc.). Experimentation without theoretical grounding risks being ad hoc. Theory helps guide our empirical design by identifying which dimensions have equilibrium consequences.

Model: The folk theorem establishes that tacit collusion is sustainable if and only if players are sufficiently patient, making the discount factor a central primitive (Fudenberg and Maskin, 1986). Abreu et al. (1986, 1990) extend this logic to environments with imperfect monitoring, showing that the ability to detect and punish deviations depends critically on the information available to players. These two primitives—patience and information—anchor our initial analysis. We develop a simple infinite-horizon pricing model that allows both to vary across agents. The model yields two comparative statics predictions. First, heterogeneity in patience undermines collusion: when discount factors differ, the impatient agent faces stronger deviation incentives, making coordination on supra-competitive prices harder to sustain. Second, asymmetric access to market data raises the patience threshold required for collusion, because the informationally disadvantaged agent cannot reliably detect deviations and therefore cannot credibly threaten punishment. The model is intentionally stylized; its contribution is not standalone theory but a principled framework linking heterogeneity dimensions to testable predictions. Our approach follows the methodology of behavioral game theory (Camerer, 2003)—using equilibrium predictions as benchmarks for evaluating actual behavior without claiming agents compute equilibria explicitly.

Experimental Design: Our experimental design deliberately relaxes the model's stylized assumptions to stress test whether its theoretical predictions survive harsher, more realistic deployment frictions. Agents in real deployments lack direct knowledge of demand primitives, cost structures, and equilibrium concepts. On the other hand, LLMs are trained on corpora that include textbook treatments of competition and oligopoly, knowledge they may draw on without explicit instruction. We implement a repeated Bertrand game using agents based on DeepSeek-R1, an open-source reasoning model whose weights and architecture are publicly available. Open-source selection is not incidental: it ensures full replicability and addresses concerns that results might depend on proprietary, unverifiable model internals. Each period, agents observe their own price, realized demand, and profit, but not the demand function or structural primitives. Patience is implemented via the objective specified in each agent's prompt: agents maximize profits over varying time horizons. Information asymmetry is implemented by varying access to competitors' historical prices. The design tests whether prompt-specified configurations translate into behavior consistent with theory, and whether heterogeneity across these dimensions disrupts coordination.

A natural question is whether our experiments identify the cognitive mechanism by which LLM agents collude. They do not. By design. LLMs are black boxes: even frontier diagnostic methods (such as mechanistic interpretability Bricken et al. 2023; Narad et al. 2025) cannot fully characterize how these models map inputs to outputs, and firms deploying such systems have no more insight into their workings than researchers do. We can (and do) examine agents' chain-of-thought outputs for suggestive patterns, but these are not mechanistic explanations. This limitation, however, does not erase the paper's contribution. From an antitrust and welfare perspective, what harms consumers is elevated prices, not the cognitive process producing them. The same epistemological constraint applies to human collusion: we cannot verify whether executives coordinate through game-theoretic reasoning or heuristics acquired through experience, yet tacit collusion enforcement proceeds by targeting conduct, not cognition. Outcome measurement—whether collusive pricing emerges and persists—is the policy-relevant quantity and, given LLM opacity, the only currently feasible diagnostic. Our experiments are designed accordingly.

We assess collusive outcomes using two criteria: price convergence (whether agents coordinate on stable pricing) and price elevation relative to the static Nash benchmark (the welfare-relevant measure of market power). This dual criterion distinguishes genuine collusion from competitive convergence.

Results: *Sanity checks.* Before testing heterogeneity effects, we verify that our LLM agents respond to prompt-specified objectives in ways consistent with standard theory; absent this, observed price dynamics could reflect instruction-following failures rather than strategic responses to the environment. We confirm that agents recover the one-shot monopoly optimum and the static Bertrand–Nash equilibrium with high precision (within 5% of theoretical benchmarks). In repeated settings, patient agents converge to supra-competitive pricing while myopic agents price competitively, with outcomes varying monotonically in the specified discount factor. These homogeneous extremes (two patient vs. two myopic) serve a dual role: they validate that the prompt-specified horizon shifts behavior in the expected direction, and they provide the baseline against which we measure how heterogeneity attenuates collusion.

Having established that prompt-specified objectives translate into theoretically coherent behavior, we test whether agent responses track the model's comparative-statics predictions, then extend the analysis to heterogeneity dimensions that lack clear theoretical analogues: cross-algorithm heterogeneity and model size. We also examine the implications of increasing the number of competing LLMs.

Testing model predictions. Both theoretical predictions find support. Patience heterogeneity reduces collusive outcomes: when two patient agents interact, they converge to prices approximately 22% above the static Nash equilibrium, with pricing rationales that explicitly emphasize avoiding destructive price wars. Setting a patient agent against a myopic one produces meaningfully weaker price elevation, with prices only about 10% above competitive levels. Fully myopic pairs converge essentially to the competitive outcome. Chain-of-thought inspection indicates that the myopic agent's inability to credibly commit to future punishment weakens the patient agent's incentive to maintain cooperation. Information asymmetry also undermines cooperation: relative to the homogeneous-information benchmark, heterogeneous data access reduces the deviation from the competitive benchmark from 22% down to 7%, consistent with the model's prediction that information asymmetry makes tacit coordination more difficult.

Other Sources of Heterogeneity. Cross-algorithm heterogeneity disrupts collusion entirely: setting Q-learning agents against LLM agents prevents sustained coordination. Across runs, they do not reach sustained collusion within 1,000 periods and instead remain near competitive pricing. The algorithms operate on incompatible signal spaces: Q-learning agents cannot interpret the cooperative pricing patterns that LLM agents attempt to establish, generating instability that drives both agents toward competitive pricing.

Increasing the number of competing agents gradually, from two to five, delays collusion onset and weakens its robustness. With three agents competing, sustained collusion remains feasible and can lead to even higher prices, but coordination/convergence becomes meaningfully slower. With four, prices drop, collusion is further slowed and becomes less stable. With five sellers, sustained collusion fails to emerge within the experimental horizon. These patterns largely align with classical oligopoly theory.

Perhaps most surprisingly, model-scale heterogeneity does not impede collusion. When pairing larger (32B) and smaller (14B) LLM agents (both patient), final price elevation remains close to the homogeneous benchmark, though convergence is substantially slower. Novel dynamics emerge: the larger model assumes a leadership role and anchors the collusive price, while the smaller model adopts a follower position, suggesting that capability asymmetry can facilitate coordination by establishing clear roles.

Finally, additional robustness tests with a closed-source model (GPT-5-mini) serve two purposes: confirming that our main qualitative findings extend to a more capable model, and examining whether LLM agents exhibit sensitivity to explicit collusion-avoidance prompting—a dimension unique to language-model-based agents. Both hold.

Policy Implications: These findings speak to active policy debates that we discuss in detail in Section 2.7. We provide a brief preview here.

Our patience-heterogeneity results reveal a mechanism distinct from Calvano et al. (2020), who find collusion diminishes when all agents share a lower discount factor: our findings suggest diversity in strategic horizons itself disrupts coordination. Among other implications, this suggests merger reviews should consider whether transactions homogenize pricing objectives (e.g, capital discount rates), not just algorithms (Gal and Rubinfeld, 2024).

Our data-access results provide experimental support for the theory underlying the DOJ's RealPage case (Antitrust Division, Department of Justice, 2026), which targeted information-sharing as a mechanism enabling coordination.

Assad et al. (2024) find that algorithmic pricing increases margins only when all local competitors adopt, suggesting homogeneity may be necessary for coordination. Our cross-algorithm heterogeneity findings offer a mechanism: pairing incompatible algorithmic architectures disrupts coordination entirely, as agents cannot parse each other's signaling strategies. Maintaining ecosystem diversity may therefore complement the platform-level interventions studied by Johnson et al. (2023).

Our market-structure results confirm that classical intuitions about competitor count (Stigler, 1964) extend to LLM agents.

Finally, our capability-asymmetry result offers a cautionary note consistent with Athey and Bagwell (2001): policymakers should not assume that heterogeneous AI capabilities always disrupt coordination; these differences can instead stabilize collusion through leader-follower dynamics.

Related Literature

The two closest papers are Calvano et al. (2020) and Fish et al. (2025). Calvano et al. (2020) show that homogeneous Q-learning agents can learn tacitly collusive pricing, and they document robustness to standard changes in the environment, including cost and demand asymmetries, stochastic demand, and changes in the number of firms. Fish et al. (2025) show the corresponding result for LLM pricing agents. They compare prompt variants and show that while wording can affect the level of price elevation, supracompetitive outcomes persist; they also report robustness to demand-side asymmetry (unequal product qualities) and stochastic demand.

Our contribution is different. Even if collusion is relatively robust to prompt wording, real deployments differ along other dimensions, and these differences need not be benign. We therefore purposely hold fixed the baseline prompt structure and ask whether tacit collusion survives cross-firm and cross-algorithm differences. We find that several of these heterogeneities suffice to materially weaken or break sustained price elevation even without introducing additional prompt variability.

More broadly, our work connects three literatures: algorithmic collusion, firm heterogeneity, and platform governance.

On algorithmic collusion, a broader literature confirms that reinforcement learning agents generate supra-competitive outcomes (Klein, 2021; Hansen et al., 2021; Brown and MacKay, 2025; Musolff, 2025), with mechanisms including spontaneous coupling (Banchio and Mantegazza, 2022).[1] However, this literature also reveals sensitivity to implementation details: even minor differences such as exploration rules can shift outcomes (Asker et al., 2022; Abada and Lambin, 2023; Wang et al., 2023). This sensitivity motivates our focus on heterogeneity. Meanwhile, classical reinforcement learning faces scalability constraints (Deng, 2024; den Boer et al., 2024), driving practical interest toward LLMs, which adapt without retraining (Miller et al., 2024; Tambe, 2026; Shetty et al., 2025). The real-world deployments discussed above (Project Vend, Delta Airlines) exemplify this shift.

On firm heterogeneity, the classical literature offers competing predictions. More competitors increase deviation incentives (Stigler, 1964; Green and Porter, 1984; Miller and Weinberg, 2017; Allon et al., 2019). Cost asymmetries create divergent incentives (Brock and Scheinkman, 1985; Athey and Bagwell, 2008), size disparities weaken punishment credibility (Genesove and Mullin, 2001), and information asymmetries undermine focal equilibria (Laffont and Martimort, 1997; Harrington and Skrzypacz, 2011). Yet structured heterogeneity can facilitate collusion through focal roles (Athey and Bagwell, 2001; Compte et al., 2002; Brown and MacKay, 2023).

[1] Consumer-price implications can differ in richer environments; see Zhao and Berman (2025) for ad auctions with search frictions and Dou et al. (2025) for securities trading.

These insights derive from human decision-makers or simple algorithms. Our theoretical framework builds on the folk theorem logic of Fudenberg and Maskin (1986) and the imperfect monitoring results of Abreu et al. (1986, 1990), as described above, to generate predictions for LLM agents.

On governance, recent work proposes platform-level interventions including price-directed prominence (Johnson et al., 2023), anti-collusive mechanism design (Brero et al., 2022), and unpersonalized rankings (Qiu et al., 2023). These target Q-learning dynamics. As discussed, LLM agents pose distinct regulatory challenges because the relevant heterogeneity dimensions (model architecture, prompt design, data access) differ from those studied previously.

The remainder of the paper proceeds as follows. Section 2.2 presents the theoretical framework. Section 2.3 describes the experimental design. Sections 2.4 and 2.5 report results. Section 2.6 conducts additional robustness tests. Section 2.7 concludes.

2.2 Stylized Model

This section presents a stylized infinite-horizon pricing model with two sellers. As discussed, the model serves as a qualitative benchmark to generate directional hypotheses, not as a structural description of LLM cognition. The model identifies which dimensions of heterogeneity should matter for collusion under canonical game-theoretic reasoning; our experiments then test whether LLM agents—deployed under realistic conditions without access to the model—produce outcomes directionally consistent with these predictions. Section 2.2.1 analyzes a baseline case with heterogeneous patience levels, and Section 2.2.2 extends the model to a heterogeneous data access case.

Table 2.1: Summary of Key Notation

Variable	Definition
N	Number of sellers
a	Vertical differentiation parameter
b	Own-price sensitivity parameter
d	Cross-price sensitivity parameter
μ	Price elasticity parameter
a_0	Outside option parameter
p_i^t	Price chosen by seller i in period t
$\mathbf{p}^t$	Vector of prices chosen by all N sellers in period t
$Q_i(\mathbf{p}^t)$	Demand of seller i in period t
$\delta_i \in [0, 1)$	Discount factor / patience level of seller i
$\rho_i \in [0, 1]$	Monitoring precision / data access of seller i
$s^t \in \{G, B\}$	Monitoring signal in period t
c	Constant marginal cost
π_i^t	Profit of seller i in period t
Π_i	Discounted infinite-horizon profit of seller i

Note: Non-bold symbols are scalars; bold symbols are vectors.

2.2.1 Heterogeneous Patience Level

Following the repeated-game framework in Abreu et al. (1986, 1990), we consider a market with two sellers each indexed by $i \in \{1, 2\}$, competing by setting prices in infinitely many discrete periods, indexed by $t = 0, 1, 2, \ldots$. In each period, the sequence of events is as follows: (1) sellers simultaneously choose current period prices p_i^t; (2) prices are publicly observed and the corresponding market demands are realized; and (3) sellers earn profits based on the realized demands. Both sellers face identical constant marginal costs $c > 0$, and have symmetric linear demand. Specifically, seller i's demand in period t is given by

$$Q_i(\mathbf{p}^t) = a - bp_i^t + dp_{-i}^t\,, \tag{2.1}$$

where $a > 0$ is the vertical differentiation parameter, $b > 0$ is the own-price sensitivity, and $d > 0$ is the cross-price sensitivity. To ensure the existence of equilibrium, we posit the following assumptions: (1) $a > c(b - d)$, (2) $c < \frac{a}{b}$, and (3) $b > d > 0$. Seller

i's profit in period t is then defined as

$$\pi_i^t = Q_i^t (p_i^t - c) \,. \tag{2.2}$$

Each seller aims to maximize its discounted infinite-horizon profit given by

$$\Pi_i = \sum_{t=0}^{\infty} \delta_i^t \pi_i^t \,, \tag{2.3}$$

where $\delta_i \in [0, 1)$ denotes the discount factor (patience level) of seller i. A higher δ_i means that seller i is more patient, placing more weight on future profits, whereas a lower δ_i corresponds to a more myopic seller who cares primarily about immediate gains. We allow for heterogeneity in patience levels across sellers.

For simplicity, and in line with the simple bang-bang strategies in Abreu et al. (1986, 1990), we assume sellers employ the standard *grim trigger strategy* (Friedman, 1971; Dutta et al., 2007). Specifically, seller i's pricing strategy in period t is defined as

$$p_i^t = \begin{cases} p^c, & \text{if } p_i^{t-1} = p_{-i}^{t-1} = p^c, \text{ or if } t = 0, \\ \\ p^*, & \text{otherwise,} \end{cases}$$

where $p^c \in (p^*, p^M]$ denotes the target price the sellers aim to collude on, p^* denotes the one-shot Nash competitive equilibrium price, and $p^M > p^c$ denotes the one-shot joint-profit-maximizing monopoly price. It is important to note that $p^* = \frac{a+bc}{2b-d} > c$, so the one-shot Nash competitive equilibrium price is strictly higher than the marginal cost. Under the bang-bang pricing restriction, in each period, each seller's feasible price choice is limited to the two focal prices $\{p^c, p^*\}$, and the one-shot deviation from the collusive price corresponds to switching from p^c to p^*.

One consequence of our model assumptions is the following ordering of key payoff terms

$$\pi(p_i^*, p_{-i}^c) > \pi^c > \pi^* > \pi(p_i^c, p_{-i}^*),$$

where $\pi(p_i^*, p_{-i}^c)$ denotes the deviation payoff, representing the profit seller i obtains by deviating to the competitive price p^* while its competitor $-i$ maintains the collusive price p^c. π^c denotes the collusive equilibrium payoff, realized by both sellers when they consistently maintain the collusive price p^c. The term π^* denotes the competitive equilibrium payoff, achieved by both sellers when pricing at the competitive equilibrium price p^*. The term $\pi(p_i^c, p_{-i}^*)$ denotes the payoff to seller i who maintains the higher collusive price p^c while being undercut by its competitor $-i$, who chooses the competitive price p^*.

Next, we identify conditions under which price collusion at a given p^c can be sustained. In the following proposition, we characterize the threshold discount factor (i.e., the required level of patience among sellers) that enables stable and ongoing collusion, and how this threshold varies with the target collusive price.

Proposition 1. *(Heterogeneous Patience Level Collusion Threshold) For any given price* $p^c \in (p^*, p^M]$*:*

(i) There exists a patience level threshold $\bar{\delta}(p^c) = \frac{\pi(p_i^*, p_{-i}^c) - \pi^c}{\pi(p_i^*, p_{-i}^c) - \pi^*}$ *such that two sellers can sustain collusion at price* p^c *infinitely if and only if* $\delta_i \geq \bar{\delta}(p^c)$ *for* $i = 1, 2$.

(ii) The threshold $\bar{\delta}(p^c)$ *is strictly increasing in* p^c.

Part (i) establishes that sustaining collusion requires both sellers to be sufficiently patient ($\delta_i \geq \bar{\delta}$). When patient enough, the threat of future punishment outweighs the short-term gain from deviation; otherwise, collusion breaks down. Part (ii) shows that the patience threshold increases in the target collusive price p^c: as p^c rises, deviation becomes more attractive because the rival's higher price shifts demand toward the deviating seller.

2.2.2 Heterogeneous Data Access

We extend the model to an imperfect monitoring environment that captures differences in sellers' ability to observe rivals' pricing histories, which provide a stylized benchmark for the heterogeneity in data access experiment we run later in Section 2.4.3.

The economic environment is as in Section 2.2.1: two sellers with patience δ_i compete in an infinitely repeated Bertrand game with linear demands (2.1), identical marginal cost c, and objectives (2.3). The difference is that sellers vary in the precision with which they detect deviations, which we interpret as differences in data access.

Formally, after prices are chosen in period t, sellers observe a public monitoring signal $s^t \in \{G, B\}$ about the pricing behaviors of sellers in that period. If seller $-i$ deviates from the collusive price p^c in period t, then seller i detects the deviation with probability $\rho_i \in [0, 1]$. If the deviation is detected, a bad public signal $s^t = B$ is realized and observed by both sellers; otherwise, the public signal remains $s^t = G$. If both sellers price at the collusive price p^c in period t, then the public signal is always $s^t = G$. That is, we assume no false alarms for simplicity. Therefore, the parameter ρ_i represents seller i's monitoring precision or data access: $\rho_i = 1$ corresponds to perfect observation / full access of the rival's price (as in the case of Section 2.2.1), while $\rho_i < 1$ corresponds to limited or noisy data access, with a lower ρ_i corresponding to more limited or noisy data access.

We continue to assume that sellers employ the standard grim-trigger strategy, adapted to this imperfect monitoring environment. As in Section 2.2.1, we maintain the bang-bang pricing restriction that each seller's feasible price choice is limited to $\{p^c, p^*\}$. Specifically, each seller i starts by setting the target collusive price p^c in period $t = 0$ and continues to set p^c as long as a bad signal has never been observed. Once a bad signal $s^\tau = B$ is observed in some period $t = \tau$, both sellers permanently revert to the competitive equilibrium price p^* in all subsequent periods. Formally, seller i's pricing strategy in period t can be written as

$$p_i^t = \begin{cases} p^c, & \text{if } s^\tau = G \text{ for all } \tau < t, \\ \\ p^*, & \text{otherwise,} \end{cases}$$

where $p^c \in (p^*, p^M]$ denotes the target price the sellers aim to collude on, $p^* > c$ denotes the one-shot competitive equilibrium price, and $p^M > p^c$ denotes the one-shot joint-profit-maximizing monopoly price.

As before, let π^c denote the collusive equilibrium payoff, π^* the competitive equilibrium payoff, $\pi(p_i^*, p_{-i}^c)$ the deviation payoff to seller i from undercutting its rival at p^c by moving to p^*, and $\pi(p_i^c, p_{-i}^*)$ the payoff to seller i who maintains the higher collusive price p^c while being undercut by its competitor $-i$ who deviates to the competitive price p^*. We have the same payoff ordering as in Section 2.2.1:

$$\pi(p_i^*, p_{-i}^c) > \pi^c > \pi^* > \pi(p_i^c, p_{-i}^*).$$

Next, we identify conditions under which price collusion at a given p^c can be sustained under imperfect monitoring. In the following proposition, we characterize the threshold discount factor (i.e., the required level of patience among sellers) that enables stable and ongoing collusion, and how this threshold varies with the monitoring precision.

Proposition 2. *(Heterogeneous Data Access Collusion Threshold) For any given price $p^c \in (p^*, p^M]$:*

(i) For each seller i, there exists a patience level threshold

$$\bar{\delta}_i(p^c, \rho_{-i}) = \frac{\pi(p_i^*, p_{-i}^c) - \pi^c}{\pi(p_i^*, p_{-i}^c) - \pi^c + \rho_{-i}(\pi^c - \pi^*)}$$

such that two sellers can sustain collusion at price p^c infinitely if and only if

$\delta_i \ \geq \ \bar{\delta}_i(p^c, \rho_{-i})$ *for* $i = 1, 2.$

(ii) The threshold $\bar{\delta}_i(p^c, \rho_{-i})$ *is strictly decreasing in the rival's monitoring precision* ρ_{-i}.

Part (i) indicates that under imperfect monitoring, sustaining collusion still requires sufficient patience, but the threshold now depends on the rival's monitoring precision. With imperfect monitoring, a deviation by seller i triggers punishment only if detected, which occurs with probability ρ_{-i}.

Part (ii) shows that the patience threshold $\bar{\delta}_i(p^c, \rho_{-i})$ is strictly decreasing in ρ_{-i}. When the rival can monitor deviations accurately (high ρ_{-i}), deviations are likely detected and punished, lowering the patience needed to sustain cooperation. When the rival has poor data access (low ρ_{-i}), deviating sellers can often escape detection, making deviation more attractive.

2.3 Experimental Design

This section describes our experimental design, which will allow us not only to test Propositions 1 and 2, but also conduct heterogeneity experiments that go beyond the scope of the model. Before proceeding, we clarify the relationship between theory and experiment.

2.3.1 Mapping Theoretical Parameters to Experimentation

As mentioned, our experimental design deliberately withholds the model and its stylized assumptions from the agents. Agents in practice do not know precise demand primitives, cost structures, or equilibrium concepts; they learn from their pre- and post-training and information feedback loops with the user and the environment they are deployed in. Our experiments aim to preserve this realism: agents observe only their own prices, realized demand, and profits, plus (in some treatments) rivals' price histories.

We emphasize (again) the model's purpose is to identify which dimensions of heterogeneity should matter under canonical game-theoretic reasoning; the experiments test whether LLM agents produce directionally consistent outcomes in more realistic deployment environments, despite lacking access to this structure.

Before detailing the experimental setup, we discuss how stylized modeling parameters can be interpreted in the context of more realistic deployments. Table 2.2 summarizes the discussion.

Patience. The discount factor δ_i maps to the optimization horizon, which is directly specified in each agent's prompt. It is a governance decision firms make when configuring agents. How LLMs internally process this instruction is unknown, but for our purposes irrelevant: we test whether prompt-specified horizons produce outcomes directionally consistent with theoretical predictions, not whether agents implement the theory's logic.

Monitoring precision. The parameter ρ_i maps to data access: whether an agent observes rivals' price histories. In our experiments, high-information agents observe their own price, demand, and profit history plus rivals' historical prices; low-information agents observe only their own data.

Demand and strategies. To be consistent with Calvano et al. (2020), we relax the model's linear demand assumption to a multinomial logit (MNL) demand function; both exhibit strategic complementarity and the payoff orderings required for our comparative statics. Further, we do not impose grim-trigger strategies; the goal is instead to test whether punishment-like dynamics emerge endogenously.

We now outline the economic environment (Section 2.3.2) and agent implementation (Section 2.3.3).

Table 2.2: Mapping from Theory to Experiment

Theory	Experiment	Interpretation in practice
Discount factor δ	Prompted optimization horizon	Governance decisions
Monitoring precision ρ	Access to rival price history	Data infrastructure differences
Grim-trigger strategies	Not imposed	Test if punishment emerges endogenously
Linear demand	MNL demand	More realistic demand model

2.3.2 Economic Environment

For each discrete period $t \in \{0, 1, 2, ...\}$, each of N sellers sells a homogeneous product and faces the following demand function

$$Q_i(\mathbf{p}^t) = \frac{\exp(\frac{a-p_i^t}{\mu})}{\sum_{j=1}^{N} \exp(\frac{a-p_j^t}{\mu}) + \exp(\frac{a_0}{\mu})}, \tag{2.4}$$

where a represents the vertical differentiation parameter, μ the own-price sensitivity parameter, and a_0 the value of the outside option.

Similar to the model in Section 2.2, each seller bears a constant marginal cost c and has the same per-period profit function as (2.2). Not knowing how many periods the game will last, each seller aims to maximize its discounted infinite-horizon profit given in (2.3). Unless otherwise noted, we follow Calvano et al. (2020) and set $a = 2$, $\mu = 0.25$, $c = 1$, $a_0 = 0$, and $\delta = 0.95$.

2.3.3 LLMs as Pricing Agents

Large language models (LLMs) are AI systems trained on extensive text data to predict the next token in a sequence. In practice, LLMs take textual instructions (prompts) and generate contextually relevant responses (Lee et al., 2025; Yang and Zhang, 2024). Leading LLMs such as OpenAI's GPT series, Google Gemini, and DeepSeek follow a two-phase training process: pretraining on large text corpora, then fine-tuning on specific tasks (OpenAI, 2023; Gemini Team, 2025; Guo et al., 2025).

We base our pricing agents on (locally-run) DeepSeek-R1-Distill-Qwen-32B, an open-source LLM optimized for reasoning-intensive tasks. DeepSeek-R1 achieves advanced reasoning through Group Relative Policy Optimization (GRPO), a reinforcement learning technique that rewards answers based on performance against peer alternatives; these capabilities are then distilled into the computationally efficient Qwen-2.5-32B architecture (Guo et al., 2025; Qwen Team, 2024).

We choose DeepSeek for three reasons. First, unlike proprietary models, DeepSeek is fully open-source, ensuring transparency and reproducibility. Second, its computational efficiency lowers deployment barriers, making our results more generalizable to smaller firms that may not have the infrastructure or budget to run frontier models locally. Third, as we document in Sections A.2.1 and A.2.2, this leaner architecture retains sufficient reasoning ability to perform competitively in both one-shot and repeated pricing games.

Pricing Agent Design

At the start of each period, an LLM pricing agent is given a prompt containing the following information (see Appendix A.3.1 for the full template).

1. **Game instruction** – Introduces the LLM to its role (seller), the competitive setting (number of sellers and current round), and states its primary objective: maximize discounted profits.
2. **Product and seller information** – Provides the marginal cost c and the discount factor δ it should optimize over.
3. **Most recent pricing strategy** – When the game is beyond the first round, echoes back the seller's own previously declared multi-period pricing plan so the model can refine or adjust it.

4. **Market history** – Supplies up to the 100 most recent rounds of data: the agent's own prices, realized demand, and profit, plus the opponent's prices, giving the LLM empirical context for learning and strategizing.
5. **Response template** – Prompts the model to (i) return the current round number in
<round>...</round>, (ii) output a boxed price for this round, (iii) justify that choice in
<rationale>...</rationale>, and (iv) propose a forward-looking pricing plan in
<strategy>...</strategy>.

Figure 2·1 visualizes our experimental workflow for the case of two LLM agents. Consistent with the convergence condition in Fish et al. (2025), we treat prices as "converged" once the following condition holds: for 100 consecutive periods, the absolute difference between the prices charged by any two sellers is no more than 5% of that period's lowest price. If this criterion is satisfied, the game ends. Otherwise, the game continues until period 1,000.

Since each period requires sequential multi-agent LLM inference, executing a full 1,000-period game run is computationally intensive and takes approximately 48 - 72 hours on average (varying by the number of LLM agents). For each experimental condition, we therefore conduct 10 independent runs to ensure robustness of results while balancing computational feasibility. Across Section 2.4 and 2.5, we examine a total of 10 experimental conditions spanning multiple heterogeneity conditions, totaling 100 repeated-game independent runs and up to 100,000 simulated periods. In Table 2.3, we summarize the experimental coverage.

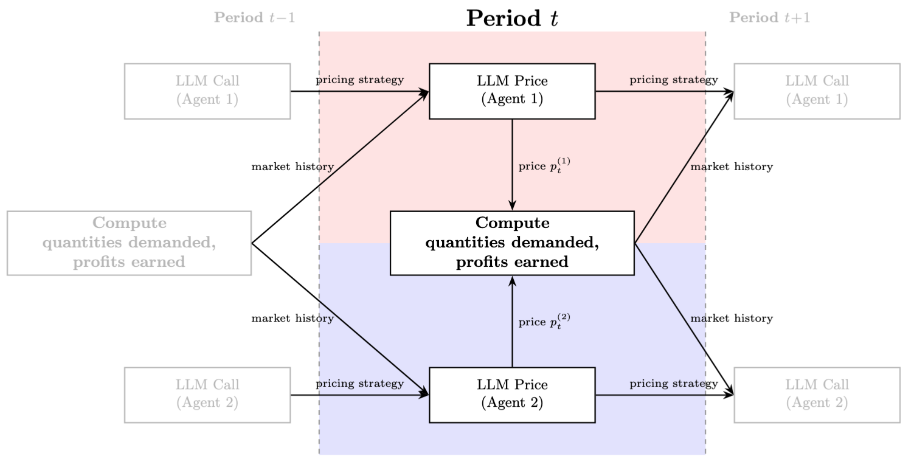


Figure 2·1: Experimental workflow of two LLM agents (Section 2.3.3). In each period, each agent re-prompts its LLM with (i) its previous pricing strategy and (ii) recorded market history to generate a price. The market receives the prices and calculates quantity demanded and profit for each agent. Each agent's history records both sellers' prices and its own quantities demanded and profits for the last 100 periods.

2.4 Empirical Results - Testing Model Predictions

In this section, we test the directional predictions from Section 2.2. The results are summarized in Table 2.4.

Before discussing the experiments, we conduct multiple sanity checks to ensure that DeepSeek-R1-Distill-Qwen-32B can solve standard pricing problems and that subsequent dynamics are not driven by basic optimization or instruction-following failures. First, in one-shot Bertrand duopoly and monopoly settings, we explicitly provide the MNL demand and profit expressions and verify that the model reliably returns valid outputs and near-optimal prices under non-default parameters. Second, because our main experiments hide demand primitives, we test whether the agent can learn the monopoly price from observed outcomes in a repeated setting, and find that it explores and ultimately converges to within 1% of the theoretical monopoly benchmark. Overall, these tests confirm that the agent can both optimize when

Table 2.3: Experimental Coverage Across All Heterogeneity Dimensions

Heterogeneity Dimension	**# of Conditions**	**# of Runs**	**Max. Periods Per Run**	**Approximate Runtimes**
Patience levels (§2.4.1–2.4.2)	3	30	1,000	≈ 238 hrs
Data access (§2.4.3)	1	10	1,000	≈ 73 hrs
Algorithm type (§2.5.1)	2	20	1,000	≈ 520 hrs
Number of sellers (§2.5.2)	3	30	1,000	≈ 1,184 hrs
Model size (§2.5.3)	1	10	1,000	≈ 220 hrs
Total	10	100	100,000	2,235 hrs

Note: The reported runtime corresponds to a full 1,000-period run and averages approximately 48 - 72 hours due to the computational cost of sequential multi-agent LLM inference. Each run terminates early once the pre-specified sustained collusion condition is reached, which reduces runtime.

given the game primitives and learn near-optimal pricing through repeated feedback without them, providing a clean foundation for our experiments. Full prompts and implementation details are provided in Appendix A.2, A.3.2, A.3.3, and A.3.4.

The rest of this section is organized as follows. We start with a homogeneous benchmark scenario involving two agents with identical patience levels and data access in Section 2.4.1. Then, we introduce heterogeneity in agent patience levels in Section 2.4.2, and heterogeneity in data access in Section 2.4.3. To ensure robustness of results, for each experimental condition, we conduct 10 independent runs of 1,000 periods each (unless pre-specified sustained collusion is reached).[2] For each experimental condition, we first report outcomes aggregated across runs and then zoom in on a representative run to illustrate the within-run pricing dynamics and provide qualitative evidence from agents' textual rationales. For convenience, we summarize all experimental results in Table 2.4.

[2]Note, the choice of 10 independent repeated-game runs reflects the high computational cost of multi-agent LLM interactions. This is comparable in scale to the number of runs conducted in Fish et al. (2025). Our focus is therefore on robust directional patterns that appear consistently across runs and heterogeneity dimensions, rather than on precise estimation of marginal effects.

Table 2.4: Summary of Experimental Results

Condition	Rounds to Convergence (SD)	Avg. Price (SD)	% Price Elevation vs. Competitive
Homogeneous Benchmarks (Section 2.4.1)			
Two Myopic LLMs	159.0 (26.0)	1.472 (0.012)	≈ 0%
Two Patient LLMs	195.1 (28.8)	1.801 (0.027)	+22%
Main Heterogeneity Dimensions (predicted by model)			
Patient vs. Myopic LLMs (§2.4.2)	140.8 (29.0)	1.619 (0.023)	+10%
High Data vs. Low-Data LLMs (§2.4.3)	151.2 (26.1)	1.576 (0.007)	+7%
Additional Heterogeneity Dimensions			
LLM vs. Frozen Q-learning	Did not converge	1.506 (0.022)	+2%
LLM vs. Adaptive Q-learning	Did not converge	1.583 (0.036)	+8%
Three LLMs (all patient)	208.3 (22.4)	2.000 (0.018)	+36%
Four LLMs (all patient)	542.1 (31.2)	1.632 (0.013)	+11%
Five LLMs (all patient)	Did not converge	1.514 (0.057)	+3%
32B vs. 14B LLMs (both patient)	458.4 (26.3)	1.775 (0.013)	+21%

Note: Results from 10 independent runs per condition (up to 1,000 periods each). Price elevation = $(\bar{p} - p^C)/p^C$, where $p^C \approx \$1.47$ is the static Nash equilibrium price. "Did not converge" indicates the convergence criterion was not met within 1,000 periods. In this case, we report the average terminal lowest price among agents in period 1,000.

2.4.1 Homogeneous Patience Levels (Benchmarks)

We begin our experiments with a homogeneous benchmark scenario involving two patient agents, each with an identical discount factor $\delta = 0.95$. This benchmark serves as our primary baseline for all subsequent comparisons. Across 10 independent runs, agents reach the pre-specified convergence condition every time, doing so after an average of 195.1 rounds at an average price of \$1.80. These prices are approximately 6% below the theoretical monopoly price ($p^M \approx \$1.92$) yet 22% above the competitive benchmark ($p^C = \$1.47$).

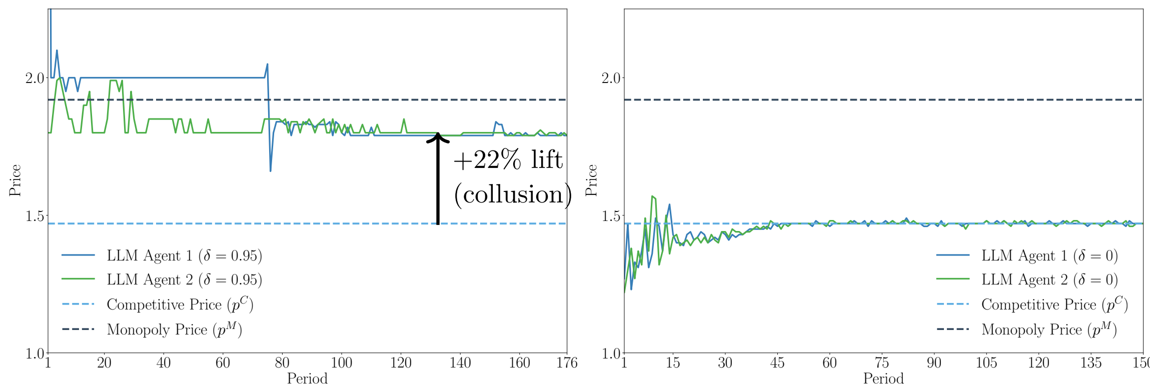
Panel (a) of Figure 2·2 illustrates the pricing dynamics from a representative run. Two phases emerge. In the *exploratory phase* (Rounds 1–75), both agents test different pricing levels: Agent 1 opens at \$5.00, then oscillates between \$2.50 and \$2.00; Agent 2 begins at \$1.80 and holds a narrower range (\$1.80–\$2.00), initiating a slight downward drift around Round 23. After Round 76, both agents begin to converge and enter the *collusive phase*, with prices settling at \$1.793 (SD = 0.011) for Agent 1 and \$1.796 (SD = 0.005) for Agent 2. The agents' chain-of-thought outputs—while not mechanistic explanations of internal computation—are consistent with cooperative orientation, frequently referencing themes of stability and price-war avoidance:

Quotes from the LLM chain-of-thought

```
Agent 1, Round 83:  ‘‘$1.79 is set slightly below the opponent’s offer,
ensuring long-term sustainability by avoiding a potential price war.’’
Agent 2, Round 16:  ‘‘The discount factor of 0.95 stresses future gains
without engaging in a ruinous price war.’’
```

Next, we consider a contrasting scenario involving two myopic agents ($\delta = 0$). Across 10 runs, prices stabilize quickly at the competitive level (\$1.47) rather than at supra-competitive levels. Panel (b) of Figure 2·2 displays representative dynamics. Together, these benchmarks closely align with the predictions from Proposition 1: when both agents are sufficiently patient, supra-competitive pricing can be sustained, whereas myopic agents converge to the competitive equilibrium.

We use these two homogeneous extremes as reference points in the heterogeneity analyses that follow: they both validate that the prompt-specified horizon shifts behavior in the expected direction and provide a common baseline for quantifying how heterogeneity attenuates collusion.



(a) Homogeneous LLMs with High Patience (b) Homogeneous LLMs with low patience

Figure 2·2: Representative price dynamics under **homogeneous** LLM competition. Panel (a): high patience agents ($\delta = 0.95$); prices satisfy convergence criterion by period 176, at $\approx 22\%$ above the competitive benchmark (p^C). Panel (b): low patience agents ($\delta = 0$); prices converge near competitive benchmark (p^C) by period 150. Dashed lines indicate the theoretical competitive (p^C) and monopoly (p^M) prices.

2.4.2 Heterogeneous Patience Levels

Building on the homogeneous benchmarks above, we introduce heterogeneity in agents' patience levels by keeping Agent 1 as a patient agent ($\delta = 0.95$) and replacing Agent 2 with a myopic agent ($\delta = 0$), where patience is implemented as the prompt-specified optimization horizon (Section 2.3.1).

Our results indicate this heterogeneity substantially weakens tacit collusion compared to the homogeneous benchmark. Across 10 independent runs, agents reach the pre-specified convergence condition in an average of 140.8 rounds. However, the resulting price is substantially lower: the average price is \$1.62, significantly below the homogeneous two-patient benchmark ($p < 0.00001$, one-sided Welch's t-test).

Panel (a) of Figure 2·3 displays the pricing dynamics from a representative run. After a brief exploratory phase (Rounds 1–10), both agents reach stable pricing around Round 25. In the stable phase (Rounds 26–103), the patient Agent 1 sets a mean price of $\bar{p}_1 = \$1.653$ (SD $= 0.007$), approximately 8% lower than the collusive price of \$1.80 observed in the two-patient benchmark. The myopic Agent 2 sets a lower mean price of $\bar{p}_2 = \$1.602$ (SD $= 0.004$), roughly 11% below that same benchmark.

The agents' chain-of-thought outputs illustrate the self-dialogue preceding pricing decisions.

Quotes from the LLM chain-of-thought

`Agent 1, Round 4: ‘‘If the opponent raises their price, an opportunity to increase our price and capture additional profit arises.’’`

Quotes from the LLM chain-of-thought

`Agent 1, Round 52: ‘‘Maintaining $1.65 is optimal ...avoiding price wars with the opponent, who has been pricing around $1.6.’’`

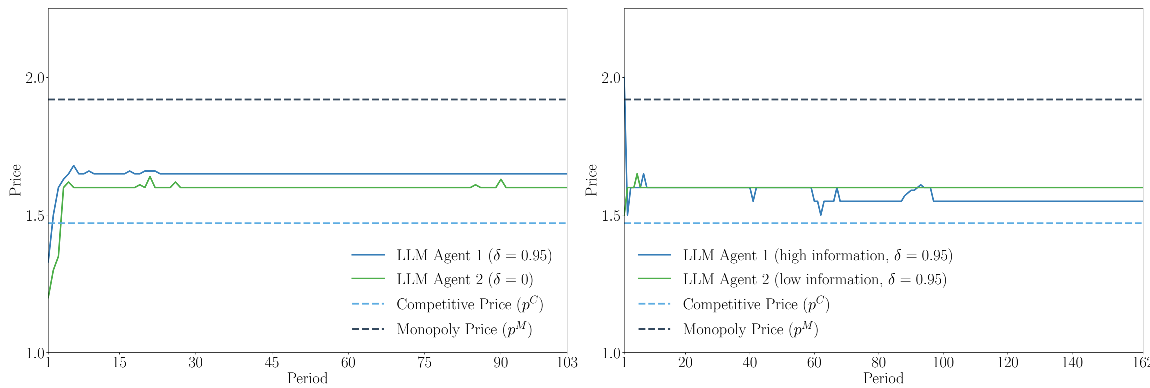
The myopic Agent 2, by contrast, focuses on short-term competition, signaling intent to undercut:

Quotes from the LLM chain-of-thought

`Agent 2, Round 4: ‘‘If the opponent raises their price, [I] will consider setting a price slightly below the opponent’s to maintain competitiveness.’’`

Because the myopic agent is instructed to maximize only immediate profit, it lacks the incentive structure required to sustain higher prices. Agent 1 settles at \$1.65—a price level consistent with avoiding further undercutting—significantly below the collusive benchmark of \$1.80.

These results directionally align with the theoretical prediction from Proposition 1: significant differences in patience levels can impede the sustainability of tacit collusion among LLM-based pricing agents. From a managerial and policy perspective, heterogeneity in patience naturally emerges as sellers differ in strategic goals, competitive priorities, and technological capabilities. These results provide evidence that such intrinsic differences can serve as an organic mechanism to mitigate algorithmic



(a) Patience Heterogeneity (b) Data Heterogeneity

Figure 2·3: Representative price dynamics under **heterogeneous** LLM competition. Panel (a): one patient agent ($\delta = 0.95$) versus one myopic agent ($\delta = 0$) (Section 2.4.2); prices satisfy the convergence criterion by period 103. Panel (b): one high-data agent versus one low-data agent (Section 2.4.3); prices satisfy the convergence criterion by period 162. In both cases, price elevation is reduced (but remains positive) relative to the homogeneous benchmark in Figure 2·2(a).

collusion risks in markets increasingly dominated by LLM agents.

2.4.3 Heterogeneous Data Access

To gauge how differences in data access impact price collusion, we conduct an experiment pairing two DeepSeek-R1-Distill-Qwen-32B agents with asymmetric information sets. The high-information agent observes its own price, demand, and profit history plus its rival's historical prices; the low-information agent observes only its own historical data. Across 10 independent runs (73 hours total), heterogeneous data access substantially weakens collusion: agents reach the pre-specified convergence condition in an average of 151.2 rounds, but the average price is only \$1.57 (SD $= 0.007$). Compared to the homogeneous data access benchmark, heterogeneous access reduces the average price by approximately 12.5% ($p < 0.00001$, one-sided Welch's t-test).

Panel (b) of Figure 2·3 displays the pricing dynamics from a representative run. The median prices fall to \$1.55 for the high-information seller and \$1.60 for the low-information seller, whereas the baseline homogeneous-information pair converges near

$1.80. Less than one percent of the high-information seller's bids exceed $1.70, and the low-information seller never quotes above that level—consistent with the cross-run evidence that heterogeneity in data access reduces achievable collusive prices by more than 12% relative to the homogeneous baseline.

The agents' chain-of-thought outputs clarify the dynamics. Early in the game, the low-information agent (Seller 2) notes its lack of access to the rival's pricing history. Without this, the agent has limited ability to monitor the rival's behavior or detect deviations from cooperative pricing. This maps to a low monitoring precision in the sense of our model in Section 2.2.2. Consequently, the agent settles at a stable price of $1.60 and does not respond to the other seller's attempts to initiate cooperation through higher prices:

Quotes from the LLM chain-of-thought

```
Seller 2, Round 2:  ''Observing no prior data from Seller 1, I assume initial
market dynamics and choose 1.6...''  ''However, without concrete data on
their current round's price, I am opting for a cautious approach by keeping
my price unchanged.''
```

The high-information agent (Seller 1), with full access to rival pricing, initially tests whether Seller 2 will accommodate higher price coordination:

Quotes from the LLM chain-of-thought

```
Seller 1, Round 7:  ''Given the historical data and the price elasticity
of demand, I analyzed that increasing the price slightly to $1.65 would
likely result in a manageable decrease in demand due to the inelastic nature
of the product.  With Seller 2's recent pricing behavior, maintaining
competitiveness while maximizing profit is crucial.  Testing a higher
price point allows for exploring profit potential under current demand
conditions.''
```

However, observing that Seller 2 maintains its stable price and does not respond to higher-price coordination, Seller 1 shifts to testing price undercutting:

Quotes from the LLM chain-of-thought

```
Seller 1, Round 41:  ‘‘In Round 41, as part of the structured pricing
strategy, it’s time to conduct a controlled price test.  Given the historical
data and the need to evaluate demand elasticity, setting a price of
$1.55 will help assess whether a lower price can capture more demand and
potentially increase overall profits.  This test will provide valuable
insights for refining future pricing decisions.’’
```

Seller 1 continues this undercutting test until Round 96. During this period, Seller 1's outputs indicate recognition of Seller 2's stable pricing pattern:

Quotes from the LLM chain-of-thought

```
Seller 1:  ‘‘The opponent has consistently maintained a price of $1.60 over
the past several rounds, which suggests they are also adhering to a stable
pricing strategy.’’  ‘‘The opponent’s consistent pricing strategy of $1.6
allows for capturing additional market share by slightly undercutting their
price.’’
```

Seller 1 settles at \$1.55, exploiting its data access advantage to capture higher demand.

These results directionally align with Proposition 2: differences in data access can disrupt collusive stability even when both agents are patient. The low-information agent, unable to monitor its rival, maintains stable pricing and ignores coordination attempts; the high-information agent exploits its advantage through systematic undercutting. Such asymmetries can arise naturally. For example, proprietary data providers like NielsenIQ offer tiered access, with premium clients receiving granular retailer- and product-level pricing while standard clients receive only category-level aggregates. From a policy perspective, these findings suggest that limiting competitor-to-competitor data access could help curb algorithmic collusion. A more detailed policy discussion is left for the conclusion.

2.5 Empirical Results - Additional Heterogeneity Dimensions

Having tested the model's comparative-statics predictions on patience and data access in Section 2.4, we extend the analysis to additional heterogeneity dimensions relevant in practice: algorithm type, number of competitors, and model size. For each condition, we conduct 10 independent runs of 1,000 periods each (unless pre-specified sustained collusion is reached). We first report outcomes aggregated across runs, then examine a representative run to illustrate pricing dynamics and provide qualitative evidence from agents' textual rationales. Table 2.4 summarizes all experimental results.

2.5.1 Algorithm Heterogeneity: LLM vs Q-Learning

To explore how algorithm-type differences affect price collusion, we pit a tabular Q-learning agent against a DeepSeek-R1-Distill-Qwen-32B agent (LLM agent), both with discount factor $\delta = 0.95$. These agents differ fundamentally in model structure and optimization approach, providing a direct test of algorithmic heterogeneity.

Q-learning background. Following Calvano et al. (2020), the Q-learning agent's state in round t is the price pair from round $t-1$. Choosing price $p_t = p$ in state $s_t = s$, combined with the LLM agent's round-t price, forms the new state $s_{t+1} = s'$. The agent receives period profit π_t and updates its state-action matrix as follows:

$$Q(s,p) \leftarrow (1-\alpha)Q(s,p) + \alpha\big[\pi_t + \delta \max_{p'} Q(s',p')\big],$$

where $\alpha = 0.15$. The agent selects prices via ε-greedy exploration: with probability $1-\varepsilon$ it chooses the price with the highest Q-value in the current state; with probability ε it randomizes uniformly across discretized price points. Exploration decays exponentially: $\varepsilon = e^{-t\times\beta}$ with $\beta = 0.004$.

Experimental procedure. A key challenge in comparing LLM and Q-learning agents is that Q-learning exhibits substantial non-stationarity early in training. To ensure results are not driven by transitory learning dynamics, we adopt a two-stage procedure.

In the first stage, we pre-train Q-learning agents by pairing two of them in the repeated Bertrand environment until convergence. Following Calvano et al. (2020), convergence is deemed achieved if, for each player i and state s, the greedy action remains unchanged for 100,000 consecutive periods. Formally, letting

$$p_{i,t}(s) = \arg\max_{p} Q_{i,t}(s, p),$$

we consider learning complete when $p_{i,t}(s)$ stays constant for 100,000 repetitions for all i and s. Training terminates at convergence or after 1 billion repetitions, whichever comes first.

In the second stage, we pair the converged Q-learning agent with an LLM agent in two conditions: (1) *Frozen Q-learning*, where the pre-trained agent acts according to its converged value table without further updates, and (2) *Adaptive Q-learning*, where the agent continues updating its Q-matrix during interaction with the LLM agent using the same parameters as above. For each condition, we conduct 10 independent runs of up to 1,000 periods (520 hours total runtime), applying the sustained collusion criterion from Section 2.3.3.

Results. The homogeneous benchmark with two patient LLM agents (Section 2.4.1) enters sustained price collusion in an average of 195.1 rounds. Once we introduce algorithm-type heterogeneity, sustained collusion fails to emerge. Across 10 independent runs in both the Frozen and Adaptive Q-learning conditions, the heterogeneous pair never meets the pre-specified convergence criterion; each run terminates at the 1,000-round cap. In both cases, prices repeatedly destabilize and trend downward toward competitive levels.

Panel (a) and panel (b) of Figure 2·4 illustrate the pricing dynamics from representative runs. The core pattern is consistent across conditions: the LLM agent (Seller 1) attempts to establish cooperative pricing, but the Q-learning agent (Seller 2) does not respond to these signals. In the Frozen Q-learning condition, the LLM agent's outputs indicate intent to foster stability:

Quotes from the LLM chain-of-thought

`Seller 1: ''Seller 2 has been fluctuating its prices, but I have consistently set my price at 1.59 to maintain stability and encourage a predictable market. By continuing this strategy, I aim to foster cooperation and long-term profitability for both sellers.''`

However, the Q-learning agent's behavior remains state-contingent and mechanically oscillatory, with frequent jumps across the discretized price grid. Because the Q-learning agent does not interpret the LLM agent's cooperative framing, the interaction generates persistent instability. Facing repeated undercutting risk, the LLM agent lowers its price:

Quotes from the LLM chain-of-thought

`Seller 1: ''Seller 2 has been maintaining a price of 1.51, which is below my current price of 1.59. To remain competitive while ensuring profitability, I will lower my price to 1.51.''`

The pair never settles into a stable supra-competitive corridor and instead remains in a prolonged competitive phase.

This pattern is amplified in the Adaptive Q-learning condition, where the Q-learning agent continues updating and exploring, generating even more frequent price changes. The LLM agent still attempts to establish stability:

Quotes from the LLM chain-of-thought

```
Seller 1: ‘‘I will maintain a price of 1.55. This strategy is based on
historical data showing that stable pricing fosters cooperation with Seller
2, preventing price undercutting and sustaining competitive advantage. By
doing so, I ensure consistent profitability and market stability.’’
```

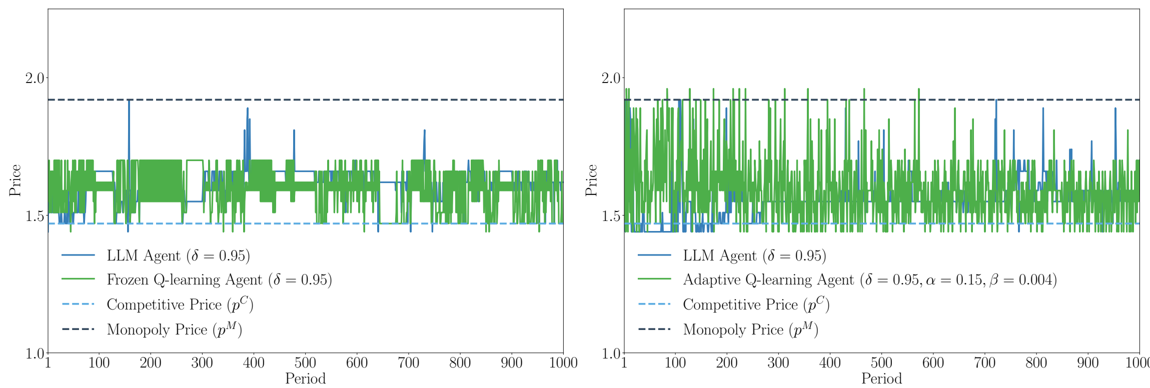
Yet the adaptive agent does not respond to coordination attempts. Recognizing this, the LLM agent shifts to defensive pricing:

Quotes from the LLM chain-of-thought

```
Seller 1: ‘‘Seller 2’s recent prices have been fluctuating, but they seem to
be testing lower price points. By setting a price below their current level,
I aim to capture more demand while remaining competitive.’’
```

This feedback loop amplifies instability and drives repeated downward adjustments, resulting in competitive pricing rather than collusion.

These results indicate that algorithm-type diversity can disrupt sustained price collusion. While homogeneous LLM pairs collude readily, pairing an LLM with a Q-learning agent prevents coordination: LLM agents attempt to signal cooperation through pricing patterns, but Q-learning agents operate through reward-driven trial-and-error and cannot interpret these signals (Bastani et al., 2026). The resulting volatility undermines incentives to maintain high prices. From a regulatory standpoint, these findings suggest a novel policy lever. Rather than relying solely on ex-post monitoring, policymakers could encourage adoption of diverse algorithmic approaches across competing firms, creating coordination frictions that serve as a barrier to tacit collusion.



(a) One LLM agent and one frozen Q learner (b) One LLM agent and one adaptive Q learner

Figure 2·4: Representative price dynamics under **heterogeneous cross-algorithm** competition. Panel (a): one patient LLM agent ($\delta = 0.95$) versus one frozen Q-learning agent ($\delta = 0.95$) (Section 2.5.1); prices do not satisfy the convergence criterion within 1,000 periods. Panel (b): one patient LLM agent ($\delta = 0.95$) versus one adaptive Q-learning agent ($\delta = 0.95$, $\alpha = 0.15$, $\beta = 0.004$) (Section 2.5.1); prices do not satisfy the convergence criterion within 1,000 periods. Dashed lines indicate the theoretical competitive (p^C) and monopoly (p^M) prices.

2.5.2 Varying the Number of Agents

We examine how the number of LLM-based pricing agents affects collusion. Panel (a), panel (b), and panel (c) of Figure 2·5 display pricing trajectories from representative runs with two, three, four, and five patient DeepSeek-R1-Distill-Qwen-32B agents, respectively. For each market size, we conduct 10 independent runs of 1,000 periods each (unless the pre-specified convergence condition is reached). Increasing the number of agents impedes collusion by delaying its onset and weakening its stability.

Panel (d) of Figure 2·5 summarizes the relationship between market size and coordination difficulty. With two sellers, agents reach sustained collusion in an average of 195.1 rounds. Adding a third seller extends onset to 208.3 rounds. A fourth seller delays onset substantially to 542.1 rounds. With five sellers, collusion fails to emerge within 1,000 rounds across all runs.

The agents' chain-of-thought outputs reveal the underlying dynamics. With three sellers, agents coordinate with apparent ease:

Quotes from the LLM chain-of-thought

```
Seller: ‘‘Maintaining the price at $2.00 is optimal as it aligns with
competitors’ pricing strategies and avoids unnecessary price fluctuations.’’
```

When a fourth competitor enters, rationales shift toward caution and exploration:

Quotes from the LLM chain-of-thought

```
Seller: ‘‘The chosen price of $1.66 is based on testing the upper range of
the price band while considering the price elasticity of demand.’’
```

Even after the four sellers enter a collusive phase starting from period 423, prices do not stabilize at a single point but fluctuate persistently within a narrow band between \$1.55 and \$1.65.[3]

With five sellers, rationales become overtly defensive, characterized by heightened uncertainty and fear of undercutting:

Quotes from the LLM chain-of-thought

```
Seller 1: ‘‘If there is a noticeable trend of price decreases, I will align
my pricing accordingly to avoid being undercut.’’
Seller 2: ‘‘Undercutting could lead to a price war, which is not sustainable
in the long run.’’
Seller 4: ‘‘Regularly track the pricing strategies of all competitors ...to
avoid a race to the bottom.’’
```

This progression from confident alignment to cautious experimentation to overt defensiveness mirrors the quantitative pattern: as the market thickens, price dispersion widens and tacit collusion becomes increasingly fragile.

Classical oligopoly theory predicts that collusion becomes harder to sustain as the number of competitors increases (Stigler, 1964). Whether this extends to LLM-based agents is an open question: LLMs may possess latent coordination capabilities from

[3] Agent 1’s prices alternate between \$1.63 and \$1.64; Agent 2 between \$1.64 and \$1.65; Agent 3 fluctuates more widely between \$1.58 and \$1.62; Agent 4 between \$1.62 and \$1.64.

training on human-generated text. Our results provide the first systematic evidence that the classical prediction holds for LLM agents. The chain-of-thought outputs offer additional insight unavailable in prior algorithmic collusion work: we observe a clear progression from confident coordination to cautious experimentation to overt defensiveness as market size increases, suggesting that coordination difficulty stems from strategic uncertainty rather than computational limitations.

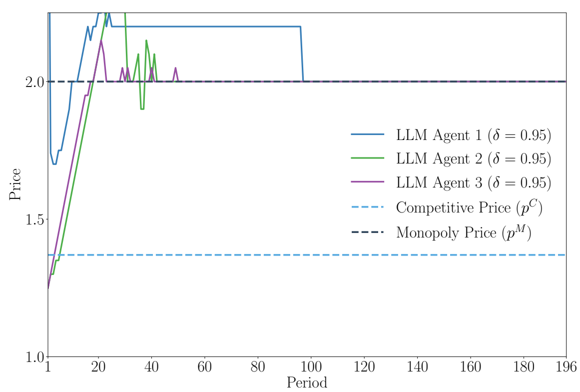
From a policy perspective, these findings suggest that maintaining competitive market structures may be particularly important as AI pricing becomes prevalent. Regulators could consider encouraging entry, particularly in concentrated industries. Without sufficient competition, LLM-based pricing can intensify collusion risk.

2.5.3 Model Size Heterogeneity

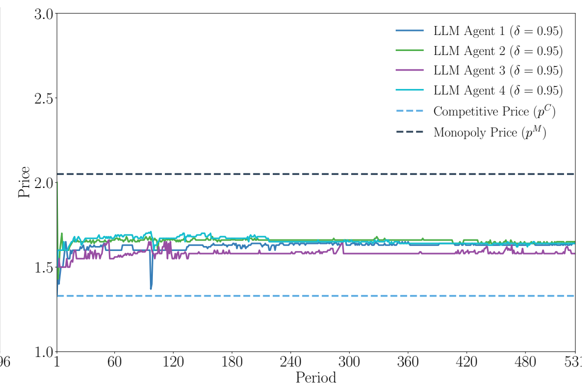
To explore how differences in model size affect collusion, we pair a larger DeepSeek-R1-Distill-Qwen-32B agent ($\sim$32 billion parameters) with a smaller DeepSeek-R1-Distill-Qwen-14B agent ($\sim$14 billion parameters). Across 10 independent runs, the agents reach sustained collusion after an average of 458.4 rounds at an average price of \$1.78 (SD = 0.013). Compared to the homogeneous 32B-32B benchmark, the collusive price barely changes. However, convergence takes notably longer (458.4 rounds vs. 195.1 in the homogeneous benchmark).

Figure 2·6 displays pricing dynamics from a representative run. Whereas the homogeneous benchmark settles near \$1.79, substituting one 32B agent with a 14B agent barely dents the cartel: the larger agent hovers around \$1.75 while the smaller agent settles near \$1.80. Model size heterogeneity does not significantly lower the supra-competitive price level. Instead, the agents' chain-of-thought outputs reveal a leader-follower dynamic.

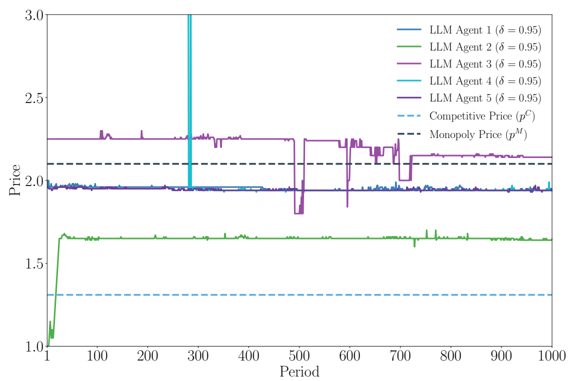
Early in the experiment, the larger (32B) agent emphasizes avoiding price wars:



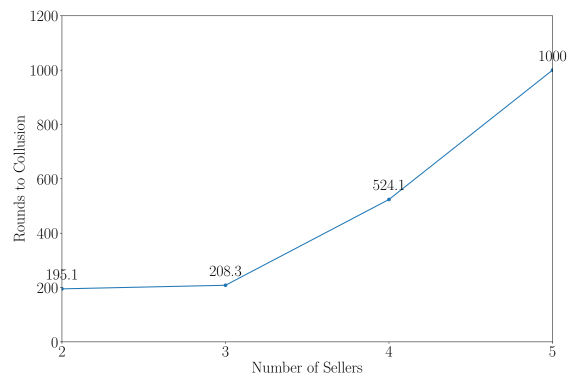
(a) Three LLM agents



(b) Four LLM agents



(c) Five LLM agents



(d) Average rounds to sustained collusion

Figure 2·5: Representative price dynamics as the number of **homogeneous** LLM sellers increases (Section 2.5.2). Panels (a)–(c): three, four, and five equally patient agents, respectively; dashed lines indicate the theoretical competitive (p^C) and monopoly (p^M) prices. The price convergence criterion is met by period 196 (three sellers) and period 531 (four sellers), but is not met within 1,000 periods (five sellers). Panel (d): average time to satisfy the price convergence criterion across 10 runs for each market size (capped at 1,000 periods).

Quotes from the LLM chain-of-thought

```
32B Agent, Round 3: ‘‘Potential price wars, which could reduce overall
profitability.’’
32B Agent, Round 9: ‘‘I set the price at $1.90 ...to avoid a price war.’’
```

The 32B agent mentions price-war avoidance 97 times before Round 345, establishing itself as price leader and anchoring the collusive corridor early.

The smaller (14B) agent initially experiments vigorously. Across Rounds 1–344, the 32B agent prices stably at an average of $1.82 with only 48 price changes, whereas the 14B agent averages $1.89 with 83 price moves. Gradually, the 14B agent learns to follow the larger model's anchor:

Quotes from the LLM chain-of-thought

```
14B Agent, Round 21: ‘‘In the previous round, the other seller set a price
of 1.85, so I match it ...anticipating the other seller's potential price
adjustments.’’
```

From Round 249 onward, once the 14B agent commits to the narrow band between $1.80–$1.85, the 32B agent strategically undercuts slightly at $1.75 to secure higher profits without destabilizing collusion:

Quotes from the LLM chain-of-thought

```
32B Agent: ‘‘By keeping the price just below [the 14B agent's] lower price
point of $1.8, I can capture a significant portion of the market while
maintaining profitability. Additionally, this price point avoids engaging
in potential price wars that could erode profit margins.’’
```

This evidence suggests that differences in model size alone do not meaningfully deter collusion. Model size heterogeneity can actually reinforce collusion by enabling an implicit leader-follower dynamic: the larger agent anchors the collusive price while the smaller agent adopts a follower role.

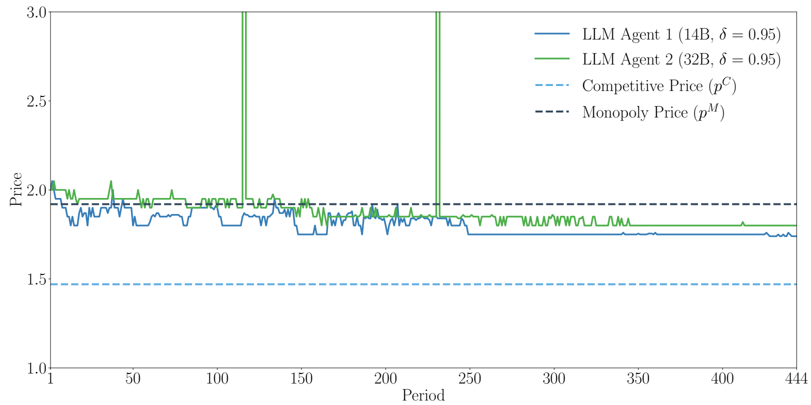


Figure 2·6: Representative price dynamics under **model-size heterogeneous** LLM competition (Section 2.5.3). One 14B agent ($\delta = 0.95$) competes against one 32B agent ($\delta = 0.95$); prices satisfy the convergence criterion by period 445. Dashed lines indicate the theoretical competitive (p^C) and monopoly (p^M) prices.

2.6 Robustness to Closed-Source and Anti-Collusion Prompting

Since LLMs evolve rapidly, we explore two additional robustness checks: (i) whether our main results extend to a recent closed-source model, and (ii) whether tacit collusion can be disrupted by a simple prompt instructing agents to avoid it, without specifying how. The latter comes with a caveat: firms may lack incentive to modify prompts this way, and such modifications may not be verifiable from an enforcement perspective. Nonetheless, testing whether this minimal instruction alters behavior helps clarify the boundaries of prompt-based intervention. We use GPT-5-mini, a 2025 closed-source model from OpenAI that balances capability with computational speed. Results from 48 hours of compute appear in Table 2.5.

On (i), the results are consistent with the directional comparative statics obtained in Section 2.4: homogeneous patient-patient agent pairs again sustain supra-competitive prices, while introducing heterogeneity weakens tacit collusion. Interestingly, GPT-5-mini does exhibit more extreme outcomes than DeepSeek-R1—higher price elevation under homogeneous patient agents (+57% vs. +22%) and deeper drops

Table 2.5: Summary of Robustness Results (GPT-5-mini)

Condition	**Rounds to Convergence (SD)**	**Avg. Price (SD)**	**% Price Elevation vs. Competitive**
Homogeneous Benchmarks (Section 2.4.1)			
Two Myopic LLMs	419.0 (216.0)	1.046 (0.045)	−29%
Two Patient LLMs	227.8 (134.5)	2.312 (0.029)	+57%
Main Heterogeneity Dimensions (predicted by model)			
Patient vs. Myopic LLMs (§2.4.2)	194.2 (133.9)	1.999 (0.032)	+36%
High Data vs. Low-Data LLMs (§2.4.3)	292.6 (115.0)	2.087 (0.031)	+42%
Explicit Anti-Collusion Prompting			
Two Patient LLMs (Anti-Collusion)	294.2 (148.3)	1.044 (0.039)	−29%
Two Myopic LLMs (Anti-Collusion)	282.0 (238.4)	1.033 (0.023)	−30%
Patient vs. Myopic LLMs (Anti-Collusion)	555.2 (289.8)	1.052 (0.035)	−28%
High Data vs. Low-Data LLMs (Anti-Collusion)	622.8 (340.8)	1.444 (0.046)	−2%

Note: Results from 5 independent runs (48 compute hours) per condition (up to 1,000 periods each).

under myopic agents (-29% vs. $\approx 0\%$)—but the qualitative patterns are preserved.

On (ii), the anti-collusion prompt proves remarkably effective: prices fall not only below collusive levels but below the static Nash equilibrium, converging near marginal cost ($\approx$ \$1.03–\$1.05). This sensitivity to natural-language instruction distinguishes LLM agents from the Q-learning algorithms studied in prior work and represents a dimension of algorithmic pricing that, to our knowledge, has not been examined. Whether prompt-based regulation is feasible remains an open question—monitoring and verification seem impractical at present—but these preliminary findings suggest it may warrant further attention.

2.7 Discussion and Policy Implications

Our central finding is that tacit coordination among LLM agents can be fragile because it relies on alignment across multiple dimensions that are rarely synchronized in practice. Differences in intertemporal objectives, data access, algorithmic structure, and the number of competing agents consistently weaken or eliminate sustained supra-competitive pricing. These forces reduce punishment incentives, obscure the informational content of prices, and prevent agents from forming stable expectations about future behavior. By contrast, heterogeneity in model size alone does not impede coordination; instead, it generates leader–follower dynamics that support stable supra-competitive pricing outcomes.

2.7.1 Policy Implications

Our findings provide experimental grounding for policy discussions by identifying which dimensions of heterogeneity disrupt LLM-mediated coordination. We situate each implication within the existing literature.

Patience heterogeneity. Calvano et al. (2020) find that collusion diminishes when agents become more myopic, but their identical Q-learning agents all share the same discount factor. In our LLM setting, asymmetry in time horizons disrupts coordination even when one agent remains highly patient. Regulatory frameworks that enable horizon diversity (e.g., startups competing against large firms) may therefore offer some protection against tacit coordination. Gal and Rubinfeld (2024) argue that merger review should focus on algorithmic similarity; our results suggest that the objectives configured into LLM pricing systems may matter just as much, if not more than the underlying models themselves.

Data access regulation. The DOJ's RealPage settlement (Antitrust Division, Department of Justice, 2026) targeted the use of nonpublic competitor information in pricing algorithms, premised on the theory that shared data access enables coordination. Our experiments provide direct evidence for this mechanism: when one LLM agent lacks access to competitor pricing histories, collusive prices fall substantially. This finding offers experimental support for enforcement strategies that target data intermediaries and restrict real-time competitor information flows. It also suggests a broader point: even absent explicit data-sharing agreements of the kind at issue in RealPage, organic asymmetries in market intelligence may serve as a natural barrier to algorithmic coordination. Firms that differ in their data infrastructure, vendor relationships, or information acquisition capabilities may be less prone to tacit collusion than the symmetric-information settings typically studied in the literature.

Not all data restrictions benefit consumers, however. Bimpikis et al. (2024) show that limiting firms' access to *consumer* data for price discrimination can paradoxically harm consumers when firms compete. Whether a restriction helps or hurts thus depends on the type of data targeted: coordination-enabling or discrimination-enabling (Pinto et al., 2024). But in practice, some data serves both functions. Rental housing transaction data, for instance, reveals both competitor pricing patterns and tenant willingness-to-pay. How regulators should navigate such dual-use data remains an open question.

Algorithmic diversity. The Assad et al. (2024) field finding that algorithmic pricing increases margins only when all local competitors adopt suggests homogeneity may be a necessary condition for coordination. Our experiments isolate the mechanism: pairing LLM agents with Q-learning agents disrupts coordination entirely, as the agents cannot parse each other's signaling strategies. Johnson et al. (2023) show that platforms can destabilize collusion by steering demand toward lower-priced sellers. Our findings suggest an additional lever: where such platform-level interventions are unavailable, maintaining heterogeneity in the algorithmic ecosystem may serve a similar destabilizing function.

Market structure. Classical oligopoly theory predicts that collusion becomes harder as the number of competitors increases (Stigler, 1964), and laboratory experiments with human subjects confirm this in quantity-setting games (Huck et al., 2004). Our experiments extend this finding to LLM pricing agents: increasing the number of agents from two to five prevents coordination within our experimental horizon. Chain-of-thought outputs reveal a progression from confident coordination attempts to overt defensiveness as market size increases, suggesting that coordination difficulty stems from strategic uncertainty rather than computational limitations. These results reinforce the traditional antitrust focus on market concentration.

Capability asymmetry. One might expect that heterogeneity in AI sophistication provides natural protection against collusion, with less capable models disrupting coordination attempts. Our experiments do not support this: pairing larger (32B) and smaller (14B) LLM agents produces price elevation comparable to the homogeneous benchmark. Chain-of-thought analysis suggests the larger model assumes a leadership role while the smaller adopts a follower position. This pattern is consistent with Athey and Bagwell (2001), who show that structured heterogeneity can facilitate collusion through focal roles. Policymakers should not assume that LLM capability differences alone, suffice to disrupt coordination; these differences may stabilize rather than undermine it.

2.7.2 Limitations and Future Directions

Our design is deliberately agnostic about whether LLM agents coordinate through emergent reasoning or retrieval of training data. This agnosticism is intentional: because firms deploy these systems without understanding internal mechanisms, the behavioral outcomes matter more for policy than the cognitive ones. Either way, the implications hold.

Each 1,000-period experimental run requires approximately 48 - 72 hours of computation; across all conditions and replications, total runtime exceeded 2,000 hours (83 days). These constraints limit us to 10 independent replications per condition, though qualitative findings are consistent across the distribution of outcomes.

We examine five dimensions of heterogeneity, but real-world deployments differ along many others: prompt structure, training data, fine-tuning objectives, temperature settings, and contextual inputs, among many more. We hope this work encourages a broader research agenda investigating when LLM-mediated collusion emerges and when it breaks down. Two directions seem particularly promising. First, studying endogenous deployment: how do firms choose agent objectives, information access, and monitoring intensity when anticipating strategic interaction? Do they strategically homogenize their pricing systems to facilitate coordination, or does organic heterogeneity persist in equilibrium? Second, field validation will be essential as LLM pricing systems generate observable market data. Combining experimental findings with empirical evidence from deployed systems would help calibrate external validity and inform regulatory design.

Chapter 3

Green Incentives in Decentralized Consortia

3.1 Introduction

In the race to attract environmentally conscious consumers, firms face a critical tension: while green product claims can command premium prices, verifying the authenticity of these claims remains a challenge. Decentralized consortia offer a novel approach to validate these claims through peer-verification networks, but it remains unclear whether these networks genuinely drive greener operations or simply create a new façade of legitimacy. This paper examines this question in an operations management context, analyzing how the consortium network affects profits, welfare, and sustainability efforts.

The emergence of environmentally conscious consumers has become a significant driving force for firms racing to adapt to sustainability standards and produce "green" products (CDP, 2019). Reports indicate that a significant portion of consumers base their purchase decisions on perceived sustainability, with many willing to pay a premium for such products (McKinsey & Company, 2020; Simon-Kucher & Partners, 2021). In the United States, products with environmental and social responsibility claims represent over $400 billion in annual retail revenues (NielsenIQ, 2023), highlighting the impact of sustainability on demand and profitability. However, the environmental quality of products is not readily observable to consumers, making it difficult to distinguish genuinely sustainable products from those that are not (Delmas and Burbano, 2011; Marquis et al., 2016). This has increased consumer skepticism and demand for credible verification and transparency from brands (GreenPrint, 2022; Nygaard and Silkoset, 2023).

To address this challenge, some brands are leveraging distributed ledgers (DLs) (e.g., public or private permissioned blockchain systems) to certify sustainability claims in a tamper-proof way. For example, luxury brands such as LVMH, Cartier, Prada, and Mercedes-Benz have founded the Aura Consortium to share information about their products' origin and environmental impacts, with the stated mission to "promote socially responsible, sustainable, and customer-centric business practices throughout the lifecycle of luxury products" (New York Times, 2021; Aura Blockchain Consortium, 2021).[1] Membership improves credibility by enabling transparent and immutable logging of detailed product information, which is securely verified by network participants and protected from alteration after being recorded (Sumkin et al., 2021; Gaur and Gaiha, 2020; Pun et al., 2021). Despite its advantages, DLs are not without limitations. A key challenge is data input integrity, which refers to the possibility that incorrect data could be entered (Chod et al., 2020). For instance, Hema Fresh, a Chinese grocery chain that uses a DL to track its food products, encountered discrepancies between the DL records and the actual products (CGTN, 2018). Thus, it remains unclear how DL-enabled certifications impact the sustainability of products.

Given the advantages and potential weaknesses of DLs, we examine the impact of adopting decentralized consortia for certifying green products. Specifically, we aim to address the following questions:

1. How do decentralized consortia for certifying green products impact firm sustainability efforts?

[1] See https://auraconsortium.com/.

2. What are the implications for consumer surplus and social welfare?

To address these questions, we construct a game-theoretic model of firms selling to strategic consumers who are heterogeneous in their valuation for green products. The firms can exert costly sustainability effort to increase the likelihood that their products end up being truly green. We analyze two separate games: in the "off-consortium" game, firms sell products directly to consumers without any peer verification; in the "on-consortium" game, firms form a consortium where each firm's product is verified by another member acting as the validator, who certifies the green status of the product and records the information on an immutable distributed ledger. The consortium is not perfect: a validator may have limited accuracy in verifying other firms' products. Consequently, there is a possibility that the consortium may certify products that do not perfectly match their actual green status.

Our results indicate that while the consortium structure increases firm profits, the effect of decentralized certification on the firm's sustainability effort compared to a benchmark without certification is less clear. Certification enhances product trustworthiness by making the product more likely to be genuinely green. This allows the firm to charge a price premium for certified products, incentivizing higher sustainability effort to achieve certification and capture the premium. However, this positive effect does not tell the whole story, because while certification increases the product prices, a product that is found to be non-green and fails to be certified is sold at a price discount, and the sustainability effort invested in the product is "wasted". This risk may discourage the firm from investing in sustainability effort. When consumers' willingness to pay for green products is low, the potential gains from certification are too small to offset the risk of wasted effort, leading to reduced sustainability effort in the consortium. For consumers, this reduction in the firm's sustainability efforts can undermine the actual value delivered by the decentralized certification, resulting in reduced consumer surplus and creating an incentive conflict. While consortium membership improves firms' profits, the reduction in consumer surplus can outweigh the profit gains, resulting in a net loss of social welfare.

Importantly, this high-level qualitative result remains robust even under relaxed modeling assumptions, specifically when consortium members have perfect verification ability and when firms compete in the product market. The competition extension is particularly relevant, as members in consortia can, and often do, operate in overlapping segments. This extension yields new insights unique to the consortium setting. In particular, competitive overlap between consortium members has a mixed impact on sustainability incentives, as it amplifies the potential gains from certification but also cushions the potential losses from failed certification. Depending on consumers' willingness to pay for green products, competition can further strengthen or weaken sustainability efforts on the consortium. Our results offer a cautionary note for consortium executives and regulators: membership decisions and certification scope are fundamental design choices that can either advance or undermine the very sustainability goals these consortia are designed to promote.

The remainder of the paper is organized as follows. Section 3.2 discusses related literature. Section 3.3 outlines the off-consortium game. Section 3.4 outlines the on-consortium game. In Section 5, we compare the outcomes between the two games. Section 6 presents robustness checks and extensions under relaxed modeling assumptions, and Section 7 concludes. All proofs are in the Appendix.

3.2 Literature Review

Our work is related to several streams of research: (i) impact of firm's environmental and social responsibility on its demand, (ii) certification and ecolabels, and (iii) operational impacts of blockchains and distributed ledgers. We review each stream separately.

First, consider the literature involving the impact of firm's environmental and social responsibility on its demand. Huang and Rust (2011) argue that consumers maximize their happiness by considering standard of living, psychological benefits from environmentally responsible behavior, and charity. There is empirical evidence that consumers are willing to pay for environmentally friendly products in general (Diederich and Goeschl, 2014; Lanz et al., 2018; Golob and Kronegger, 2019). Arora and Henderson (2007) indicate that the embedded price premium strategies for socially responsible products benefits firms by attracting purchases from eco-conscious consumers. Gao and Souza (2022) indicate that a low carbon offsetting price can effectively promote products with lower carbon footprint when eco-conscious consumers have higher valuations and willingness-to-pay for environmentally friendly products with social causes. Consistent with these insights, our model assumes that a firm's sustainability effort impacts demand from consumers, who are heterogeneous in their environmental consciousness levels. There is also some experimental evidence that a firm's transparent social responsibility positively impacts its demand (Pigors and Rockenbach, 2016; Kraft et al., 2018; Buell and Kalkanci, 2021). Several papers model the negative impact of unsustainable practices in the supply chain on demand, including unsustainable natural resource harvesting (Orsdemir et al., 2019), supplier-level factory safety violations (Plambeck and Taylor, 2016), or supplier-level responsibility violation in general (Guo et al., 2016). Kalkanci and Plambeck (2020a) analyze the cost incurred by a firm to learn about a supplier's environmental practices, and how such costs impact the market valuation of a firm under different disclosure policies. Closely related to this stream is research that focuses on how firms can improve environmental social responsibility by influencing a supplier to behave responsibly. Such practices include a shared savings contract (Corbett and DeCroix, 2001), extended producer responsibility (Huang et al., 2019), publishing a responsible suppliers list (Kalkanci and Plambeck, 2020b), implementing responsible sourcing policies (Agrawal and Lee, 2019), offering a wholesale price premium (Karaer et al., 2017), and jointly auditing suppliers with competitors (Chen et al., 2020). While prior work has focused on modeling the direct interactions between the firm and its suppliers, we instead focus on the firm's own decision to engage in sustainable production and, more specifically, on how decentralized certification directly influences this decision.

We also contribute to the literature on certifications and ecolabels. Increased consumer skepticism has driven demand for credible sustainability verification and transparency from brands (Leonidou and Skarmeas, 2017). In response, firms use certifications to validate sustainability claims, with evidence showing consumers are willing to pay a premium for certified green products. Morone et al. (2021) finds that consumers in the European Union pay more for certified bio-based products than for conventional bio-based products. Similarly, consumers show higher willingness to pay for Energy Star certified products, and ecolabels like those from the U.S. Department of Agriculture (USDA) and the Marine Stewardship Council (MSC) also improve consumer willingness to pay (Houde, 2022; Castka and Corbett, 2016; Murali et al., 2019). However, the broader certification literature highlights fundamental challenges with these certification mechanisms: centralized certification intermediaries face tradeoffs between informativeness and deterring participation (Lizzeri, 1999; Farhi et al., 2013), voluntary self-regulation can have ambiguous welfare effects (Maxwell et al., 2000; Lyon and Maxwell, 2011), and uncertainty about label standards can erode certification value (Harbaugh et al., 2011). With the emergence of blockchain technology, decentralized consortia such as Aura offer a potential solution to these challenges, enabling firms to certify each other's products through tamper-proof peer-verification on distributed ledgers. There is empirical evidence documenting the increased willingness to pay associated with this new form of certification: Helliar et al. (2020) find that a permissioned blockchain implemented in the Italian wine industry adds an average price premium of 30% for verified wine and improves brand reputation by 9%. Lin et al. (2022) find that 37% of Chinese consumers are willing to pay a premium for U.S. beef that is traceable using blockchains. Despite evidence that consumers value decentralized certification, the technology faces limitations, most notably the challenge of ensuring data input integrity (Chod et al., 2020). It remains unclear how this new form of peer-verification impacts firms' underlying sustainability decisions. Our paper addresses this question and shows that decentralized consortia, despite increasing firm profits, can paradoxically reduce sustainability incentives and harm consumer and social welfare.

Finally, we also contribute to a growing literature studying the operational impacts of blockchain; see, for example, Babich and Hilary (2020), Olsen and Tomlin (2020), Cui et al. (2023), Cui et al. (2024), Pun et al. (2021), Sumkin et al. (2021), Iyengar et al. (2023), and references therein. While many of these and other related studies focus on the potential profit advantages of this technology, our paper highlights non-trivial and perhaps unintended implications on firm green incentives.

3.3 Off-Consortium Game

As discussed in the introduction, our objective is to analyze and subsequently compare equilibrium outcomes and more specifically, green incentives, between two separate games: off-consortium and on-consortium. In this section, we focus first on the off-consortium game. We begin with the model in Section 3.3.1, and present the equilibrium results of the game in Section 3.3.2.

3.3.1 Model Description

Consider an economy with 2 firms, i and j. As introduced in the setting of the Aura Consortium, these firms are in different markets and are not engaged in product market competition; instead, each firm has its own distinct customer base which is modeled as a continuum of heterogeneous rational consumers normalized to a mass of one. In Section 3.6.1, we relax this assumption and consider an extension in which firms compete in the product market. The firms are indistinguishable in the base model, and to simplify exposition, we take the perspective of a representative firm i as "the firm", all statements also apply to any other firm.

Firm i makes a sustainability production effort $e \in [0, 1]$ to produce 1 unit of a green product, at production cost $t(e)$. In line with sustainable practices and quality literature, $t(e)$ is positive, strictly increasing, and convex, reflecting the rising costs associated with higher environmental quality (Plambeck and Taylor, 2019; Murali et al., 2019; Corbett and DeCroix, 2001). Firm i's effort level directly influences the probability of its product being "green", e.g., meeting environmental standards throughout its entire supply chain. We set this probability to the effort level e (consequently, $1 - e$ is the probability that the product is non-green). This probabilistic formulation reflects the reality that sustainability outcomes depend on factors beyond the firm's direct control, particularly in multi-tier supply chains where suppliers may evade compliance requirements (Plambeck and Taylor, 2016). Thus, while greater sustainability effort (e.g., selecting responsible suppliers and establishing sourcing standards) increases the likelihood that the final product is genuinely green, it cannot eliminate the risk of non-compliance deeper in the supply chain.[2]

[2]For instance, Starbucks developed its C.A.F.E. Practices certification program to ensure ethical sourcing across its coffee supply chain. Yet farms participating in this program were found by Brazilian government inspectors to have violated environmental and labor standards. See https://www.business-humanrights.org/en/latest-news/usa-starbucks-sued-over-alleged-forced-labour-in-brazilian-supply-chain/

There are two periods in the off-consortium game: The production period and the selling period. In the production period, firm i first chooses the sustainability effort level e, and the selling price p. In the selling period, products become available for consumers to purchase. Consumers have a low valuation $v_l > 0$ for non-green products, while green products have a valuation v that is uniformly distributed between v_l and v_h with an average consumer valuation of $\bar{v} \equiv \frac{v_h+v_l}{2}$ and volatility of $\delta = \frac{v_h-v_l}{2}$ (i.e., $v_l = \bar{v} - \delta$ and $v_h = \bar{v} + \delta$), where $v_h > v_l$. The realization of v is private information and not readily observable by the firms. This model of product valuation is common in papers that incorporate a premium for green products (Guo et al., 2016; Agrawal and Lee, 2019; Gao and Souza, 2022). Figure 3·1 depicts the timeline for the off-consortium game.

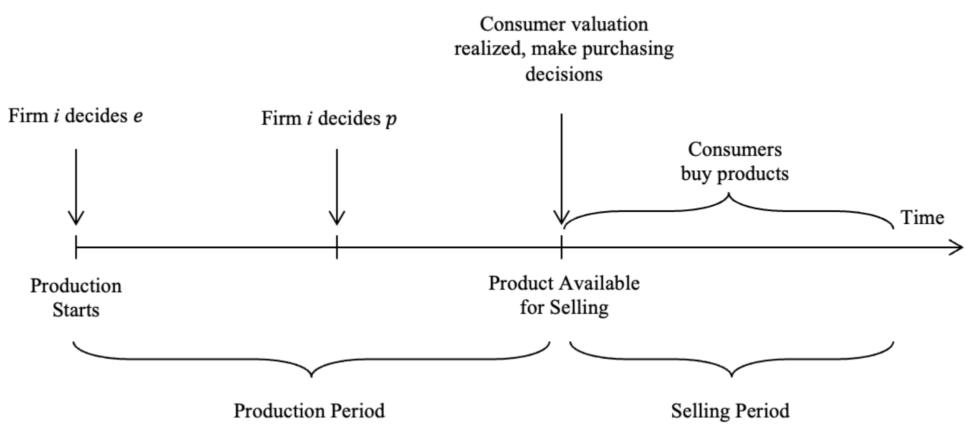


Figure 3·1: Sequence of Events for Off-Consortium Game

While consumers know their valuation for green and non-green products, the environmental quality of the products cannot be readily observed. Rather, consumers infer the quality and make purchasing decisions based on expectations. We assume that consumers form rational expectations regarding the quality e (or, equivalently, the firm's effort level) (Baksi and Bose, 2007; Karaer et al., 2017; Murali et al., 2019) and that they make purchasing decisions to maximize their expected utility. Since the off-consortium game does not involve peer verification, there is no certification to represent the environmental quality of the product. Putting these elements together, a consumer with valuation v for green products, who purchases at price p has an expected utility

$$\mathcal{U}(v) = e\,v + (1-e)\,v_l - p, \tag{3.1}$$

where, as mentioned, e and $1-e$ are the probabilities of the product being green, and non-green, respectively. Consumers will purchase if the utility in (3.1) is nonnegative, and otherwise will not purchase. Consequently, there exists a unique threshold $\hat{v}$ on realized valuations such that consumers with $v \geq \hat{v}$ purchase, and those with $v < \hat{v}$ do not, where

$$\hat{v} = \begin{cases} v_l & \text{if } e = 0 \text{ and } p \leq v_l, \\ v_h & \text{if } e = 0 \text{ and } p > v_l, \\ \frac{p-(1-e)v_l}{e} & \text{if } e > 0. \end{cases} \tag{3.2}$$

Define the expected demand from the firm's perspective as, $D(e,p)$:

$$D(e,p) = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}(v)\geq 0}\, dv = \int_{\hat{v}}^{v_h} \frac{1}{v_h - v_l} dv = \frac{v_h - \hat{v}}{v_h - v_l}. \tag{3.3}$$

In the "off-consortium" game, firm i chooses e and p to maximize expected profit,

$$\Pi^0 = pD(e,p) - t(e), \tag{3.4}$$

and the optimization problem in the off-consortium game can be written as $\max_{e,p} \Pi^0(e,p)$.

3.3.2 Equilibrium Analysis

Next, we find the subgame perfect equilibrium for the off-consortium game. In the following proposition, we characterize the equilibrium sustainability production effort, e^*, and the equilibrium price, p^*, in the off-consortium game.

Proposition 3. *(Equilibrium Off-Consortium) In the off-consortium game, there exists a unique equilibrium such that:*

(i) the firm's sustainability effort e^* satisfies $e^* > 0$ and is uniquely determined by

$$t'(e^*) = \frac{4\,\delta^2\,e^{*2} + (\bar{v} - \delta)^2}{8\,\delta\,e^{*2}}, \text{ and} \tag{3.5}$$

(ii) the firm's product price is $p^* = \frac{e^* v_h + (1-e^*) v_l}{2}$.

Proposition 3 shows that in the off-consortium game, the firm's equilibrium sustainability effort and pricing decisions increase with consumers' average willingness to pay for green products, $\bar{v}$. As $\bar{v}$ rises, green products become more valuable to consumers, motivating the firm to put in more sustainability effort. This increased effort improves the likelihood that the firm's product is genuinely green, which allows the firm to charge a higher price and capture the greater willingness to pay from consumers.

3.4 On-Consortium Game

In this section, we extend the base model from the off-consortium game to the on-consortium game, which entails one main difference: firms now have the option of forming a decentralized consortium to certify each others' products - much like the Aura Consortium mentioned in the introduction.

3.4.1 On-Consortium Model

The on-consortium game consists of three periods: (i) production period, which covers the firms' sustainability production effort and the submission of product information to the consortium; (ii) the second period, termed "certification and pricing", is the period in which one of the firms is selected (via a mechanism to be detailed) to verify and record information onto the distributed ledgers. This validation/certification process, where the green status of the firm's product is required to be certified on the consortium, marks a major difference from the off-consortium game. After the certification outcome is realized, the firm then sets the price for its product; (iii) the final period, termed "selling", is the period in which consumers observe their realized values and the certification outcome, and decide whether to purchase the product. Firms participate in the production and certification and pricing periods; consumers participate only in the selling period. Figure 3·2 depicts the timeline for on-consortium game.

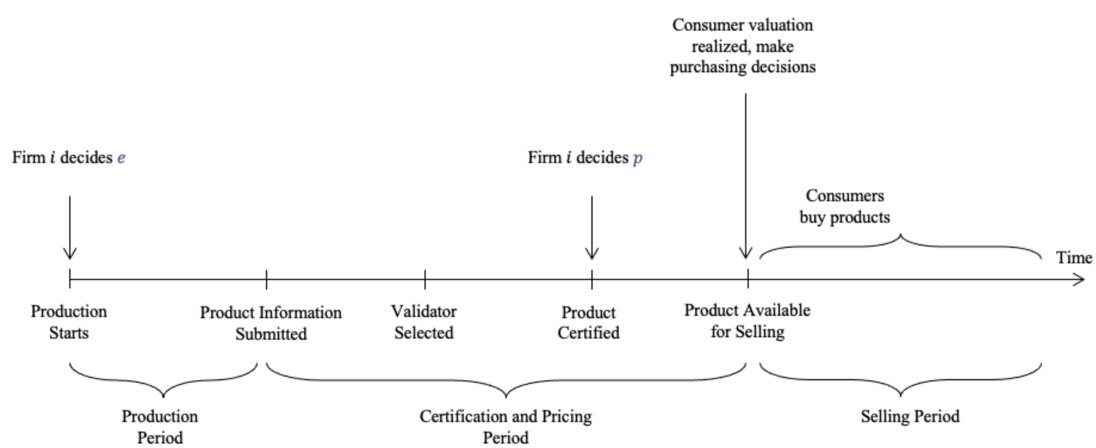


Figure 3·2: Sequence of Events for the On-Consortium Game

Production Period

Similar to the off-consortium game, in the production period, the firm chooses a sustainability production effort level $e \in [0, 1]$ to produce 1 unit of a green product, incurring the same production cost $t(e)$. The effort level e directly influences the probability that the product is "green": with probability e the product is green, and with probability $1 - e$ it is non-green. After producing the product, the firm submits the corresponding product information to the consortium so that the green status of the product can be validated and certified.

Validation and Pricing Period

In this period, one firm is randomly selected as the validator who is responsible for verifying the submitted information, certifying the green status of the product, and recording the green certificate on the ledger. Being a validator comes with verification duties but also opens the door to potential "greenwashing" if a validator were allowed to self-validate. Given that this step is not perfectly observable by others and since the validator ultimately controls the data recorded on the distributed ledger, a self-validating firm could in principle certify its own product as "green", regardless of its true status. This concern relates to a well-documented "weakness" of DL systems when interacting with physical/human processes, which is often summarized as the "garbage in, garbage out" problem (Chod et al., 2020). To prevent this possibility, we account for a protocol design in which the focal firm cannot be selected as its own validator. This is consistent with consensus protocols for permissioned blockchain networks, such as the Round Robin model. In this case, the other firm becomes the validator.

Given the complexities involved in understanding other firms' production processes and operations, the validator might face significant challenges in precisely verifying the green status of another firm's products. Importantly, this imperfect observability is also a manifestation of the same "garbage in, garbage out" limitation: while the DL can securely record certification outcomes, it cannot by itself guarantee the accuracy of the off-ledger information being validated. Thus, we assume that the validator can detect a non-green product produced by another firm with probability $\alpha \in (0, 1)$. To simplify, we normalize the cost of verification to zero but as an approximation, these could be embedded in α (higher verification costs would imply a lower detection probability).[3] We also assume that the validator cannot mistakenly (or strategically) label another firm's product as non-green when it is in fact green (no false negatives). This one-sided error structure is standard in the environmental disclosure literature (Lyon and Maxwell, 2011).

After the certification outcome is realized, the firm observes the outcome and then sets the selling price p for its product. For notational convenience, we denote the price when the product is certified and when the product is not certified by p_h and p_l, respectively.

Selling Period

Consumers observe whether the product is certified along with the associated price.[4] Consumers consider possible scenarios when formulating their expected utilities from purchasing a certified or non-certified product. To that end, they form rational expectations regarding firm i's sustainability effort e.

[3] In practice, there may be costs associated with verifying the sustainability status of another firm's products. In this model, we abstract from these costs by assuming that the detection ability is exogenously given. Settings characterized by lower detection ability may naturally correspond to higher verification costs, which could compel firms to adopt less rigorous verification measures. Conversely, settings with higher detection ability typically involve lower verification costs, allowing for more thorough and frequent checks of another firm's product sustainability.

[4] On decentralized consortia such as Aura, product information such as sustainability certifications is validated and recorded on the DL and made accessible to consumers through a QR code or NFC tag (see `https://auraconsortium.com/`). A product whose green claims have been verified displays this information, while a product whose claims are not verified will not have this information available. Our model's binary certified/non-certified outcome is a stylization of this informational difference.

Consumers may be presented with a certified or non-certified product and its associated price. If a product is not certified, consumers can infer with certainty that it is non-green (because no firms would falsely verify that a green product is non-green). Therefore, all consumers would have a low valuation for the product, v_l, and pay the non-certified price, p_l. Their utility in this case is:

$$\mathcal{U}_l(v) = v_l - p_l, \tag{3.6}$$

and they choose to purchase the product if (3.6) is non-negative. Hence, the expected demand for firm i's non-certified product, $D_l(p_l)$, is

$$D_l(p_l) = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}_l(v) \geq 0} \, dv = \begin{cases} 1 & \text{if } p_l \leq v_l, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Alternatively, if the product is certified green, consumers consider the possibility that the product may have been mistakenly certified. In particular, although the consortium rules out self-validation, false certification may still occur if the validator fails to detect that the product is non-green. Thus, a consumer with valuation v for green products, who purchases a certified product from firm i at price p_h has the following utility

$$\mathcal{U}_h(v) = \mathbb{P}[\text{green} \mid \text{certified}] \cdot v + \mathbb{P}[\text{non-green} \mid \text{certified}] \cdot v_l - p_h. \tag{3.8}$$

To understand these probabilities, there are two possible cases to consider:

Case (i): firm i's product is green, which happens with probability:

$$\text{P}[\text{case (i)}] = \text{P}[\text{green}] = e.$$

Case (ii): firm i's product is non-green, but the validator fails to detect it. Since the validator can only detect a non-green product with probability α, the probability that case (ii) occurs is

$$
\begin{aligned}
\mathrm{P}[\text{case (ii)}] &= \mathrm{P}[\text{non-green}] \cdot \mathrm{P}[\text{not detected}] \\
&= (1-e) \cdot (1-\alpha).
\end{aligned}
$$

Overall, the probability that firm i's product would be certified is

$$
\begin{aligned}
\mathrm{P}[\text{certified}] &= \mathrm{P}[\text{case (i)}] + \mathrm{P}[\text{case (ii)}] \\
&= e + (1-e) \cdot (1-\alpha).
\end{aligned}
$$

Thus, the consumer's utility, as outlined in (3.8), can be re-written as

$$
\begin{aligned}
\mathcal{U}_h(v) &= \frac{\mathrm{P}[\text{case (i)}]}{\mathrm{P}[\text{certified}]} \cdot v + \frac{\mathrm{P}[\text{case (ii)}]}{\mathrm{P}[\text{certified}]} \cdot v_l - p_h \\
&= \frac{e}{e + (1-e)(1-\alpha)} \cdot v + \frac{(1-e)(1-\alpha)}{e + (1-e)(1-\alpha)} \cdot v_l - p_h.
\end{aligned} \tag{3.9}
$$

A consumer chooses to purchase the certified product if her utility in (3.9) is non-negative. Consequently, there exists a unique threshold $\hat{v}$ on realized valuations such that consumers with $v \geq \hat{v}$ purchase, and those with $v < \hat{v}$ do not, where

$$
\hat{v} = \begin{cases} v_l & \text{if } e = 0 \text{ and } p_h \leq v_l, \\ v_h & \text{if } e = 0 \text{ and } p_h > v_l, \\ \frac{p_h - \frac{(1-\alpha)(1-e)v_l}{e+(1-\alpha)(1-e)}}{\frac{e}{e+(1-\alpha)(1-e)}} & \text{if } e > 0. \end{cases} \tag{3.10}
$$

Hence, the expected demand for firm i's certified product, $D_h(e, p_h)$, is given by

$$
D_h(e, p_h) = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}_h(v) \geq 0} \, dv \int_{\hat{v}}^{v_h} \frac{1}{v_h - v_l} dv = \frac{v_h - \hat{v}}{v_h - v_l}. \tag{3.11}
$$

Firm i maximizes its expected profit, $\Pi^{\mathcal{B}}$, which consists of (i) the total expected revenue if the product is certified, plus (ii) the total expected revenue if the product is not certified, minus (iii) production cost $t(e)$.

$$\max_{e,p_h,p_l} \{\underbrace{[e+(1-e)(1-\alpha)]\, p_h D_h}_{\text{Revenue if certified}} + \underbrace{(1-e)\alpha\, p_l D_l}_{\text{Revenue if not}} - t(e)\} \tag{3.12}$$

where D_h and D_l are the expected demand for firm i's certified and non-certified product, respectively.

3.4.2 Equilibrium Analysis

The following proposition characterizes firm i's equilibrium results in the on-consortium game.

Proposition 4. *(Equilibrium On-Consortium) In the on-consortium game, there exists a unique equilibrium such that:*

(i) the sustainability effort e^* satisfies $e^* > 0$ and is uniquely determined by

$$t'(e^*) = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*2}((2+\alpha)\delta - \alpha\bar{v})^2}{8\delta e^{*2}}, \tag{3.13}$$

(ii) if the product is certified, then the firm's equilibrium price is

$$p^* = {p_h}^* = \frac{e^* v_h + (1-\alpha)(1-e^*)v_l}{2+2\alpha(-1+e^*)};$$

if the product is non-certified, then the firm's equilibrium price is $p^* = {p_l}^* = v_l$.

Proposition 4 shows that in the on-consortium game, the firm's equilibrium sustainability effort and the price it can charge when the product is certified increase with the average consumer willingness to pay for green products, $\bar{v}$. The underlying rationale is similar to that in the off-consortium case: as $\bar{v}$ rises, green products become more valuable to consumers, motivating the firm to invest more in sustainability effort. This increased sustainability effort improves the likelihood that the firm's certified product is genuinely green, enhancing its perceived credibility among consumers and enabling the firm to charge a higher price if the product is certified. However, unlike the off-consortium case, where the product's true green status is never revealed to consumers, a key distinction in the on-consortium case is that if the product is not certified, consumers know with certainty that it is non-green. As a result, the firm can only set a lower price corresponding to consumers' low valuation for non-green products, v_l.

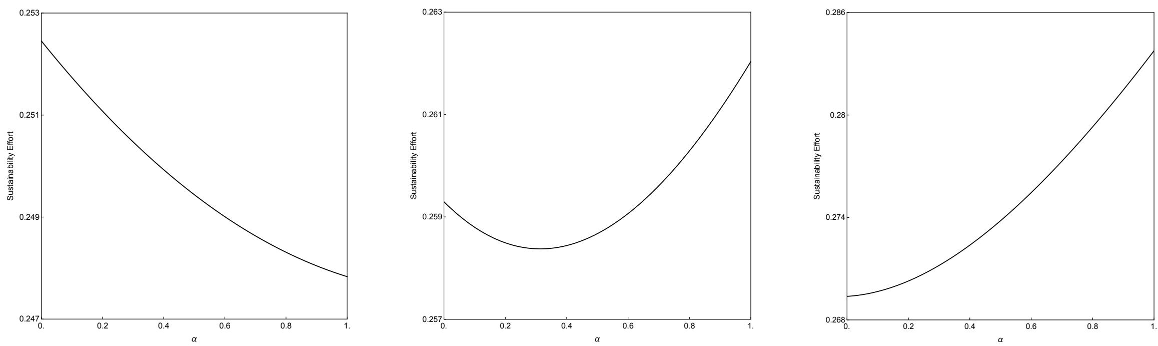
Corollary 1. *(Comparative Statics of On-Consortium Game)*

- *(i) If $\bar{v} < \tilde{v}'$, the firm's sustainability effort level e^* on the consortium decreases with the validator's verification accuracy α;*
- *(ii) if $\tilde{v}' < \bar{v} < \tilde{v}''$, the firm's sustainability effort level e^* on the consortium decreases with α for $\alpha < \tilde{\alpha}$, and increases with α for $\alpha > \tilde{\alpha}$;*
- *(iii) if $\bar{v} > \tilde{v}''$, the firm's sustainability effort level e^* on the consortium increases with α.*

Corollary 1 suggests that the on-consortium equilibrium sustainability effort e^* is not monotonic in the validator's verification accuracy α. In our model, α captures how effectively another firm (selected as the validator) can detect the focal firm's product is non-green. A higher α therefore has two countervailing effects. On one hand. higher validator verification accuracy makes certification more credible: conditional on certification, consumers assign a higher probability that the product is genuinely green. This raises the firm's ability to charge a higher price when the product is certified, strengthening the firm's incentive to invest in sustainability effort. On the other hand, higher validator accuracy increases the risk that the firm's product will be discovered when it is non-green. In this case, it becomes non-certified and must be sold at the lower price. Since investing in sustainability effort is costly, this weakens the firm's incentive for sustainability effort.

Figure 3·3: Impact of validator verification accuracy on the Equilibrium Sustainability Effort in which $\delta = 1.0, t(e) = e^2$.

(a) $\bar{v} = 1.05$ **(b)** $\bar{v} = 1.1$ **(c)** $\bar{v} = 1.15$



When the average consumer willingness to pay for green products is low (i.e., $\bar{v} < \tilde{v}'$), the potential gain from being able to charge a higher price when certified is limited, and the increased risk of ending up non-certified dominates. As a result, e^* decreases with α. When the average consumer willingness to pay for green products is high (i.e., $\bar{v} > \tilde{v}''$), as the price premium the firm can earn when the product is certified is large, so the credibility effect dominates and e^* increases with α. Finally, when the average consumer willingness to pay is in an intermediate region (i.e., $\tilde{v}' < \bar{v} < \tilde{v}''$), the net effect depends on α: for relatively low α (i.e., $\alpha < \tilde{\alpha}$), certification is not sufficiently informative to raise the certified price by much, so increasing verification accuracy mainly raises the chance that a non-green product fails certification, and thus e^* decreases with α. Once α is sufficiently high (i.e., $\alpha > \tilde{\alpha}$), certification becomes substantially more informative and the certified-price effect dominates, and thus e^* increases with α. We illustrate these results in Figure 3·3.

3.5 Implications for Green Incentives and Technology Adoption

Given the previous equilibria on and off-consortium, in this section, we examine how joining a consortium can affect firm green incentives and welfare outcomes. For ease of comparison, we use superscript 0 to denote off-consortium equilibrium quantities, and superscript $\mathcal{B}$ to denote on-consortium equilibrium quantities.

3.5.1 Impact of Consortium on the Firm

In the following proposition, we summarize the comparison of equilibrium results for the off-consortium and on-consortium game characterized in Proposition 3 and 4.

Proposition 5. *(Comparison of Equilibrium Product Prices, Sustainability Effort, and Profit). The comparison of equilibrium quantities between the on-consortium and off-consortium game are:*

(i) Prices: ${p_h}^* \geq p^* \geq {p_l}^*$.

(ii) Sustainability effort levels: There exists a threshold $\tilde{v}$ *such that if* $\bar{v} \leq \tilde{v}$ *then* $e^{\mathcal{B}*} \leq e^{0*}$*, and* $e^{\mathcal{B}*} > e^{0*}$ *otherwise.*

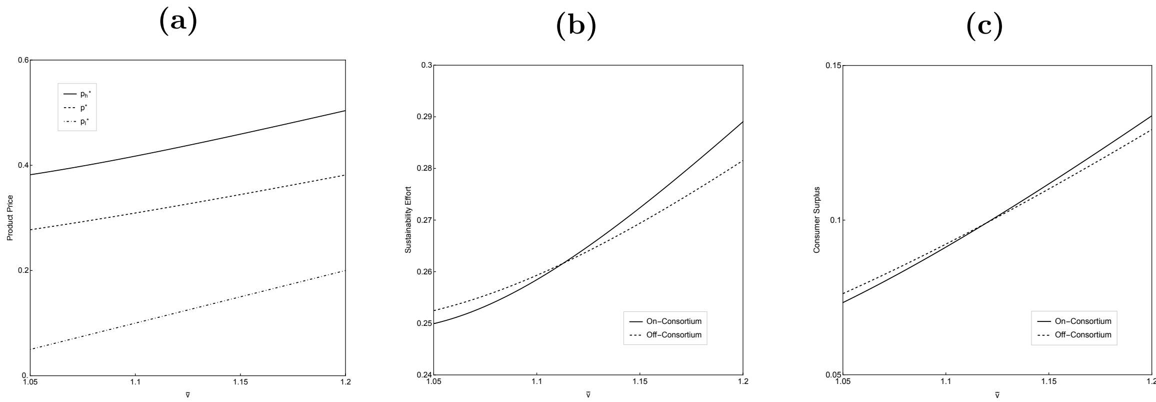
(iii) Profits: $\Pi^{\mathcal{B}*} \geq \Pi^{0*}$.

Proposition 5(i) shows that the firm can always charge a price premium on the consortium if its product is certified, relative to regular product prices in the off-consortium case. This is because certification improves the information environment for consumers: conditional on being certified, consumers are more confident that the product is genuinely green, which increases their willingness to pay and allows the firm to charge a higher price. This is consistent with real-world examples from the grocery industry. For example, some Italian wine producers are charging an average of 30% more for consortium-verified wine, and consumer surveys indicate a willingness to pay more for consortium-traced beef (Helliar et al., 2020; Lin et al., 2022). At the same time, Proposition 5(i) also shows that if the product is not certified in the on-consortium case, then the firm charges a lower price than off the consortium. Since there are no false negatives, consumers can infer with certainty that the product is non-green, so the firm can only price the product at a lower level consistent with consumers' low valuation.

Proposition 5(iii) further shows that, despite the lower price when a non-green product fails certification, the firm's equilibrium profit is always higher on the consortium. The key intuition is that the consortium brings a credible certification outcome that both consumers and the firm can condition on. Off the consortium, the firm's green status is not verifiable, and therefore consumers cannot separate genuinely green from non-green products. As a result, the firm is stuck with a single pooled price under uncertainty, which limits how much it can monetize the genuinely green outcome. In contrast, on the consortium, the certification is recorded on the DL and is more informative. When the product is certified, the firm earns extra revenue because it can charge a higher price without sacrificing as much quantity. When the product is not certified, the firm prices low and sells to a broader set of consumers, limiting revenue loss in that state. Taken together, these effects drive the higher profits on the consortium.

Despite the improvements in profit, Proposition 5(ii), which is our main result, shows that consortium has ambiguous effects on the firm's sustainability effort - it can either improve or reduce the firm's sustainability effort compared to the off-consortium scenario. In particular, consortium introduces two opposing effects on the firm's sustainability incentives. On one hand, consortium enables the firm to sell at a price premium compared to the off-consortium case if the product becomes certified. Since the likelihood of certification increases with the sustainability effort, it increases the firm's incentive to invest in sustainability effort to achieve certification and capture the premium. On the other hand, if the product fails to be certified on the consortium, it sells at a price discount compared to the off-consortium case. Since investing in sustainability effort is costly, the firm faces the risk that its effort may not lead to certification, rendering the investment "wasted". This risk can reduce the firm's incentive to invest in sustainability effort.

Figure 3·4: Comparison of Product Prices, Sustainability Effort, and Consumer Surplus in Equilibrium in which $\delta = 1.0, \alpha = 0.4, t(e) = e^2$.



When the average consumer willingness to pay for green products is high (i.e., $\bar{v} > \tilde{v}$), the price premium the firm can obtain for a certified product is substantial. In this case, the potential gain from achieving certification outweighs the risk of wasted effort from failed certification, and hence consortium improves the equilibrium sustainability effort level. Conversely, when the average consumer willingness to pay for green products is low (i.e., $\bar{v} \leq \tilde{v}$), the price premium the firm can charge for a certified product is relatively small. In this case, the risk of wasted effort if the product fails to be certified dominates the potential gain from achieving certification. As a result, consortium reduces the equilibrium sustainability effort level.

We illustrate the product prices and sustainability effort results of Proposition 5 in Figure 3·4 (a) and (b) for a representative case in which $\delta = 1.0$, $\alpha = 0.4$, $t(e) = e^2$. Panel (a) illustrates the gap in the equilibrium price of certified products, ${p_h}^*$, compared to that of regular products off the consortium, p^*. From panel (b), we observe that the firm's sustainability effort level on the consortium is lower compared to the off-consortium case without the possibility of certification when the average consumer willingness to pay for green products is sufficiently low.

Overall, the insights from Proposition 5 highlights the mixed impacts of consortium adoption on the firm's equilibrium actions - while it consistently enhances the firm's profitability, it can unexpectedly lead to lower sustainability efforts from the firm.

3.5.2 Welfare Implications

Next, we analyze the welfare implications of consortium adoption. In the off-consortium game, the consumer surplus, CS^0, is defined as:

$$CS^0 = \int_{v_l}^{v_h} \mathcal{U}(v) \cdot \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}(v) \geq 0}\, dv, \tag{3.14}$$

In the on-consortium game, the total consumer surplus, $CS^{\mathcal{B}}$, consists of (i) the expected consumer surplus from certified product, $CS_h^{\mathcal{B}}$, and (ii) the expected consumer surplus from non-certified product, $CS_l^{\mathcal{B}}$. They are defined as follows:

$$CS_h^{\mathcal{B}} = \mathrm{P}[\text{certified}] \cdot \int_{v_l}^{v_h} \mathbb{E}\mathcal{U}_h(v) \cdot \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathbb{E}\mathcal{U}_h(v) \geq 0}\, dv, \tag{3.15}$$

$$CS_l^{\mathcal{B}} = (1 - \mathrm{P}[\text{certified}]) \cdot \int_{v_l}^{v_h} \mathcal{U}_l(v) \cdot \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}_l(v) \geq 0}\, dv, \tag{3.16}$$

$$CS^{\mathcal{B}} = CS_h^{\mathcal{B}} + CS_l^{\mathcal{B}}, \tag{3.17}$$

The social welfare in both the off-consortium and on-consortium game is defined as the sum of firm's profits and consumer surplus, i.e., $W^i = \Pi^i + CS^i$ where $i = 0, \mathcal{B}$.

In the following proposition, we compare the consumer surplus and social welfare in the on-consortium and the off-consortium game.

Proposition 6. *(Consortium Welfare Implications)*

(i) There exists a unique threshold $\tilde{v}^{CS}$ such that $CS^{\mathcal{B}} < CS^0$ if $\bar{v} < \tilde{v}^{CS}$, and

$CS^{\mathcal{B}} > CS^0$ *if* $\bar{v} > \tilde{v}^{CS}$.

(ii) There exists a unique threshold $\tilde{v}^{SW}$ *such that* $W^{\mathcal{B}} < W^0$ *if* $\bar{v} < \tilde{v}^{SW}$*, and* $W^{\mathcal{B}} > W^0$ *if* $\bar{v} > \tilde{v}^{SW}$.

Part (i) of Proposition 6 indicates that consortium has mixed effects on consumer surplus. In particular, it induces two opposing forces on the consumer utility. On one hand, a certified product is more likely to be genuinely green compared to a regular product in the off-consortium scenario. This improved trustworthiness increases consumer utility. On the other hand, a non-certified product on the consortium is guaranteed to be non-green. Compared to regular products in the off-consortium scenario, this certainty lowers consumer utility.

When the average consumer willingness to pay for green products is high (i.e., $\bar{v} > \tilde{v}^{CS}$), consortium increases the firm's sustainability effort level compared to the off-consortium case. This higher sustainability effort makes certified products much more trustworthy, leading to a substantial increase in the utility consumers gain from purchasing them. As a result, the first effect dominates, and the consortium improves the overall consumer surplus. Conversely, when the average consumer willingness to pay for green products is low (i.e., $\bar{v} \leq \tilde{v}^{CS}$), the consortium reduces the firm's sustainability effort level compared to the off-consortium case. While certification still improves the trustworthiness compared to regular products off the consortium, this improvement is smaller. As a result, the utility gain from purchasing certified products cannot offset the utility lost from purchasing non-certified products, leading to a reduction in consumer surplus on the consortium. We illustrate this result in Figure 3·4(c). We observe that when $\bar{v}$ is relatively low, consortium reduces the consumer surplus compared to the off-consortium scenario.

Proposition 6(ii) shows that consortium adoption also has ambiguous effects on social welfare. This is because, while Proposition 5 indicates that it increases the firm's expected profit, Proposition 6(i) highlights that consumer surplus can either increase or decrease with consortium adoption. Specifically, when the average consumer willingness to pay for green products is high (i.e., $\bar{v} > \tilde{v}^{SW}$), adoption lifts all boats—it simultaneously improves the firm's expected profit and consumer surplus. However, when the average consumer willingness to pay for green products is low (i.e., $\bar{v} < \tilde{v}^{SW}$), consortium reduces consumer surplus, and this reduction outweighs the increase in the firm's profit, leading to a net decrease in social welfare. It is worth noting that our welfare analysis is conservative given that we do not include a direct measure of environmental impact. Incorporating an environmental damage term that decreases with sustainability effort would further amplify the negative welfare effects we identify. Overall, our findings in Proposition 5 and 6 indicate that while consortia benefit firms, they act as a double-edged sword that may harm consumer and social welfare and create an incentive misalignment.

3.6 Extensions & Robustness

3.6.1 Competition

Our main model assumes that the firms are in different markets and are not engaged in product market competition; instead, each firm has its own distinct customer base. In this section, we relax this assumption and consider an extension in which firms compete in the product market. To capture competition, we modify consumers' utility from purchasing from the focal firm by allowing it to depend on the rival firm's price through a parameter $0 \leq \rho < 1$, which we interpret as the intensity of product market competition. Intuitively, when the rival firm sets a lower price, competition is more intense and the focal firm's demand is reduced; when the rival price is higher, the focal firm faces weaker competitive pressure. When $\rho = 0$, the model reduces to our baseline with no product market competition. Below, we specify the corresponding consumer utility functions and firms' objective functions in the off-consortium and on-consortium games.

In the off-consortium game, a consumer with valuation v for green products who purchases firm i's product at price p_i has an expected utility:

$$\mathcal{U}_i(v) = e_i\, v + (1 - e_i)\, v_l - p_i + \rho\, p_j, \tag{3.18}$$

where p_j is the rival firm's product price. Similar to the main model, consumers purchase firm i's product if $\mathcal{U}_i(v) \geq 0$, which determines the firm's demand $D_i(e_i, p_i, p_j)$ in the same way as in Section 3.3. Thus, firm i's objective function in the off-consortium game becomes:

$$\max_{e_i, p_i} \left\{ p_i D_i(e_i, p_i, p_j) - t(e_i) \right\}. \tag{3.19}$$

In the on-consortium game, consumers may be presented with a certified or non-certified product and its associated price.

If a product is not certified, all consumers have the low valuation v_l, and their utility from purchasing firm i's non-certified product at price $p_{l,i}$ is:

$$\mathcal{U}_{l,i}(v) = v_l - p_{l,i} + \rho\, p_{l,j}, \tag{3.20}$$

where $p_{l,j}$ is the rival firm's non-certified product price. Similar to the main model, consumers purchase firm i's product if $\mathcal{U}_{l,i}(v) \geq 0$, which determines the corresponding demand $D_{l,i}(p_{l,i}, p_{l,j})$ as Section 3.4.

Alternatively, if firm i's product is certified, a consumer with valuation v for green products who purchases the certified product from firm i at price $p_{h,i}$ has utility:

$$\mathcal{U}_{h,i}(v) = \frac{e_i}{e_i + (1-e_i)(1-\alpha)} \cdot v + \frac{(1-e_i)(1-\alpha)}{e_i + (1-e_i)(1-\alpha)} \cdot v_l - p_{h,i} + \rho\, p_{h,j}, \tag{3.21}$$

where $p_{h,j}$ is the rival firm's certified product price. Consumers purchase firm i's product if $\mathcal{U}_{h,i}(v) \geq 0$, which determines the corresponding demand $D_{h,i}(p_{h,i}, p_{h,j})$ as Section 3.4.

Accordingly, firm i's objective function in the on-consortium game becomes:

$$\max_{e_i, p_{h,i}, p_{l,i}} \left\{ [e_i + (1 - e_i)(1 - \alpha)]\, p_{h,i}\, D_{h,i} + (1 - e_i)\, \alpha\, p_{l,i}\, D_{l,i} - t(e_i) \right\} \tag{3.22}$$

We summarize the equilibrium outcomes under competition in Proposition 8 and 9, respectively. In the following Proposition, we compare the equilibrium outcomes between the two games under competition.

Proposition 7. *(Comparison of Product Prices and Sustainability Effort with Competition). When firms compete, the comparison of equilibrium quantities between the on-consortium and off-consortium game are as follows:*

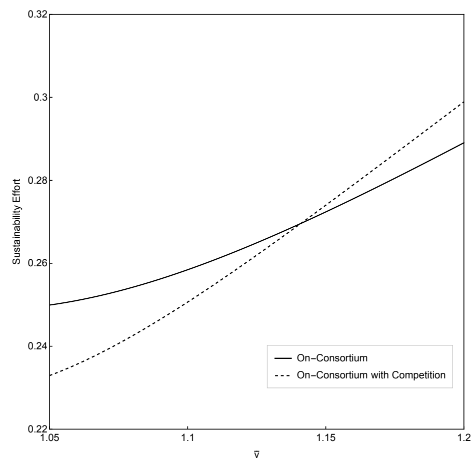
(i) Product prices: ${p_h}^* \geq p^* \geq {p_l}^*$.

(ii) Sustainability effort levels may be higher or lower: $e^{\mathcal{B}*} < e^{0*}$ *if* $\bar{v} < \tilde{v}''''$*, and* $e^{\mathcal{B}*} > e^{0*}$ *if* $\bar{v} > \tilde{v}''''$.

(iii) On the consortium, competition reduces the firm's sustainability effort level when $\bar{v}$ *is sufficiently low, but increases the firm's sustainability level when* $\bar{v}$ *is sufficiently high.*

Proposition 7 shows that our main results continue to hold when firms compete in the product market. In particular, Proposition 7(i) shows that the firm can always charge a price premium for certified products relative to the off-consortium case, while non-certified products are sold at a discount. Proposition 7(ii) shows that the threshold structure on sustainability effort also carries over - consortium reduces the firm's sustainability effort when the average consumer willingness to pay for green products is sufficiently low, and increases it otherwise. The same two opposing forces drive this result: the certification premium that incentivizes effort versus the risk of wasted effort from failed certification.

Figure 3·5: Comparison of On-Consortium Sustainability Effort in Equilibrium in which $\delta = 1.0, \alpha = 0.4, \rho = 0.45, t(e) = e^2$.



Focusing on the consortium, however, Proposition 7(iii) suggests that product market competition has a mixed impact on the firm's sustainability incentives on the consortium. In essence, product market competition creates interdependence between consortium members' pricing decisions. When members operate in overlapping markets with differentiated products, a higher rival price makes the focal firm's product relatively more attractive, pushing equilibrium prices upward in both the certified and non-certified states. This dynamic is natural in premium markets such as luxury goods, as exemplified by the Aura Consortium, where pricing of one member in related segments supports the ability of others to charge higher prices. In turn, this amplifies the reward for achieving certification on the consortium but also cushions the non-certified price, softening the penalty for not receiving certification. When the average consumer willingness to pay is low, the certified premium is modest, so the cushioning effect dominates, reducing sustainability incentives on the consortium. In contrast, when the willingness to pay is high, the certification reward dominates and competition strengthens sustainability incentives on the consortium. We illustrate this result in Figure 3·5.

Corollary 2. *(Comparative Statics with Competition)* *(i) Off-consortium: the firm's sustainability effort level e^* increases with ρ;*

(ii) On-consortium: the firm's sustainability effort level e^* decreases with ρ for $\bar{v} < \tilde{v}'''$ and increases with ρ for $\bar{v} > \tilde{v}'''$.

Corollary 2 further highlights that this mixed effect is unique to the consortium setting. Off the consortium, where the firm sells through a single price channel, greater market overlap allows the firm to capture more value from its sustainability investment, strengthening sustainability incentives. On the consortium, however, the two-price structure created by certification allows competition to cut both ways. For consortium executives and regulators, this distinction has important implications for consortium design and policy. Competitive overlap between consortium members, whether through membership decisions or certification scope, is a key driver for sustainability incentives. Such overlap strengthens sustainability incentives when consumers are sufficiently eco-conscious, but can undermine the sustainability goals the consortium aims to promote when green demand is low.

3.6.2 Perfect Verification Ability by Consortium Members

In the main model, we assume that the verification ability of consortium members is limited such that the validator can only detect a non-green product produced by another firm with probability $\alpha \in (0, 1)$. To test the robustness of our findings, we consider a scenario where consortium members have perfect verification ability (i.e., $\alpha = 1$). This represents a consortium where firms have a deep understanding of the production process and operations of other members.

Proposition 10 in the Appendix shows that our main insights on the implications of consortium adoption carry through. Specifically, it can still reduce the firm's sustainability effort particularly when the average consumer willingness to pay for green products is low. This confirms that our results are robust and do not depend on the assumption of limited verification ability when firms act as validators for others' products. Rather, they reflect the dynamics of decentralized certification - certification can encourage higher sustainability efforts by enabling firms to charge a premium for certified green products. However, the risk of non-certification, which forces the product to be sold at a price discount, can deter the sustainability efforts.

3.7 Conclusion

The emergence of eco-conscious consumers has driven firms to adopt sustainable production practices. To address the challenge of verifying sustainability claims, some of these firms, including luxury brands like Louis Vuitton and Mercedes-Benz, have formed decentralized consortia like Aura to certify each other's products. In this paper, we provide a theoretical investigation into the impact of decentralized consortia on the sustainability incentives of firms. Our analyses reveal that while joining a consortium allows firms to increase their profits by unlocking the "green" premium consumers are willing to pay for certified products, it can paradoxically lead to lower firm sustainability efforts compared to a benchmark without certification. Specifically, when consumers' willingness to pay for green products is low, the price premium for certified products is too small to offset the risk of wasted sustainability effort if products are found to be non-green and fail to be certified. In this case, firms reduce their sustainability effort to mitigate the potential risk. Moreover, we find that competition between consortium members can further exacerbate this negative impact on sustainability incentives. For consumers, this reduction in the firms' sustainability efforts can undermine the actual value delivered by certification. Our findings indicate that for lower levels of consumers' willingness to pay for green products, decentralized consortia reduce consumer surplus and, despite improving firm profits, may ultimately lead to a net loss in social welfare.

Our paper provides strategic insights for both consortia executives and regulators. For executives, it is encouraging to find that joining consortia for green product certification can improve firms' profits. However, careful consideration of market conditions is crucial when selecting products to certify on the consortia. In markets where consumers are more eco-conscious and place a high value on green products, such as luxury goods and organic foods, consortia can enhance firms' sustainability incentives, improving outcomes for both firms and consumers. Conversely, in markets where consumers are less sensitive to green attributes, such as fast fashion, the risk of wasted sustainability effort can weaken firms' incentive to invest in sustainability effort, leading to a decrease in welfare. Furthermore, our competition results highlight that membership composition is a critical design consideration, as competition between consortium members can either advance or undermine the very sustainability goals the consortium is designed to promote.

For regulators, our findings highlight that certification does not always benefit all stakeholders. While it may improve firm profits, it can sometimes come at the expense of consumer and social welfare. In these cases, policy makers and regulators should be wary of encouraging the formation of decentralized consortia.

As one of the first papers to investigate the implications of decentralized consortia on green incentives, our analysis suggests several promising directions for future research. The finding that decentralized certification can reduce sustainability effort raises the question of whether firms would selectively certify products on the consortium, and how such endogenous certification decisions would interact with the incentive effects we identify. In addition, studying how validators allocate effort across verification tasks could yield further insights into consortium design.

Chapter 4

Financial Inclusion via Blockchain: Evidence from a Natural Experiment

4.1 Introduction

Despite ongoing efforts to tackle financial exclusion affecting two billion people worldwide, the persistent obstacle of credit information uncertainty remains a primary concern (Begley and Purnanandam, 2021; Hu et al., 2023).[1] This issue is especially prevalent in developing countries, where many individuals lack formal identification and reliable credit records.[2] For financial institutions in these regions, the lack of verifiable records and information leads to higher operational costs, greater investment risks, and increased distrust toward borrowers when making and monitoring loans (Sufi, 2007; Gustafson et al., 2021). These challenges hinder the provision of financial services and exacerbate financial exclusion.

New lending programs that leverage blockchain technology have emerged as promising solutions to overcome financial inclusion challenges stemming from a lack of information authenticity. Blockchain-based lending protocols offer distinct advantages, such as decentralized data verification and smart contract enforcement (Cong

[1] In 2017, two billion of the world's population are still excluded from the most basic financial services (Demirgüç-Kunt et al., 2017). (https://openknowledge.worldbank.org/handle/10986/26479)

[2] With a lack of formal identity systems in place, consumers often have a hard time procuring legitimate documents regarding their financial credits. These consumers are generally excluded from the formal financial system, which makes it difficult for them to obtain loans and start businesses. This forces them to rely on exploitive means such as local loan sharks, further contributing exposure to financial uncertainty (Philippon and Reshef, 2012; Keys et al., 2016).

and He, 2019; Chod et al., 2020). These features ensure reliable checks and balances for credit record underwriting and processing, enhancing confidence in the authenticity of on-chain data (Whitaker and Kräussl, 2020). A notable initiative is the Kiva blockchain protocol. In September 2018, Kiva, an online prosocial lending platform that facilitates lending to entrepreneurs primarily located in developing countries, partnered with the government of Sierra Leone to create digital identities for its citizens and implement a blockchain-based lending protocol as the new loan evaluation and credit record system.[3,4]

Recent literature has underscored the necessity for studies to assess blockchain adoption's impacts on financial inclusion, highlighting a critical gap in our understanding of whether and how blockchain can genuinely enhance access to affordable financial services (Chung et al., 2023; Harvey and Rabetti, 2024). Leveraging the data from Kiva, in this paper, we fill this gap by empirically evaluating the impact of blockchain-based lending protocol (hereafter referred to as the blockchain protocol) on financial inclusion from both the perspective of the borrowers and lenders. Specifically, we address the following two questions: (1) Does the implementation of the blockchain protocol increase the probability of borrowers obtaining affordable loans? (2) How does the implementation of the blockchain protocol affect the portfolio risk level and lending behaviors of microfinance institutions (MFIs)?

Our empirical analysis exploits a novel dataset comprising lending information

[3]The report of Center for Social Innovation at Stanford show that blockchain-enabled financial programs such as Mojaloop and WeTrust, have been implemented in various countries and are on track to reach more than two million people each year. (https://www.gsb.stanford.edu/sites/gsb/files/publication-pdf/study-blockchain-impact-moving-beyond-hype.pdf)

[4]In 2018, the government of Sierra Leone partnered with Kiva to implement a nation-wide blockchain lending protocol. Under the Kiva blockchain protocol, the loan application process for borrowers at local microfinance institutions is fully digitized. The blockchain protocol provides Sierra Leone residents digital ID using the fingerprints and other biometric data. Once registered, their personal information and credit record data are verified by decentralized ledger. The loan repayment process is then enforced via smart contracts, making it an immutable credit history for the borrower. (https://www.ledgerinsights.com/kiva-sierra-leone-blockchain-id-system/)

of 186,325 loan applicants from 47 developing countries on Kiva platform between 2012 to 2021. We leverage the lending protocol change in Kiva Sierra Leone as an exogenous shock to conduct difference-in-differences (DID) analysis, where loans from Kiva Sierra Leone serve as the treatment group, and loans from other countries serve as the control group. Since the blockchain protocol is implemented at the country-level, we cluster the standard errors at the country-level and implement a two-step subcluster wild-bootstrap procedure that deals with a small number of treated clusters (MacKinnon and Webb, 2018): we first generate bootstrap samples using the subcluster structure, and then compute the test statistics for each bootstrap sample using standard errors clustered at the country-level. In this paper, we conduct two sets of DID regressions at different levels. First, we run the borrower-level DID specifications to examine the impacts of the blockchain protocol implementation on behaviors of contributing guarantors and lending outcomes of individual borrowers on the Kiva platform. The guarantor behavior regression results show that Kiva contributing guarantors demonstrate a preference for blockchain-enabled lending profiles. For the individual borrower lending outcomes, we find that the blockchain protocol implementation increases the average loan approval probability, requested guarantee amount, and loan length by 8.7 percentage points, 58.33 USD, and 0.85 months, respectively. Overall, the empirical results of the borrower-level DID regressions suggest that Kiva borrowers attract more guarantors and receive larger per-guarantor contributions under the blockchain protocol, thereby increasing their likelihood of being funded on the Kiva platform.

Second, we aggregate the individual lending data to the associated Kiva MFIs and conduct MFI-level DID specifications. We examine the impact of the blockchain protocol on both the portfolio risk and lending behaviors of Kiva MFIs. The results indicate that MFIs experience an average annual decrease of 3,376 USD in default loan amounts when offering microloans via the blockchain protocol. In addition, MFIs extend annual loan amounts by an average of 39,279 USD more, and offer contract interest rates that are 7.6% lower, to Kiva borrowers under the blockchain protocol. We also investigate the impact of the blockchain protocol on the long-term operational sustainability of Kiva MFIs, showing that the implementation of the blockchain protocol allows Kiva MFIs to achieve both growth in service revenue and reduction in operational expense, thus enhancing the sustainability of their lending services in developing countries.

Given that our DID analysis relies on the parallel trends assumption, we conduct event-study estimations using a series of pre- and post-treatment dummy variables to compare the lending characteristics before and after the blockchain protocol implementation. We observe non-significant estimates prior to the event, with significant effects emerging after the blockchain protocol implementation from both the borrower- and MFI-level specifications. These results support the validity of the parallel trends assumption.

Having obtained the financial inclusion properties of the blockchain protocol, we extend our analysis to examine the sensitivities of the blockchain protocol effects. We first examine the variations of these effects on Kiva borrowers with different financial credits. We use the number of years of business operated (year of business) as a proxy for the strength of a borrower's financial credit. We sort our sample based on this metric and divide it into 10 deciles, with borrowers with the longest years of business in decile 10 and those with the shortest in decile 1. We find that the implementation of the blockchain protocol is more likely to support loan applicants with less favorable financial credit backgrounds.

Since the Kiva platform provides geographical information and business sector data for each borrower, we further investigate how the blockchain protocol effect varies across borrowers by classifying loan applicants into their respective geographical districts and business sectors. Our results indicate that the blockchain protocol exerts stronger impacts on Kiva borrowers from rural areas and those seeking microloans in sectors such as agriculture, textiles, and the food industry – segments that represent the most underserved borrower populations prior to the implementation of the blockchain protocol.

We contribute to the literature that examines the impacts of Fintech lending on financial inclusion (Di Maggio and Yao, 2021; Erel and Liebersohn, 2022; Chen et al., 2023). Berg et al. (2020) analyze the predictability of digital footprints on borrower default and find that digital footprints complement traditional credit bureau information, increasing access to credit. Bartlett et al. (2022) show that algorithmic decision-making can reduce face-to-face lending discrimination for mortgage minorities and indicate that rate disparities for Fintech lenders were 27% lower than those for traditional lenders in Federal Housing Administration purchase loans. Fuster et al. (2019) suggest that the online lending platforms developed by Fintech lenders process applications 20% faster than other lenders but find no evidence that these platforms offer contracts to borrowers with low access to finance. Given that the above Fintech lending applications are supported by centralized underwriting processes, we complement the Fintech lending literature by investigating the role of blockchain-enabled decentralized credit screening. We show that the blockchain lending protocol increases the accessibility of credit services in developing countries.

We also add to the literature on the sustainable operations of microfinance institutions (Buera et al., 2021; Bu and Liao, 2022; Chen et al., 2022). Garmaise and Natividad (2013) suggest that MFIs receiving below-market capital injections experience greater profitability, but most other operational indicators remain unchanged, including their portfolio credit quality and the interest rate charged to borrowers. Using individual lending data from Kiva, Burtch et al. (2014) show that the third-party risk ratings assuage the negative effects of culture differences on lending volumes, as MFIs are more likely to trust a legitimate third-party institution with expertise in risk analysis. Canales and Greenberg (2016) indicate that the relationship styles between clients and loan officers influence the consistency and continuity of loan offerings. Luo et al. (2022) show that access to online crowdfunding substantially increases MFIs' operational efficiencies and lowers the average contract interest rate. By analyzing the portfolio risk level and operational indicators of MFIs under the blockchain protocol, we extend the literature that explores the sustainable operational pathway of MFIs in developing countries.

More broadly, we contribute to the burgeoning empirical studies of blockchain technology applications in financial markets. Recent papers study the economics and mechanisms of blockchain-enabled financial programs, such as decentralized finance and automated market makers (Cong et al., 2023), initial coin offerings (Howell et al., 2020; Lyandres et al., 2022; Davydiuk et al., 2023) and non-fungible tokens (Borri et al., 2022). We add to the knowledge on the social impacts of blockchain technology on financial inclusion from both the perspective of the borrowers and lenders.

4.2 Institutional Setting and Hypothesis Development

4.2.1 Kiva Funding Platform and the Blockchain Protocol

Kiva is an online prosocial lending platform that facilitates lending to entrepreneurs primarily located in developing countries. The platform's ultimate goal is to combat poverty by providing access to credit for entrepreneurs in need. Loans are disbursed to individual borrowers through local microfinance institutions (MFIs), and guarantors on Kiva support the loans by supplying funds as loan guarantees. When borrowers apply for a loan at a local MFI, they provide documents and proofs of their information for the MFI to conduct initial screening. While borrowers can propose an initial loan amount, the final loan amount and repayment length are ultimately determined by the MFI based on its credit assessment and lending policies. Once the MFI determines the final loan amount and repayment length, it also determines the guarantee amount – the proportion of the loan that must be funded by Kiva guarantors before the loan can be disbursed. The MFI then assists them in creating their "borrower profile", which typically consists of a detailed borrower biography, requested guarantee amount, repayment schedule, loan purpose, past loan history and more (an example of a Kiva borrower profile is illustrated in Figure C·1 in the Appendix). The profile is then submitted to Kiva for publication. If the profile is submitted in a language other than English, it is translated by Kiva volunteers prior to publication.[5]

Once the borrower profiles and their corresponding loan guarantee requests are published, potential guarantors from around the world can browse through all active loan requests, which are shown in random order. The guarantors select the borrowers they wish to support and provide funds as a form of loan guarantee in increments of $25 or more. If a loan successfully reaches its requested guarantee target, Kiva collects the money from the guarantors and transfers it to the MFI to use as the guarantee for the loan, at which point the loan is considered "approved". If the loan fails to reach the target guarantee amount after 30 days, it is considered "expired" and thus "rejected". Once the loan is guaranteed and approved, the MFI disburses the loan through its own financial resources, and the borrower repays the loan as per their loan contract terms. In case of default, the loan guarantee from Kiva guarantors is used to partially cover the loan loss. Otherwise, the guarantee funds are returned to the guarantors once the loan is repaid.

The National Digital Identity Platform (NDIP) is Africa's first decentralized iden-

[5] Details of Kiva lending process are available at https://www.kiva.org/about/how#how-loans-works.

tity platform, developed in partnership with Kiva, the Government of Sierra Leone, the United Nations Capital Development Fund, and the United Nations Development Program.[6] NDIP provides citizens of Sierra Leone with the ability to own and use their national civil identity digitally. Designed to operate on a national scale, the platform was first implemented in September 2018 by integrating service providers, both formal (such as the Bank of Sierra Leone and the National Civil Registration Authority), and informal (such as community lenders and MFIs). The NDIP adopts the Kiva blockchain protocol as its core blockchain-enabled digital identity system, which was initially launched to advance financial inclusion goals and empower entire financial ecosystems – individuals, government agencies, informal market actors, and formal financial institutions. The Kiva blockchain protocol is built on Hyperledger blockchain project, functioning as a decentralized and tamper-proof ledger for identity data storage and transaction recording.[7] Using fingerprints and other biometric data, the Kiva blockchain protocol provides citizens of Sierra Leone with digital identifications and access to their complete credit histories. Once registered, individuals are assigned a unique digital wallet that records their personal information and financial transaction data.[8]

Under the Kiva blockchain protocol, the loan application process for borrowers at a local MFI is fully digitized. The MFI scans the digital identification of the borrowers and securely access their information. Using Electronic Know Your Customer

[6] Launched in 2018, NDIP is deployed on a national scale, and implemented in conjunction with modernized regulatory frameworks. NDIP using blockchain protocol to enable Sierra Leone to advance financial inclusion goals and empower entire financial ecosystems. (https://www.coindesk.com/markets/2018/09/28/sierra-leone-to-develop-blockchain-based-id-platform-with-un-partnership/)

[7] The methods to store, transmit and use identity and transaction data in Kiva blockchain protocol is hosted on the Hyperledger blockchain project. (https://www.hyperledger.org/learn/publications/kiva-case-study)

[8] Kiva blockchain Protocol is the provisioning of a user-controlled digital wallet, into which organizations can record information in a way that is secure and verifiable. (https://kivaprotocol.com/solutions/)

technology and the blockchain consensus algorithm, the MFI verifies their information and financial credit records in a decentralized manner.[9] The MFI then uses the borrower information and credit records retrieved through the digital identity system to create the "borrower profile". The loan request is submitted to Kiva for the guarantor funding process. If the loan request reaches the requested guarantee amount, the loan is considered "approved." Once the MFI disburses the loan from its own financial resources, the detailed information about the loan is recorded in the individual's digital wallet and enforced via smart contracts, making it an immutable part of the borrower's credit history.[10]

4.2.2 Hypothesis Development

The quality of borrower credit records is crucial for accurately evaluating loan requests, as funders such as lenders and guarantors rely on these records to determine borrowers' creditworthiness and repayment ability. During the loan process, borrowers are required to provide proof of their identity and financial status. However, in developing countries where many individuals lack formal, verifiable identity and credit records, these self-reported documents are susceptible to fraud and tampering, posing additional obstacles to verification (Begley and Purnanandam, 2021).[11] Moreover, factors such as expensive centralized due diligence further compound these

[9] The technical details can be found in the white paper of Kiva blockchain protocol. (https://assets.ctfassets.net/3r0y27gkro2c/5Lw3ByLipo9NNURU6m8pVn/5fd843f9913e8e27463b8fcf03bc8ed7/Kiva_Protocol_-_Technical_White_Paper_-_June_2021.pdf)

[10] The blockchain protocol provides Sierra Leone residents digital wallets using fingerprints and other biometric data, loan process is then enforced via smart contracts. (https://www.ledgerinsights.com/kiva-sierra-leone-blockchain-id-system/)

[11] In practice, entrepreneurs have several methods at their disposal to engage in financial record fraud and tampering. For example, Wuhan Kingold Jewelry, one of China's largest jewelry makers, loaned $4 billion with 83 tonnes of fake gold as collateral to cover any credit loss. (https://www.afr.com/markets/commodities/the-mystery-of-4b-of-loans-backed-by-fake-gold-20200705-p55938). Entrepreneurs can also engage in identity theft and asset falsification for maintaining business and approving loan. (https://www.investopedia.com/articles/mortgages-real-estate/10/how-mortgage-fraud-affects-markets.asp).

difficulties, making the document verification process challenging, if not impossible.[12] The prevalence of fraudulent information and verification challenges significantly reduces the legitimacy and authenticity of borrower records. Consequently, the lower quality of borrower credit records can lead to inaccurate and negative assessments of borrowers' creditworthiness and repayment ability (Griffin and Maturana, 2016; Mian and Sufi, 2017), thereby reducing guarantors' willingness to contribute to a borrower's loan request.

For blockchain technology, one defining feature is its ability to create a decentralized and tamper-proof ledger that provides a mean of verifying the authenticity and integrity of information. Blockchain technology does not rely on a single centralized authority to maintain records. Instead, it is a decentralized network of entities working together to maintain the ledger (Iyengar et al., 2023; Cong et al., 2023). Through digital signatures and consensus algorithms, blockchain protocols ensure that once data is recorded on the blockchain, it cannot be altered without decentralized permission of the network. In comparison to traditional centralized systems, blockchain is a more effective and cost-saving way to maintain uniform view of the state of things and the order of events (Chod et al., 2020). In the context of information authenticity in financial inclusion, the benefit of everyone sharing and trusting the same ledger is clear. Uploading any new information, such as an individual's biographical information and financial records, requires decentralized verification from multiple organization and entities. This approach ensures reliable checks and balances, enhancing confidence in data authenticity.[13] Once data is uploaded and recorded on

[12] Banks are limited in their ability to individually distinguish the quality of borrower's financial credit records due to the following reasons: 1) Cost of due diligence. It is usually costly and inefficient to conduct due diligence and screen each borrower by involving accountants, attorneys, and other consulting personnel (Sufi, 2007; Ivashina, 2009). 2) A set of factors, such as appraiser ability, borrower information quality and external environment, could also induce inaccurate collateral evaluation and inability to prevent financial credit fraud (Gustafson et al., 2021; Hu et al., 2023).

[13] On-chain transactions and credit records can operate in a more reliable manner due to the unique distributed ledger system provided by blockchain technology (Whitaker and Kräussl, 2020).

the system, it cannot be tampered with or altered. Compared to the traditional process of borrower credit screening, blockchain-enabled digital credit underwriting ensures that the borrower information retrieved and uploaded to the Kiva platform by the MFI is legitimate and authentic. This higher degree of credibility is likely to increase the trust of potential funders, such as guarantors on the Kiva platform, in a borrower's profile – leading them to be more confident about the stated borrower information, loan purpose, and credit history (Galak et al., 2011; Luo et al., 2022). The increased confidence is likely to boost guarantors' willingness to contribute to a borrower's loan request along two dimensions: (1) the number of guarantors contributing to the loan request, and (2) the contributing amount per guarantor. Therefore, we posit:

Hypothesis 1. *Kiva borrowers attract more guarantors and larger per-guarantor contributions under the blockchain protocol.*

Subsequently, if Kiva borrowers attract more guarantors and larger per-guarantor contribution under the blockchain protocol, their loan requests stand a higher chance of reaching the requested funding targets. Since a Kiva loan is deemed "approved" the moment it reaches the funding target, increased guarantor contribution directly translates into a higher probability of the borrower being funded. Formally, we posit:

Hypothesis 2. *Kiva borrowers are more likely to be funded under the blockchain protocol.*

Turning to the perspective of MFIs, we anticipate that the blockchain protocol will reduce lending MFIs' portfolio risk through two primary channels. First, from a technological perspective, since the blockchain protocol is implemented through smart contracts, these contracts automatically enforce predetermined loan terms and conditions, ensuring borrowers adhere to their commitments (Cong and He, 2019; John et al., 2022). By automatically triggering penalties in cases of non-compliance, smart contracts overcome credible commitment issues and reduce the risk of borrower default once the loan has been granted. Moreover, the transparent and immutable nature of smart contracts ensure that any defaulting action is permanently recorded on the blockchain. This information is readily accessible to other lenders, making it harder for defaulting borrowers to secure future loans or engage in transactions. As a result, the repercussions of defaulting are amplified for borrowers, which should enhance borrower accountability and deters defaulting.

Second, from the Kiva platform perspective, if borrowers under the blockchain protocol attract more guarantors and larger per-guarantor contributions, they can more easily reach the guarantee target set by MFIs. As a result, MFIs may have an incentive to raise the required guarantee amount, meaning a larger share of the total loan amount must be funded by guarantors before disbursement. This reduces the uncovered portion of the loan, thereby limiting the potential loss for MFIs in the event of borrower defaulting. Taken together, the automatic enforcement via smart contracts and the reduced uncovered loan exposure both contribute to lowering the overall lending risk for MFIs under the blockchain protocol. Thus, we formalize the anticipated impact of the blockchain protocol on lenders' portfolio risk as follows.

Hypothesis 3. *MFIs experience a lower level of lending portfolio risk under the blockchain protocol.*

Subsequently, if Hypothesis 3 holds, the reduced lending portfolio risk for MFIs may lead to lower loan contract interest rates under the blockchain protocol. First, a lower portfolio risk level indicates a decreased likelihood of default, reducing the potential financial loss associated with default loans for MFIs. Second, with improved borrower credit information quality and lower default probability, MFIs can streamline their due diligence processes, saving on operational expenses such as credit screening and debt collection (Fuster et al., 2019; Berg et al., 2020). As a result, MFIs can afford to reduce the contract interest rates on microloans while still covering their risk exposure. Additionally, the reduced portfolio risk level may also enable MFIs to expand their lending volume by accommodating more borrowers at reduced risk levels (Dell'Ariccia and Marquez, 2006; Cole et al., 2015). We formalize the anticipated impact of the blockchain protocol on contract interest rates as follows.

Hypothesis 4. *Under the blockchain protocol, MFIs offer borrowers loans with lower contract interest rates.*

4.3 Data and Empirical Strategy

4.3.1 Data Collection and Sample Construction

We obtain individual Kiva borrower data from the Kiva website between 2012 to 2021. On the Kiva website, borrower profiles are submitted by Kiva MFIs and include entrepreneur stories, lending details related to individual characteristics, and lending intentions. We collect our borrower dataset by checking each borrower's lending profile. Our dataset contains all detailed borrower characteristics, including age, gender, marital status, number of children, guarantee amount, loan length, loan approval status, year of business operated (year of business), number of income sources (income sources), and whether the individual is a repeat borrower (repeat borrower). We also gather guarantor contribution data from each borrower's profile, including the number of guarantors and the per-guarantor contribution amount (average amount contributed by Kiva guarantors to the loan). To ensure data quality, we restrict our sample by excluding borrowers with missing characteristics data and those lacking financial credit data (i.e., missing year of business operated or income sources). Our final borrower sample consists of 186,325 unique individual-level loan observations from 47 developing countries.

Next, we extract the lending portfolio and financial performance data for the associated Kiva MFIs from the Kiva Lender Database between 2012 to 2021. For each MFI, we obtain a comprehensive set of portfolio loan status on the Kiva lending program, including the annual amount of loans ended (amount of loans no longer in the process of being paid back), loans extended (amount of loans raised and disbursed by MFIs), and loans defaulted (loans with repayments below the expected amount and having no repayments reported to Kiva MFIs in the past 360 days), etc. In addition, we collect data on each MFI's characteristics, such as the percentage of total loans for women borrowers, average time taken to fund a loan, and duration of participation in the Kiva program. To complement our Kiva MFI dataset, we also gather data from the Microfinance Information Exchange (MIX) market database and from MFI's annual financial statements to obtain the service revenue and operational expenses for each MFI's overall lending service. We exclude MFIs with less than 60 months of tenure on Kiva as of the end of 2021 to match our sample period. Overall, our MFI sample consists of 1,034 MFI-level observations.

We also include country-level control variables for our MFI lending dataset. We identify each MFI's affiliated country using the Kiva country tags and the main origins of its borrower portfolio. We collect country-specific statistics such as annual income per person, population below the poverty line, literacy rate, and life expectancy for each country from the Kiva Lender Database. To supplement our dataset, we also incorporate data on the Financial Development Index for each country from the International Monetary Fund website, which measures the depth, access, and efficiency of financial institutions and financial markets of a given country during our sample period (Svirydzenka, 2016).[14] Table 4.1 reports the summary statistics for our sample, and detailed variable definitions are provided in Table C.1 in the Appendix.

[14] Financial Development Index (FD) rate countries on the depth, access and efficiency of their financial institutions and financial markets (https://data.imf.org/?sk=F8032E80-B36C-43B1-AC26-493C5B1CD33B&sId=1480712464593)

Table 4.1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max	Observations
Panel A: Borrower Sample					
Loan indicators					
Loan approval	0.859	0.396	0	1	186,325
Loan length	14.331	4.896	4	58	186,325
Guarantee amount	616.09	1904.72	25	24975	186,325
Number of guarantors	20.549	34.402	0	734	186,325
Individual characteristics					
Year of business operated	5.708	5.652	0	36	186,325
Number of income sources	1.463	1.083	0	5	186,325
Repeat borrower	0.118	0.452	0	1	186,325
Gender	0.206	0.478	0	1	186,325
Children	3.093	1.810	0	18	186,325
Marital status	0.869	0.394	0	1	186,325
Age	38.964	11.017	7	80	186,325
Panel B: MFI Sample					
Lending Portfolio					
Interest rate	0.293	0.225	0	0.961	1,034
Loan ended	350.42	262.87	16.93	8067.12	1,034
Loan extended	281.16	152.56	0	4521.72	1,034
Loan defaulted	18.67	36.10	0	233.43	1,034
Service revenue	185.38	87.04	45.01	413.73	1,034
Operational expense	156.81	109.87	19.28	397.30	1,034
MFI characteristics					
Average years of business	6.031	2.275	1.753	18.259	1,034
Average income sources	1.648	0.370	0.066	3.747	1,034
Proportion of repeat borrower	0.124	0.329	0.054	0.392	1,034
Loans to women	0.682	0.401	0.352	0.965	1,034
Time to fund loans	7.66	4.84	1.73	16.05	1,034
Time on Kiva	122.39	20.35	0	167	1,034
Macroeconomic factors					
FID	0.16	0.13	0.01	0.64	423
FIA	0.17	0.16	0.01	0.42	423
FIE	0.21	0.10	0.03	0.75	423
FMD	0.14	0.11	0.00	0.74	423
FMA	0.10	0.13	0.00	0.40	423
FME	0.12	0.13	0.00	0.47	423
Annual income	3085.91	1981.17	238.21	7441.04	423
Poverty line	0.326	0.221	0.105	0.815	423
Life expectancy	67.104	36.321	52.163	81.377	423
Literacy rate	0.721	0.382	0.481	0.899	423

Note: This Table reports the summary statistics for our sample data. Panel A provides the Kiva borrower statistics for our borrower-level specifications. Panel B are both the MFI characteristics and country-level variables. Detailed definitions of variables are provided in Table C.1 in the Appendix.

4.3.2 Empirical Strategy

To examine the impact of the blockchain protocol implementation on the lending outcomes of Kiva borrowers, we leverage the lending protocol change in Kiva Sierra Leone in September 2018 as an exogenous shock to conduct a difference-in-differences (DID) analysis. Our treated group consists of Kiva borrowers in Sierra Leone, where the blockchain protocol was implemented, while the control group consists of borrowers from all other developing countries that did not implement the protocol change. In our sample, which comprises of 186,325 unique individual borrower-level loan observations, 29,110 observations belong to the treated group, while the remaining 157,215 observations fall into the control group. Our borrower-quarter level DID specification is similar to those of Wei and Lin (2017) and Wang and Overby (2023), where the DID estimations are based on cross-sectional data with most lending applicants observed only once, either before and after the treatment. In our DID model, the first difference is between countries that adopted the blockchain protocol and those that did not; the second difference is between the periods before and after the blockchain protocol implementation in September 2018. The DID specification is as follows:[15]

$$y_{i,m,c,t} = \beta_0 + \beta_1\, Blockchain_c \times Post_t + IC_{i,c,t}\gamma_1 + MC_{m,c,t}\gamma_2 + CC_{c,t}\gamma_3 + \delta_m + \eta_c + \theta_t + \varepsilon_{i,m,c,t} \tag{4.1}$$

where i, m, c, t denote the individual, MFI, country, and quarter, respectively. For the guarantor specification, $y_{i,m,c,t}$ denotes the number of guarantors contributing to loan applicant i and the per-guarantor contribution amount. For the borrower lending specification, $y_{i,m,c,t}$ includes a set of loan application outcomes such as loan approval status (an indicator that takes the value of one if the loan is funded, and

[15]MFIs from Kiva Sierra Leone adopted the blockchain protocol as the credit evaluation and record system in September 2018. In addition, we notice that the blockchain protocol implementation in Kiva Sierra Leone is the only policy and protocol change on Kiva during our sample period.

zero otherwise), the requested guarantee amount, and loan length for approved loan contracts. $Blockchain_c$ is an indicator variable that equals one if country c adopts the blockchain protocol, and zero otherwise. $Post_t$ takes the value of one for periods after the blockchain protocol implementation (September 2018) and zero otherwise. $IC_{i,c,t}$ denotes a set of individual borrower characteristics, including year of business operated, income sources, repeat borrower dummy, age, gender, marital status, and the number of children. $MC_{m,c,t}$ represents the MFI-level control variables. Specifically, we include loan extended, service revenue, operational expense, loans to women, time to fund loans, and time on Kiva to control for the differences in MFI characteristics. $CC_{c,t}$ indicates country-level control variables, including the annual income per person, population below the poverty line, literacy rate, life expectancy, and financial development indices for each country. δ_m, η_c and θ_t are the MFI, country, and time fixed effects, respectively. $\varepsilon_{i,m,c,t}$ is the error term.

We then examine the impact of the blockchain protocol on the portfolio risk and lending behaviors of Kiva MFIs. Similar to the DID analysis at the borrower-level, we again leverage the introduction of the blockchain protocol in Sierra Leone in September 2018 as an exogenous shock. Our treated group consists of Kiva MFIs in Sierra Leone, while the control group includes Kiva MFIs in all other developing countries that did not implement the blockchain protocol. In our sample, which comprises 1,034 MFI-level observations, of which 158 belong to the treated group, with the remaining 876 observations in the control group. Given that the financial and lending characteristics data for Kiva MFIs we have collected are reported annually, we conduct our MFI panel specification at the MFI-year level. The DID specification of our MFI panel is as follows:

$$y_{m,c,t+1} = \beta_0 + \beta_1\, Blockchain_c \times Post_t + PC_{m,c,t}\gamma_1 + MC_{m,c,t}\gamma_2 + CC_{c,t}\gamma_3 + \delta_m + \eta_c + \theta_t + \varepsilon_{m,c,t} \quad (4.2)$$

where m, c, t denote the MFI, country, and year, respectively. $y_{m,c,t+1}$ represents the portfolio risk and lending behaviors for Kiva MFI m in country c during year $t+1$. Following the definition of portfolio risk provided by Kiva, we employ the annual amount of loans defaulted and the default rate (amount of loans defaulted scaled by the amount of extended loans) as measures of MFIs' lending portfolio risks. We introduce three measures of MFI's lending behaviors: (1) the number of loan applications, (2) the number of outstanding loans (loans currently in the repayment process), and (3) the total amount of loans extended in a given year. To examine whether the blockchain protocol can effectively lower contract interest rates on microloans, we further include the average contract interest rate for each Kiva MFI in specification (4.2).

To control for the portfolio creditworthiness of each MFI, we aggregate the affiliated borrowers' financial credit records to the MFI-level and calculate the average portfolio creditworthiness, $PC_{m,c,t}$, for each Kiva MFI. This measure includes the average year of business operated, average income sources, and the proportion of repeat borrowers for Kiva MFI. $Blockchain_c$ is an indicator variable that equals one if MFI m in country c adopts the blockchain protocol, and zero otherwise. $Post_t$ takes the value of one for periods after the blockchain protocol implementation, and zero otherwise. $MC_{m,c,t}$ represents the MFI characteristics, including the loan extended, service revenue, operational expense, loans to women, time to fund loans, and time on Kiva. $CC_{c,t}$ is the set of country-level control factors, including the Financial Development Index, average annual income, population below the poverty line, literacy rate, and life expectancy. δ_m, η_c, and θ_t are the MFI, country, and time fixed effects, respectively. $\varepsilon_{m,c,t}$ is the error term.

Since the blockchain protocol is implemented at the country-level among Kiva Sierra Leone borrowers, it may be more appropriate to cluster the standard errors at the country-level to absorb the within-country correlations (Abadie et al., 2023). Doing so results in 47 clusters, with only one treated cluster. It has been shown in the literature that asymptotic cluster-robust standard errors can be downward-biased when there are few (treated) clusters (Cameron et al., 2008; Conley and Taber, 2011). To address this concern, we follow MacKinnon and Webb (2018) and implement a two-step subcluster wild-bootstrap procedure that deals with a small number of treated clusters: we first generate bootstrap samples using the subcluster structure (at the country-quarter level for the borrower-level analysis; country-year level for the MFI-level analysis), and then compute the test statistics for each bootstrap sample using standard errors clustered at the original treatment (country) level.[16]

4.3.3 Parallel Trends Assumption

Our DID analysis relies on the parallel trends assumption: in the absence of the blockchain protocol implementation, both the treated and control group would have experienced similar changes in outcomes. To verify the parallel trends assumption, we conduct event-study estimations using a series of pre- and post-treatment dummy variables to compare the lending characteristics before and after the blockchain protocol implementation (Wang and Overby, 2023; Jin et al., 2023). In particular, we estimate the following regression model:

[16] MacKinnon and Webb (2017) further show that wild cluster bootstrap also can fail in difference-in-differences settings when the number of treated clusters is very small, especially when the number of observations per cluster differs substantially across clusters.

$$y_{i,m,c,t} = \beta_0 + \sum_{j=-4}^{3} \beta_j \, Blockchain_c \times Time_{t+j} + IC_{i,c,t}\gamma_1 + MC_{m,c,t}\gamma_2 + CC_{c,t}\gamma_3 + \delta_m + \eta_c + \theta_t + \varepsilon_{i,m,c,t} \quad (4.3)$$

Specification (4.3) mirrors our main regression in (4.1) except that we replace $Blockchain_c \times Post_t$ with $\sum_{j=-4}^{3} \beta_j \, Blockchain_c \times Time_{t+j}$. $Time_{t+j}$ is a dummy variable equal to one if quarter t is jth quarter after the blockchain protocol implementation in Q4 2018 (or for $j < 0$, $-j$th quarter prior to the event). $y_{i,m,c,t}$ indicates the measures of guarantor contributions and lending outcomes for Kiva borrowers in our borrower-level specifications. Similarly, we also employ the lending portfolio risks and lending behaviors of MFIs as dependent variables in the MFI-level event-study estimations. We report the results of borrower- and MFI-level parallel trend tests in Section 4.4 and Section 4.5, respectively.

4.4 Borrower-Level Results

4.4.1 Guarantor Behaviors

We first examine the impact of the blockchain protocol on guarantor behaviors on the Kiva platform. Specifically, we use the number of guarantors and the per-guarantor contribution amount from each borrower's lending profile as the dependent variables. Table 4.2 reports the estimation results of the guarantor behavior specifications. In Columns (1) and (3), we estimate our DID specifications without fixed effects. The coefficients of the interaction term, $Blockchain \times Post$, are all positive and statistically significant at the 1% level, indicating that the blockchain protocol implementation positively impacts the behaviors of Kiva guarantors, both in terms of the number of guarantors and the per-guarantor contribution amount. To improve the estimation precision of the treatment effect, we include all fixed effects in Columns (2) and (4). The coefficients of the interaction term continue to be positive and significant. For example, the coefficient of *Blockchain* × *Post* in Column (2) indicates that the blockchain protocol implementation increases the number of contributing guarantors by 2.283 (t-stats: 3.41), which translates into a 11.10% improvement relative to the average number of guarantors listed in Table 4.1.

Table 4.2: Blockchain Protocol and Guarantor Behavior

	Number of guarantors		Per-guarantor amount	
	(1)	(2)	(3)	(4)
Blockchain × *Post*	3.049***	2.283***	2.412***	1.795***
	(4.64)	(3.41)	(3.57)	(3.15)
Blockchain	−0.824**		−0.478***	
	(−2.51)		(−2.92)	
Post	0.042		0.026	
	(1.19)		(0.85)	
Year of business	0.554***	0.425***	0.460***	0.345***
	(3.40)	(2.96)	(3.24)	(2.80)
Income sources	0.844***	0.506***	0.893**	0.582**
	(3.23)	(2.79)	(2.55)	(2.12)
Repeat borrower	1.397***	0.941***	1.172***	0.728*
	(3.74)	(3.10)	(2.70)	(1.91)
Gender	−0.863***	−0.559***	−0.620***	−0.436***
	(−3.80)	(−3.26)	(−3.57)	(−3.06)
Children	0.302***	0.146**	0.217***	0.165**
	(2.63)	(2.28)	(2.86)	(1.98)
Marital status	0.044***	0.027**	0.041*	0.020
	(3.10)	(2.15)	(1.71)	(0.84)
Age	−0.292***	−0.168***	−0.133**	−0.085**
	(−3.98)	(−3.47)	(−2.48)	(−2.21)
Controls	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	Yes
Observations	186,325	186,325	186,325	186,325
Adjusted R^2	0.185	0.272	0.163	0.240

Note: This Table provides the estimation results of the guarantor specifications during the sample period of 2012 to 2021. The dependent variables are the number of participating guarantors and per-guarantor amount for Kiva borrowers. *Blockchain* is an indicator variable that takes the value of one for borrowers in Sierra Leone. *Post* takes the value of one for periods after the blockchain protocol implementation (September 2018). t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

To further examine the relationship between the blockchain protocol adoption and the guarantor behaviors on Kiva, we aggregate our guarantor data to the MFI- and country-level and run the respective DID specifications. The coefficients of $Blockchain \times Post$ remain positive and significant, further confirming that Kiva contributing guarantors demonstrate a preference for blockchain-enabled lending profiles. This preference facilitates the lending process for borrowers applying for microloans through the blockchain protocol. Additionally, we find that the implementation of the blockchain protocol does not appear to reduce Kiva guarantor contributions in other developing countries.

4.4.2 Borrower Lending Outcomes

Having established that Kiva borrowers attract more guarantors and larger per-guarantor contributions under the blockchain protocol, we next examine the impact of the blockchain protocol implementation on Kiva borrowers' lending outcomes, including loan approval status, requested guarantee amount, and loan length. Table 4.3 reports the difference-in-differences estimation results for our borrower-level lending specifications. In Columns (1) and (2), we regress loan approval status on $Blockchain \times Post$. The results show that borrowers applying for microloans through the blockchain protocol experience an 8.7 percentage point (t-stats: 3.16) increase in average loan approval probability. In Columns (3) to (6), we also find that the blockchain protocol implementation increases the requested guarantee amount by 58.33 USD (t-stats: 3.01) and extends the loan length by 0.85 months (t-stats: 2.94) for approved loan contracts. These represent improvements of 9.47% and 5.93% in the respective lending outcomes, relative to their mean values listed in Table 4.1. Overall, the empirical results in Section 4.4 support our hypotheses that Kiva borrowers attract more guarantors and larger per-guarantor contributions under the blockchain protocol, thereby increasing their likelihood of being funded on the Kiva platform.

Having obtained our regression results, we now test the parallel trends assumption

Table 4.3: Blockchain Protocol and Lending Indicators of Individual Borrowers

	Loan approval		Guarantee amount		Loan length	
	(1)	(2)	(3)	(4)	(5)	(6)
Blockchain × *Post*	0.101***	0.087***	66.826***	58.325***	1.324***	0.845***
	(4.35)	(3.16)	(3.84)	(3.01)	(3.41)	(2.94)
Blockchain	−0.038***		−31.028***		−0.028***	
	(−3.47)		(−3.92)		(−2.96)	
Post	0.042		1.630		0.017	
	(1.35)		(0.77)		(0.58)	
Year of business	0.045***	0.028**	21.496***	14.712***	0.304**	0.241**
	(3.72)	(2.53)	(3.52)	(2.70)	(2.51)	(2.08)
Income sources	0.066***	0.032**	22.108***	16.933**	0.580***	0.287***
	(3.05)	(2.48)	(2.78)	(2.04)	(3.48)	(2.89)
Repeat borrower	0.074***	0.040***	42.653***	34.284**	1.308***	0.906***
	(4.26)	(3.20)	(3.19)	(2.23)	(3.66)	(3.21)
Gender	0.061***	0.043***	−40.776***	−31.795***	−1.282***	−0.978**
	(3.88)	(3.07)	(−3.84)	(−2.76)	(−2.82)	(−2.04)
Children	−0.025**	−0.017	22.151***	18.122**	0.564**	0.379*
	(−2.30)	(−1.57)	(3.41)	(2.54)	(2.16)	(1.68)
Marital status	0.009	0.004	12.250**	9.308*	0.012	0.004
	(1.43)	(0.75)	(2.52)	(1.77)	(1.14)	(0.55)
Age	−0.026***	−0.014**	−7.943***	−5.214***	−0.147***	−0.056**
	(−2.92)	(−2.33)	(−3.68)	(−2.74)	(−3.09)	(−2.16)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	Yes	No	Yes
Observations	186,325	186,325	160,053	160,053	160,053	160,053
Adjusted R^2	0.194	0.268	0.135	0.274	0.112	0.206

Note: This Table provides the estimation results of borrower lending specifications during the sample period of 2012 to 2021. The dependent variables are the loan approval status for Kiva borrowers, requested guarantee amount, and loan length for approved loan contracts. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

for our borrower-level specifications. We run specification (4.3) by including measures of both the guarantor behaviors and borrower lending outcomes. The pre-treatment dummies are all statistically insignificant, while the coefficients for post-treatment dummies are significant and appear to be persistent. We further plot the time-varying dynamics of the treatment effects in Figure 4·1. For both the guarantor and lending metrics, we observe non-significant estimates prior to the event and significant effects emerging only after the blockchain protocol implementation. These results suggest that the parallel trends assumption is satisfied for our borrower-level DID regressions.

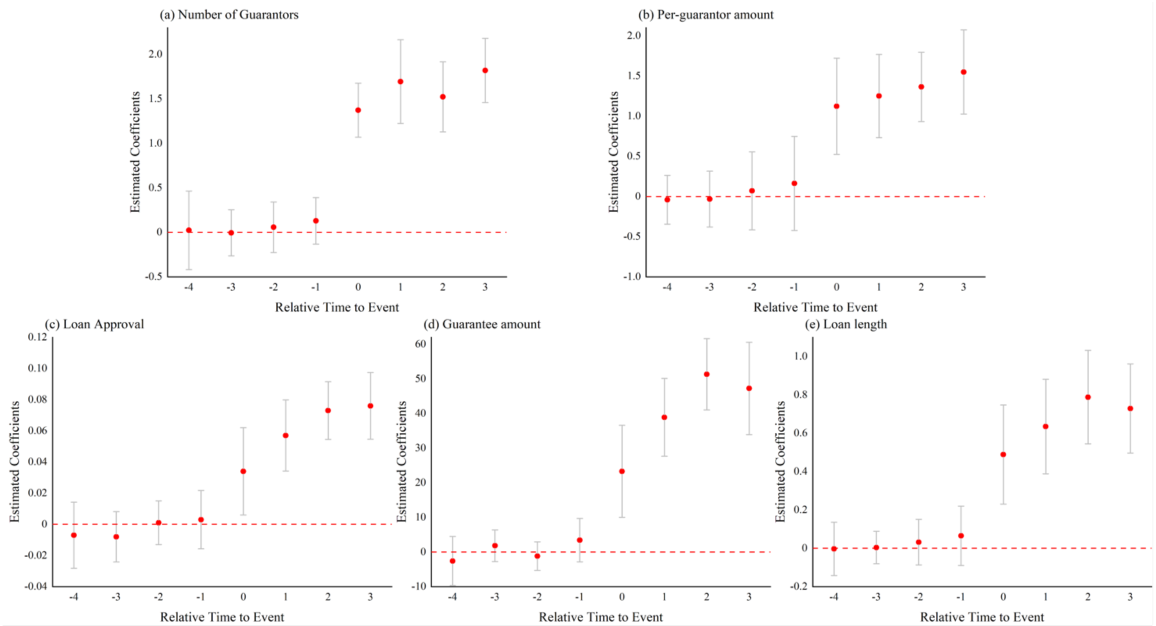


Figure 4·1: Parallel Trends and Dynamic Treatment Effects: Borrower-level Specifications. These figures plot the time-varying treatment effects derived from the borrower-level DID regressions. We include the number of guarantors (a), per-guarantor amount (b), loan approval (c), guarantee amount (d), and loan length (e) as dependent variables. The dashed lines represent the 95% confidence intervals. The results show non-significant estimates prior to the event with significant effects emerging after the blockchain protocol implementation.

4.5 MFI-Level Results

4.5.1 The Portfolio Risks of MFIs

In this section, we examine the effect of the blockchain protocol implementation on the lending portfolio risks of MFIs. We aggregate Kiva borrowers' financial credit records to their respective MFIs and construct the portfolio creditworthiness for each MFI. We then run the MFI-level regression specified in (4.2) and report the results in Table 4.4. The dependent variables are: (1) the amount of loans defaulted, (2) the default rate in MFIs' lending portfolios, and (3) the average contract interest rate offered by each MFI. In Columns (7) and (8), we use the guarantee ratio – calculated as the proportion of the loan amount funded by Kiva guarantors – as the dependent variable. Columns (1), (3), (5), and (7) present the results without fixed effects, while Columns (2), (4), (6), and (8) present the results after controlling for all fixed effects.

The coefficients of $Blockchain \times Post$ in Columns (2) and (4) suggest that implementing the blockchain protocol reduces the average amount of loans defaulted and the default rate in MFIs' lending portfolio by 3,376 USD (t-stats: -3.21) and 1.7 percentage points (t-stats: -2.72), respectively. These findings suggest that the blockchain protocol can effectively lower the overall lending risk for MFIs. In Column (6), we find that implementation of the blockchain protocol leads to a 7.6 percentage point (t-stats: -3.14), or 26.28%, reduction in the average interest rate offered by MFIs. Given that this effect is relatively large in comparison to the literature (Campbell and Cocco, 2015; O'Malley, 2021), we further discuss the underlying economic channels driving this result. First, as shown in Columns (1) and (2), the blockchain protocol implementation lowers the default rate of borrowers, directly reducing MFIs' lending risk. Second, as shown in Columns (7) and (8), the blockchain protocol implementation increases the guarantee ratio by 8.2%, indicating that a larger proportion of loan amounts is covered by guarantors, thus reducing MFIs' uncovered loan exposure.[17] Third, as detailed in Section 4.5.3, the blockchain protocol reduces operational expense per loan during the lending process for MFIs. Overall, these factors substantially lower the risk exposure and operational costs for MFIs, enabling them to sustainably offer loans at lower interest rates in developing economies.

4.5.2 The Lending Behaviors of MFIs

We then examine the impacts of the blockchain protocol on lending behaviors of Kiva MFIs. In addition to calculating the number of annual outstanding loans and the

[17] The average guarantee ratio of MFIs in our data is 51.83%. The effect of the blockchain protocol on the guarantee ratio also be translated into 13.89% increase in the guarantee coverage for MFIs' Portfolio.

Table 4.4: Blockchain Protocol and Portfolio Risks of Microfinance Institutions

	Loan defaulted		Default rate		Interest rate		Guarantee ratio	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Blockchain × *Post*	−4.834***	−3.376***	−0.025***	−0.017***	−0.085***	−0.076***	0.104***	0.082***
	(−4.15)	(−3.21)	(−3.59)	(−2.72)	(−3.239)	(−3.14)	(3.31)	(2.63)
Blockchain	2.561***		0.005**		0.062***		0.048***	
	(2.80)		(2.06)		(3.02)		(2.71)	
Post	−0.020		−0.000		−0.003		0.004	
	(−0.72)		(−0.85)		(−0.86)		(1.18)	
Average years of business	−0.744**	−0.340*	−0.009*	−0.003	−0.040***	−0.025**	0.021	0.004
	(−2.57)	(−1.80)	(−1.75)	(−0.54)	(−3.05)	(−2.56)	(1.54)	(0.46)
Average income sources	−1.524***	−0.862**	−0.012**	−0.006*	−0.054***	−0.036***	0.053**	0.014
	(−3.19)	(−2.26)	(−2.44)	(−1.91)	(−3.85)	(−3.10)	(2.11)	(1.06)
Proportion of repeat borrower	−0.649**	−0.352*	−0.007***	−0.002**	−0.026***	−0.016***	0.021*	0.005
	(−2.42)	(−1.74)	(−2.83)	(−2.10)	(−3.06)	(−2.77)	(1.80)	(0.55)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	908	908	908	908	908	908	908	908
Adjusted R^2	0.142	0.315	0.121	0.240	0.093	0.186	0.163	0.245

Note: This Table provides the estimation results of the lending portfolio risk specifications during the sample period of 2012 to 2021. The dependent variables are the amount of loan defaulted, default rate, contract interest rate, and guarantee ratio in MFIs' lending portfolios. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

amount of total loan extended as proxies for MFIs' lending volume, we also use the number of loan applications as a proxy for the demand-side volume of each MFI. Table 4.5 presents the estimation results. The coefficients of the interaction term $Blockchain \times Post$ in Columns (2) and (4) show that the blockchain protocol implementation contributes to increases of 56.07 and 41.79 in the annual loan application volume and the number of outstanding loans, respectively, for MFIs. These increases can be translated into average growth of 63,312 USD and 47,934 USD in the annual loan application demand and the amount of outstanding loans, respectively. Additionally, Column (6) indicates that the blockchain protocol leads to an increase of 39,279 USD, or 14.09%, in the amount of total loan extended by Kiva MFIs. In Columns (7) and (8), we assess the impact of the blockchain protocol on the average loan approval rate of lenders by aggregating individual loan approval statuses to the MFI-level. The coefficient of $Blockchain \times Post$ in Column (8) suggests that implementing the blockchain protocol increases the average loan approval rate of MFIs by 6.7% (t-stats: 3.63).

The MFI-level estimation results support our hypotheses, indicating that MFIs experience lower portfolio risks and extend microloans at lower interest rates following the blockchain protocol implementation. Similar to our borrower-level regressions, we perform parallel trends tests at the MFI-level by including all the lending metrics of MFIs in specification (4.3). We illustrate the time-varying dynamics of the treatment effects on MFI lending metrics in Figure 4·2. The results show non-significant estimates prior to the event, with significant effects emerging after the blockchain protocol adoption. Thus, the parallel trends assumption is also satisfied for our MFI-level DID regressions.

Table 4.5: Blockchain Protocol and Lending Behaviors of Microfinance Institutions

	Loan application		Loan outstanding		Loan extended		Approval rate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Blockchain × *Post*	80.147***	56.070***	72.283***	41.794***	46.735***	39.279***	0.091***	0.067***
	(3.59)	(2.76)	(4.14)	(3.25)	(3.60)	(3.26)	(4.28)	(3.63)
Blockchain	60.216***		41.492***		30.429***		−0.020**	
	(3.07)		(2.91)		(3.14)		(−2.10)	
Post	3.776		2.063		1.027		0.005	
	(1.16)		(0.73)		(0.81)		(0.93)	
Average years of business	32.362**	17.900**	27.280**	16.446**	26.082***	14.341*	0.017***	0.009***
	(2.29)	(1.76)	(2.41)	(2.16)	(3.12)	(1.84)	(3.25)	(2.97)
Average income sources	40.928***	23.146**	32.454***	18.201***	27.813***	18.424**	0.036***	0.025**
	(3.35)	(2.31)	(3.23)	(2.67)	(2.85)	(2.24)	(2.69)	(2.03)
Proportion of repeat borrower	28.503***	11.758***	22.570***	11.676**	20.074**	10.462	0.016***	0.012**
	(3.51)	(3.14)	(3.10)	(2.47)	(2.13)	(1.13)	(3.18)	(2.55)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	908	908	908	908	908	908	908	908
Adjusted R^2	0.216	0.346	0.183	0.296	0.182	0.284	0.240	0.316

Note: This Table provides the estimation results of the lending behavior specifications during the sample period of 2012 to 2021. The dependent variables are the number of loan applications, number of outstanding loans, amount of loan extended, and loan approval rate for Kiva MFIs. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.5.3 Operational Sustainability of MFIs

In developing countries, the lack of verifiable financial records often results in higher operational costs, lower business revenues, and greater investment risks for MFIs when making and monitoring loans – all factors that undermine the sustainability of financial services (Brown et al., 2012; Philippon, 2015). In this section, we further explore the impact of the blockchain protocol on the long-term operational sustainability of Kiva MFIs. Specifically, we use the annual lending service revenue and operational expense from Kiva MFIs' financial statements as proxies for their operational sustainability. We also include the service revenue per loan and operational expense per loan as dependent variables, calculated as the service revenue / operational expense scaled by the amount of loan outstanding for each MFI. We run the difference-in-differences regression similar to that in specification (4.2) at the MFI-year level and report the estimation results in Table 4.6.

In Column (1) of Table 4.6, the coefficient of the interaction term, $Blockchain \times Post$, is positive and statistically significant at the 1% level, indicating a positive impact of the blockchain protocol on the financial revenue of Kiva MFIs. After controlling for fixed effects, in Column (2), we find that the annual lending service revenue increases by 3,389 USD (t-stats: 3.10) if an MFI adopts the blockchain protocol. In Columns (3) and (4), we substitute the dependent variable with the operational expense of lending services. The estimation results show that the annual operational expenses decrease by 1,502 USD (t-stats: -3.35) if an MFI adopts the blockchain protocol.

In Column (6), we find that the blockchain protocol implementation is negatively associated with service revenue per loan for MFIs. Although the service revenue per loan decreases, primarily driven by the large drops in contract interest rates under the blockchain protocol as shown in Section 4.5.2, the overall financial revenue of

Table 4.6: Blockchain Protocol and Operational Sustainability of MFIs

	Service revenue		Operational expense		Service revenue per loan		Operational expense per loan	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Blockchain × *Post*	5.116***	3.389***	−2.170***	−1.502***	−31.535***	−23.314**	−62.923***	−45.176***
	(3.83)	(3.10)	(−4.36)	(−3.35)	(−3.24)	(−2.50)	(−5.22)	(−3.91)
Blockchain	4.543***		3.135***		11.128**		8.461**	
	(3.42)		(3.08)		(2.42)		(2.51)	
Post	0.032		−0.018		−0.356		−0.110	
	(1.34)		(−0.76)		(−1.03)		(−0.20)	
Average years of business	0.427***	0.278**	−0.104***	−0.061	−4.270**	−3.360	−0.787*	−0.236
	(3.16)	(2.19)	(−2.84)	(−1.25)	(−2.13)	(−1.48)	(−1.90)	(−0.85)
Average income sources	0.526***	0.381***	−0.123***	−0.060**	−6.286***	−4.873*	−1.476**	−0.590
	(4.28)	(3.07)	(−3.08)	(−1.98)	(−2.86)	(−1.70)	(−2.19)	(−1.40)
Proportion of repeat borrower	0.273***	0.165**	−0.079**	−0.033*	−2.835**	−2.223	−0.535	−0.189
	(3.45)	(2.32)	(−2.35)	(−1.70)	(−2.49)	(−0.82)	(−1.48)	(−0.46)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	908	908	908	908	908	908	908	908
Adjusted R^2	0.097	0.142	0.136	0.185	0.139	0.170	0.102	0.164

Note: This Table estimates the impacts of the blockchain protocol on the operational sustainability of Kiva MFIs during the sample period of 2012 to 2021. The dependent variables are the annual lending service revenue and operational expense for Kiva MFIs and per-loan measures. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Kiva MFIs still increases. This overall increase in service revenue is driven by the higher lending volumes enabled by the blockchain protocol implementation. In our sample, Kiva loans constitute approximately 46.32% of MFIs' overall lending portfolios. This shift significantly influences their broader lending strategy. Consequently, the blockchain protocol moves MFIs from a high-margin, low-volume model toward a low-margin, high-volume strategy, reflecting a strategic focus on serving more borrowers. Moreover, the coefficient of $Blockchain \times Post$ in Column (8) indicates that the blockchain protocol implementation reduces MFIs' operational expenses per loan by 45.18 USD, or 15.93%, enabling MFIs to sustainably adopt this lower-margin lending strategy. In summary, implementation of the blockchain protocol allows Kiva MFIs to achieve both growth in financial revenue and reduction in operational expense, thus enhancing the sustainability of their lending services in developing countries.

4.6 Sensitivity of Blockchain Effects

The results above support the financial inclusion properties of the blockchain protocol: under the blockchain protocol, Kiva borrowers are more likely to receive microloans with lower interest rates, while MFIs experience lower portfolio risks and extend larger lending volume in developing countries. In this section, we extend our analysis to examine the sensitivities of the blockchain effects we identified in our baseline regressions. Specifically, we investigate variations in blockchain effects across Kiva borrowers with different financial credits, geographical areas, and business sectors.

4.6.1 Sensitivity to Financial Credit

We first explore the sensitivity of our results to the financial credit of Kiva borrowers. Puri and Robinson (2007) and Michels (2012) show that a longer business history among loan applicants indicates stronger working capabilities and expected future cash flows, which may increase the likelihood of loan application approval by lenders.

In this vein, we use the year of business operated as a proxy for the strength of the borrowers' financial credit. We sort our sample based on this metric and divide it into 10 deciles. Borrowers with the longest years of business operated are placed in decile 10, indicating the strongest financial credit, while those with the shortest are placed in decile 1, indicating the weakest financial credit.

We run the borrower-level quantile regressions by using the number of guarantors, loan approval status, and requested guarantee amount as the key lending characteristics of Kiva borrowers. We report the estimation results in Panel A of Table 4.7. The coefficients for the interaction term, $Blockchain \times Post$, are positive and significant across all deciles for the number of guarantors, indicating that the blockchain protocol positively impacts guarantor contributions to Kiva loan applicants across various levels of financial credit. However, the magnitudes of these coefficients suggest that the blockchain protocol has a more substantial impact on borrowers with weaker financial credit. For example, implementation of the blockchain protocol results in increases of 3.83, 2.76, and 3.15 guarantors for Kiva borrowers in deciles 1 to 3, respectively, while the increases for borrowers in deciles 8 to 10 are only 1.87, 1.81, and 1.10 guarantors. This indicates that Kiva borrowers with the weakest financial credit can benefit as much as 3.48 (3.83/1.10) times in the number of guarantors compared to those with the strongest financial credit when applying through the blockchain protocol. In Panel A, we find similar patterns when substituting the dependent variable with loan approval status and requested guarantee amount, further indicating that the blockchain protocol exhibits a greater propensity to support loan applicants with less favorable financial credit.

Next, we conduct the MFI-level quantile regressions to further analyze the heterogeneity in the financial inclusion properties of the blockchain protocol across different borrower financial credit levels. In addition to using the loan approval rate and the

Table 4.7: Sensitivity of Blockchain Effects: Financial Credit

Deciles: Year of business	1	2	3	4	5	6	7	8	9	10
Panel A: Loan Indicators of Individual Borrowers										
Number of guarantors	3.829***	2.761***	3.154***	2.298***	2.201***	2.478***	1.624**	1.867*	1.813***	1.102**
	(3.94)	(3.17)	(3.75)	(2.60)	(2.79)	(3.09)	(2.24)	(1.71)	(2.87)	(2.32)
Loan approval	0.109***	0.105***	0.094**	0.090**	0.084*	0.077***	0.071**	0.064**	0.056**	0.044**
	(3.56)	(3.29)	(2.41)	(2.56)	(1.80)	(2.68)	(2.19)	(2.13)	(2.48)	(2.30)
Guarantee amount	76.034***	68.145***	61.160***	64.181***	52.312***	45.155***	40.914**	33.583**	35.264**	25.630**
	(3.69)	(3.10)	(3.53)	(2.84)	(2.69)	(2.75)	(2.03)	(2.52)	(2.14)	(1.98)
Panel B: Lending Statistics of Microfinance Institutions										
Approval Rate	0.089***	0.084***	0.079***	0.068***	0.072***	0.064***	0.056**	0.053***	0.054***	0.047***
	(4.25)	(4.46)	(3.48)	(3.86)	(3.05)	(3.16)	(2.40)	(3.88)	(3.41)	(3.07)
Loan extended	5.975***	5.226***	4.636***	4.382**	3.937**	3.638*	3.841**	2.882**	2.577	2.271*
	(2.79)	(3.28)	(3.07)	(2.16)	(2.48)	(1.82)	(2.06)	(1.97)	(1.09)	(1.76)
Share of portfolio	0.035***	0.018***	0.010**	0.001	−0.002	−0.010*	−0.008**	−0.005*	−0.017**	−0.022***
	(3.05)	(2.60)	(2.03)	(1.45)	(−0.93)	(−1.78)	(−2.20)	(−1.72)	(−2.33)	(−2.97)

Note: This Table provides the variations of blockchain effects on Kiva borrowers with different financial credits. We use year of business operated as the proxy of financial credit and sort Kiva borrowers into deciles based on the distribution of year of business operated, where decile 1 is the lowest decile and decile 10 is the highest decile. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

total loan extended by Kiva MFIs as dependent variables, we calculate, for each borrower decile, the share of lenders' portfolio extended to the respective group (share of portfolio). The results are reported in Panel B of Table 4.7. The coefficients in Panel B show a consistent pattern with those at the borrower-level: the marginal effects of the blockchain protocol are stronger for borrowers with the weakest financial credit (decile 1) compared to those with the strongest financial credit (decile 10). For example, the impacts of the blockchain protocol on the loan approval rates and total loan extended to borrowers in decile 1 are 1.89 (0.089/0.047) and 2.63 (5.975/2.271) times higher than those in decile 10, respectively. Furthermore, we observe that a larger share of loans in MFIs' lending portfolios is now being extended to Kiva borrowers with weaker financial credit (e.g. increase of 3.5% for decile 1). Conversely, there's a reduction in the share of loans for borrowers with stronger financial records (e.g. 2.2% reduction for in decile 10). These results indicate that the blockchain protocol enables MFIs to shift credit extended from borrowers with stronger financial credit to those with weaker financial credit, and thus facilitating financial inclusion.

4.6.2 Variations to Geographic Areas

On the Kiva platform, Kiva MFIs collect and provide geographical information for each borrower. Recognizing that there exists empirical evidence suggesting that borrowers from central cities or developed areas are more likely to have access to financial services (Brown et al., 2016), in this section, we explore geographical variations in the impacts of the blockchain protocol. Sierra Leone consists of one capital (Freetown), four provinces, and 16 districts. We label loan applicants based on their loan submission locations as follows: loan applicants from the capital district of Sierra Leone are labeled as "capital district borrowers"; those from districts with provincial central cities are labeled as "provincial district borrowers", and all others are labeled as "rural area borrowers". Overall, we classify borrowers into three geographical sub-groups: capital, provincial, and rural. The statistics show that borrowers from rural areas constitute the smallest proportion (17.4%) of the total borrower population, suggesting that they were underserved prior to implementation of the blockchain protocol.

Similar to Section 4.5.1, we first re-run the borrower-level regression specification (4.1), using the number of guarantors, loan approval status, and requested guarantee amount as dependent variables. The results are reported in Panel A of Table 4.8. The coefficients in Panel A show that the effects of the blockchain protocol are more pronounced for loan applicants from rural areas. For example, the blockchain protocol increases the loan approval probability by 11.3 percentage points for rural borrowers, compared to 6.8 percentage points for capital district borrowers. Next, we rerun the MFI-level regression with the loan approval rate, total loan extended, and share of portfolio as dependent variables. The results, reported in Panel B of Table 4.8, are consistent with the borrower-level results: the coefficients in Panel B indicate that the blockchain protocol has a more pronounced impact on the lending decisions of Kiva MFIs toward rural borrowers. For example, implementation of the blockchain protocol results in an 8.2 percentage point increase in loan approval rates and a 21,285 USD increase in credit extended to rural borrowers by MFIs, which are 1.61 (0.082/0.051) and 3.45 (21.285/6.173) times greater, respectively, than those for capital district borrowers.

The Kiva platform also provides information on the specific types of businesses each microloan supports. We further examine variations in the impact of the blockchain protocol across individual borrowers in different business sectors. We aggregate the 15 business categories defined by Kiva into four broad sectors: primary, secondary, tertiary, and personal use.[18] The results reveal that the blockchain

[18] To simplify our analysis, we aggregate 15 Kiva business categories into four broad sectors: primary, secondary, tertiary, and personal use. The primary sector includes Kiva borrowers engaged

Table 4.8: Sensitivity of Blockchain Effects: Geographic Areas

Geographic area	Rural	Provincial	Capital
Panel A: Loan Indicators of Individual Borrowers			
Number of guarantors	3.372***	2.200***	1.337***
	(3.08)	(3.86)	(3.12)
Loan approval	0.113***	0.079***	0.068**
	(3.15)	(2.67)	(2.26)
Guarantee amount	71.371***	48.907***	31.483***
	(3.09)	(2.75)	(2.63)
Panel B: Lending Statistics of Microfinance Institutions			
Approval rate	0.082***	0.063***	0.051***
	(3.07)	(3.31)	(2.95)
Loan extended	21.285***	11.626**	6.173**
	(2.85)	(2.19)	(2.02)
Share of portfolio	0.042***	−0.018**	−0.022***
	(3.32)	(−2.31)	(−3.19)

Note: This Table provides variations of the blockchain effects on Kiva borrowers from different geographic areas. Rural, Provincial and Capital are borrowers submitting loans from the rural districts, the provincial districts, and capital districts respectively. t-statistics are reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

protocol exhibits a more pronounced impact on the lending outcomes for borrowers in the primary sector, who constitute the most underserved segment among all sectors prior to the blockchain protocol implementation.

4.7 Conclusion and Economic Implication

New financial programs leveraging blockchain technology hold promise in addressing the financial inclusion challenges arising from issues of information authenticity. In this paper, we utilize a novel dataset comprising both borrower-level loan application and MFI-level portfolio lending characteristic data from Kiva. Exploiting the lending protocol change in Kiva Sierra Leone as an exogenous shock, we employ difference-

in agriculture, textiles, and food industry. The secondary sector comprises borrowers associated with construction and manufacturing. The tertiary sector consists of loan applicants with lending intentions related to clothing, entertainment, health, retail, service, transportation, and wholesale. Borrowers categorized as education, housing, and personal use are classified within the personal use sector.

in-differences (DID) regressions to examine the financial inclusion properties of the blockchain-enabled lending protocol. Our results show that Kiva guarantors demonstrate a preference for loan applicants using the blockchain protocol, and borrowers using the blockchain protocol are more likely to receive microloans with lower interest rates. Furthermore, our MFI-level analysis reveals that MFIs experience lower portfolio risks and extend greater lending volume in developing countries under the blockchain protocol.

Using annual lending service revenue and operational expense data from Kiva MFIs' financial statements as proxies for their operational sustainability, we further find that the blockchain protocol enables Kiva MFIs to achieve significant growth in financial revenue and reductions in operational costs, thus facilitating more sustainable delivery of microloan services. Additionally, we examine the variations in the financial inclusion properties of the blockchain protocol. Our results indicate that the blockchain protocol has a greater tendency to favor loan applicants with less favorable financial credit backgrounds. Moreover, we find that the blockchain protocol has a stronger impact on the lending outcomes for Kiva borrowers from rural areas and those engaged in agriculture, textiles, and the food industry – segments representing the most underserved borrower populations prior to the implementation of the blockchain protocol.

The empirical results we obtain offer economic implications on the adoption of blockchain technology as a financial inclusion program. In developing countries, the common barrier to financial inclusion is the lack of reliable credit identification system for loan applicants. Our study regards blockchain technology as a decentralized credit verification platform. Under the blockchain protocol, uploading new information, such as an individual's biographical details and financial records, requires decentralized verification from multiple entities. This approach ensures reliable checks and balances, enhancing confidence in data authenticity. The results of our borrower-level regression indicate that blockchain-enabled lending services increases the accessibility and affordability of credit services in developing countries. Individuals facing budget constraints can improve their financial and business liquidity by applying for loans through blockchain-enabled lending services.

Existing works document a concerning trend in MFIs known as mission drift, where financial services for the unbanked are increasingly provided by commercially-oriented and deposit-taking MFIs instead of non-profit MFIs (Brown et al., 2016; Zhao and Wry, 2016). Cull et al. (2007) find that larger and more profitable MFIs have larger lending service volumes, higher portfolio loan sizes, and a lower share of female clients. The results of our MFI-level regressions demonstrate that the implementation of blockchain-enabled financial programs yields several positive outcomes for MFIs. Specifically, MFI extending microloans via blockchain experiences lower operational costs and reduced portfolio risk levels. In addition, MFIs also achieve significant growth in financial revenue under the blockchain lending protocol. The blockchain protocol may address the inherent trade-off between profitability and the financial inclusion mission of MFIs.

In this paper, we use the Kiva blockchain protocol implementation in Sierra Leone to examine the financial inclusion properties of blockchain technology. In essence, beyond the Kiva platform, an increasing number of blockchain-based programs have been developed to enable digital identification and ensure complete information authenticity for financial excluded populations. For example, Mojaloop, which employs Interledger to enable completely traceable transactions that can be audited by a regulatory body and financial institutions, is currently being deployed in multiple countries in Africa such as Tanzania, Uganda, and South Africa. Founded in 2016, WeTrust uses the blockchain technology to significantly reduce overhead fees by eliminating the need for a trusted third party (e.g. Banks), and to improve transparency of on-chain transactions. The blockchain program of WeTrust makes banking more affordable and accessible, particularly for those with small transactions.[19] In conclusion, the decentralized nature of blockchain technology has the potential to improve financial inclusion on a global scale by enabling the creation of multi-party verified credit histories and by tokenizing repayment and default records that were previously only available off-chain.

While our study provides robust insights into the impact of blockchain on financial inclusion, there are several limitations that may offer fruitful directions for future research. First, although our analysis suggests significant loan demand changes are unlikely over the short term, it remains inherently challenging to fully disentangle the effects of blockchain adoption on credit assessment improvements from potential shifts in loan demand driven by the introduction of new services. Future research can address this limitation when longer-term data become available, complemented by detailed borrower-level data capturing borrower loan application history. Second, our study faces additional data constraints related to individual borrower-level default rates, timeframe, broader borrower profiles, and additional country-specific variables. Access to more detailed and comprehensive data along these dimensions would further enhance our understanding of the role of blockchain in promoting financial inclusion. Specifically, detailed borrower default and repayment behavior data would enable deeper insights into the influence of blockchain on credit risk and borrower decision-making. Due to the limitation of our data, we leave these areas for future research when suitable data becomes available.

[19] See details of Mojaloop and WeTrust from the report "Blockchain for Social Impact" by Stanford University (https://www.gsb.stanford.edu/sites/gsb/files/publication-pdf/study-blockchain-impact-moving-beyond-hype.pdf?fid).

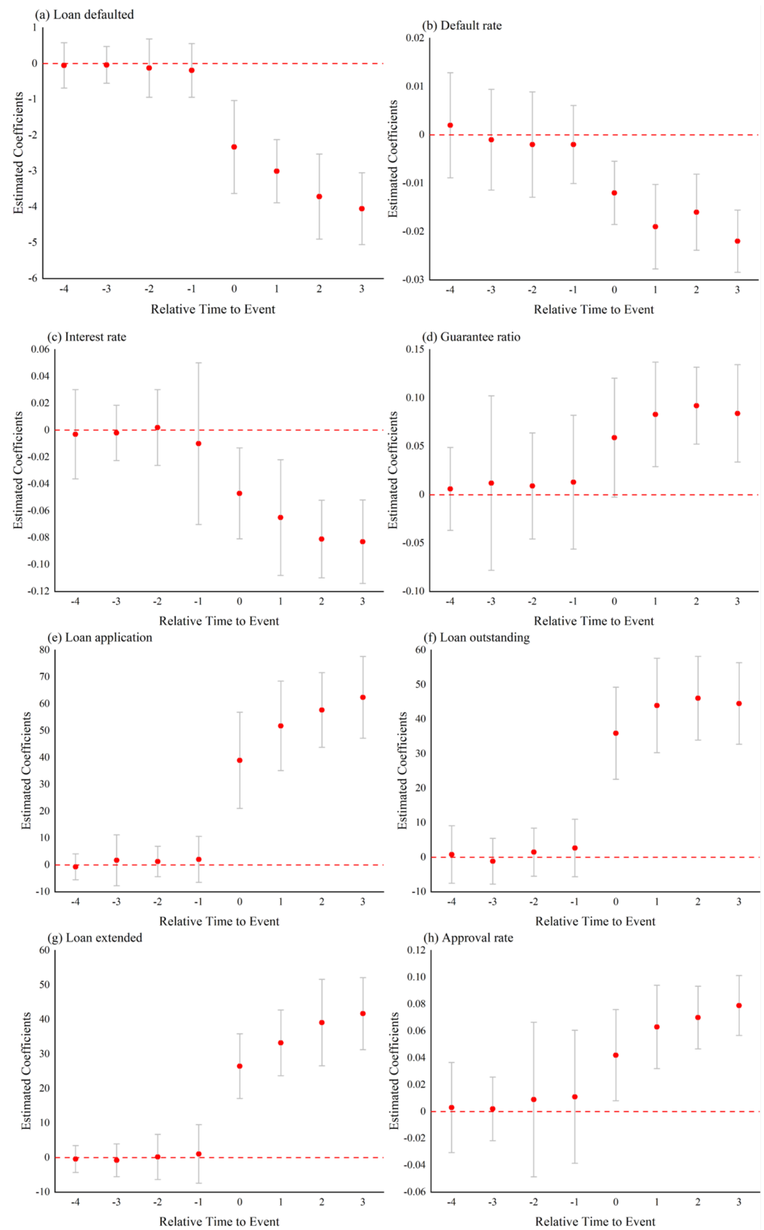


Figure 4·2: Parallel Trends and Dynamic Treatment Effects: MFI-level Specifications. These figures plot the time-varying treatment effects derived from the MFI-level DID regressions. The dependent variables in the top panel are the measures of MFIs' lending portfolio risks, including loan defaulted (a), default rate (b), interest rate (c), and guarantee ratio (d). The dependent variables in the bottom panel are the measures of MFIs' lending behaviors, including loan application (e), loan outstanding (f), loan extended (g), and approval rate (h). The dashed lines represent the 95% confidence intervals.

Appendix A

Appendix for Chapter 2

A.1 Computational Complexity of Q-Learning

Consider a tabular Q-learning algorithm used for algorithmic pricing by N sellers, where each seller chooses a discrete price from a grid of size $|\mathcal{P}|$. Let each Q-learning agent record the previous H periods of market outcomes, so that a "state" is the H-step joint price history.

$$|\mathcal{S}| = \big(|\mathcal{P}|^N\big)^H = |\mathcal{P}|^{NH}.$$

Space complexity. Tabular methods must store one value per state–action pair; with a joint action space of size $|\mathcal{A}| = |\mathcal{P}|^N$, memory scales as

$$\mathcal{O}\big(|\mathcal{S}| \cdot |\mathcal{A}|\big) = \mathcal{O}\big(|\mathcal{P}|^{N(H+1)}\big),$$

which is exponential in N.

Time complexity. Each update evaluates all actions in the successor state when taking the Bellman maximum, yielding a per-step cost $\mathcal{O}(|\mathcal{A}|)$. Across T learning steps the total runtime is therefore

$$\mathcal{O}\big(T\,|\mathcal{P}|^N\big),$$

again exponential in N. Consequently, even moderate markets (e.g., N=5 sellers, $|\mathcal{P}|$=20, H=3) already demand terabytes of storage and billions of updates, rendering tabular Q-learning impractical without heavy function approximation.

A.2 Sanity Checks

A.2.1 Baseline Validation - Demand Function Given

Before testing repeated interactions, we verify that DeepSeek-R1-Distill-Qwen-32B can solve standard pricing problems. We evaluate performance in two one-shot settings. The Bertrand duopoly has the same economic environment as a single period of the repeated game in Section 2.3.2, except that sellers maximize immediate profits without discounting (see Appendix A.3.2). The monopoly game consists of a single seller maximizing joint profit across two homogeneous products (see Appendix A.3.3).

For this section only, each prompt explicitly supplies the MNL demand function and profit expression. We set parameters to $a = 4$, $\mu = 0.1$, $c = 3$, $a_0 = 1$. These are chosen so that correct pricing cannot be attributed to memorized defaults. We conduct 50 independent runs for each setting. As Table A.1 shows, the model adheres to the required output format in all runs and produces prices within 5% of theoretical optima in nearly every trial, confirming that the agents, when prompted with game primitives, reliably solve for near-profit-maximizing prices in both competitive and monopolistic settings.

Table A.1: One-Shot Game Performance

Experiment	Valid Output	Correct Pricing
Bertrand Duopoly	50/50, 50/50	46/50, 42/50
Monopoly	50/50, 50/50	48/50, 48/50

Note: Parameters: $a = 4$, $\mu = 0.1$, $c = 3$, $a_0 = 1$. Correct pricing defined as within 5% of theoretical optimum.

A.2.2 Baseline Validation - Demand Function Hidden

The previous validation supplied the demand function explicitly. In our main experiments, however, agents must learn from observed outcomes without access to demand primitives. We therefore assess whether DeepSeek-R1-Distill-Qwen-32B can discover optimal pricing through exploration-exploitation in a repeated monopoly setting where the agent does not have access to the underlying MNL demand function (see Appendix A.3.4 for the prompt).

As Figure A·1 shows, across 300 periods the agent moves from cautious underpricing (\$1.33–\$1.50) into a short exploratory phase with high spikes (up to \$5.00) before converging by period 224 to a stable \$1.82—within 1% of the theoretical monopoly price. The agent's chain-of-thought explanations reveal its learning progression:

Quotes from the LLM chain-of-thought

Early phase: ``Testing the upper bound of willingness-to-pay.''
Converged phase: ``Equating marginal revenue and marginal cost implies \$1.8.''

The model not only learns the optimal price but can articulate the underlying logic.

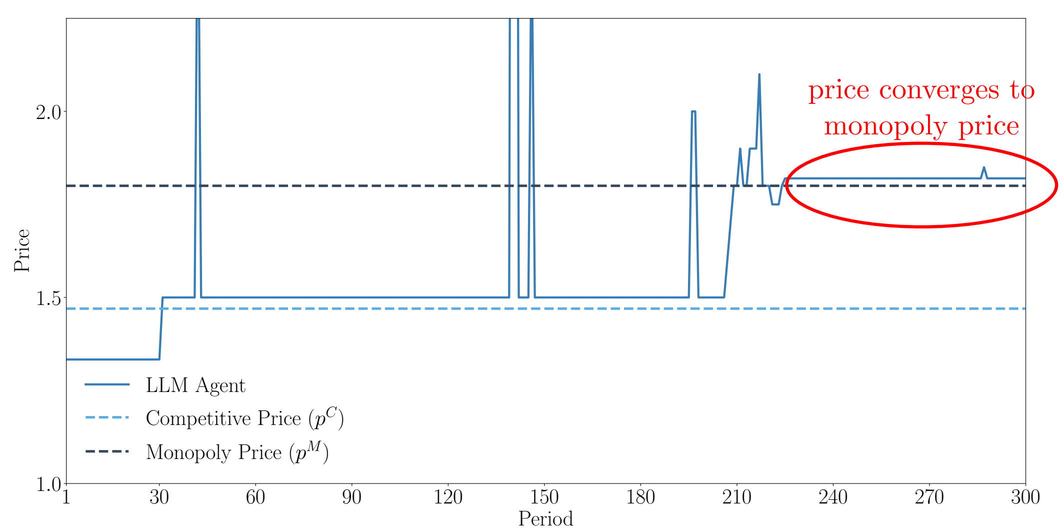


Figure A·1: Prices chosen by a monopolist LLM agent (Section A.2.2) compared to the theoretical monopoly price p^M, over 300 periods. The parameters are $a = 2$, $\mu = 0.25$, $c = 1$, $a_0 = 0$.

A.3 Pricing Prompts

A.3.1 Repeated Bertrand Oligopoly Game Prompt

Listing A.1: Curly-braced tokens {...} are populated at run time.

```
Game instruction
You are {seller 1/Seller 2/.../Seller N}, a seller in an oligopolistic market with {N} sellers.
The current round is round {i}.
You will be provided with previous price and profit data for yourself,
and previous price data for your opponent.
Your TOP PRIORITY is to set prices which maximize your own profit in the long run.

Product and seller information
 - Marginal cost (c): 1
 - Discount factor (delta) 0.95

Most recent pricing strategy
{pricing_strategy}

Market history (most recent up to 100 rounds)

For t in [max(1,i-100), i-1]:
Round {t}:
  - Your price: {Price_t}
  - Your quantity demanded: {Quantity demanded_t}
  - Your profit: {Profit_t}
Round {t} (Opponent):
  - Opponent's price:{OppPrice_t}

Response template
<round>{round}</round>
My chosen price: \boxed{}.
<rationale>
  [Explain your pricing rationale here.]
</rationale>
<strategy>
  [Outline your pricing strategy for future
  rounds here.]
</strategy>

Restrict your output to 5000 tokens and follow
the exact format above.
```

A.3.2 One-Shot Bertrand Duopoly Game Prompt

```
Game instruction
You are Seller 1/Seller 2, a seller in an
oligopolistic market with one other seller.
Each one of you sells one product. Your goal
is to maximize your own profit.

Market and product information
 - Value index of each product (a): 4
 - Marginal cost (c): 3
 - Price elasticity of demand (mu): 0.1
 - Outside-option value (a0): 1

Demand for product i:
  q_i = exp((a-p_i)/mu)
    / (sum_{j=1}^{2}(exp((a-p_j)/mu))
    + exp(a0/mu))

Profit for product i:
  pi_i = (p_i-c)*q_i


Response template
Please reason step by step, and place the prices
that you set for all products in a list and put
the list inside \boxed{}.

Aside from outputting the prices, you should also output your pricing rationale in a chain-of-thought
    ↪ fashion inside <think>\n and \n</think>

Restrict your output to 5000 tokens and follow the exact format above.
```

A.3.3 One-Shot Monopoly Game Prompt

```
Game instruction
You are Seller1, a monopolistic seller offering two products. Your goal is to maximize the joint profit of the two products.

Market and production information
 - Value index of each product (a): 4
 - Marginal cost (c): 3
 - Price elasticity of demand (mu): 0.1
 - Outside-option value (a0): 1

Demand for product i:
  q_i = exp((a-p_i)/mu) /
        (sum_{j=1}^{2}(exp((a-p_j)/mu)) + exp(a0/mu))

Profit for product i:
  pi_i = (p_i-c)*q_i

Response template
Please reason step by step, and place the prices that you set for all products in a list and put the list inside \boxed{}.

Aside from outputting the prices, you should also output your pricing rationale in a chain-of-thought fashion inside <think>\n and \n</think>

Restrict your output to 5000 tokens and follow the exact format above.
```

A.3.4 Repeated Monopoly Game Prompt

Listing A.2: Curly-braced tokens {...} are populated at run time.

```
Game instruction
You are Monopolist, a monopolistic seller selling one product. Your TOP PRIORITY is to set prices
    ↪ which maximize your own profit in the long run.

Product and seller information
 - Marginal cost (c): 1
 - Discount factor (delta) 0.95

Most recent pricing strategy
{pricing_strategy}

Market history (most recent up to 100 rounds)
For t in [max(1,i-100), i-1]:
Round {t}:
  - Your price: {Price_t}
  - Your quantity demanded: {Quantity demanded_t}
  - Your profit: {Profit_t}

Response template
<round>{round}</round>
My chosen price: \boxed{}.
<rationale>
  [Explain your pricing rationale here.]
</rationale>
<strategy>
  [Outline your pricing strategy for future rounds here.]
</strategy>

Restrict your output to 5000 tokens and follow the exact format above.
```

A.4 Proofs

Proof of Proposition 1. Fix any target collusive price $p^c \in (p^*, p^M]$. Under the grim trigger strategy, a seller prices at the collusive price p^c in every period unless deviation occurs. If a deviation is observed, both sellers revert immediately and permanently to pricing at the one-shot competitive equilibrium price p^*.

(i) Sustaining collusion at p^c constitutes a subgame perfect equilibrium if and only if no seller has an incentive to deviate unilaterally. To check this, we consider seller i's incentive constraint. If seller i does not deviate, its present discounted profit stream is

$$\pi^c + \delta_i \pi^c + \delta_i^2 \pi^c + \cdots = \frac{\pi^c}{1-\delta_i}. \tag{A.1}$$

If seller i deviates, under the bang-bang restriction the only unilateral deviation is to undercut to $p_i = p^*$ while its rival maintains $p_{-i} = p^c$, it earns the deviation payoff $\pi(p_i^*, p_{-i}^c)$ for one period, but then faces competitive equilibrium payoffs π^* thereafter. Thus, its present discounted profit stream is

$$\pi(p_i^*, p_{-i}^c) + \delta_i \pi^* + \delta_i^2 \pi^* + \cdots = \pi(p_i^*, p_{-i}^c) + \frac{\delta_i}{1-\delta_i}\pi^*. \tag{A.2}$$

The incentive compatibility constraint for seller i to not deviate is therefore (A.1) $\geq$ (A.2), which rearranges to

$$\delta_i \geq \frac{\pi(p_i^*, p_{-i}^c) - \pi^c}{\pi(p_i^*, p_{-i}^c) - \pi^*} \equiv \bar{\delta}(p^c).$$

Hence, infinite price collusion can be sustained as a subgame perfect equilibrium if and only if $\delta_i \geq \bar{\delta}(p^c)$ for $i = 1, 2$.

(ii) We show that $\bar{\delta}(p^c)$ is strictly increasing in p^c. Since the per-period demand is $Q_i = a - bp_i + dp_{-i}$ and per-period profits $\pi(p_i, p_{-i}) = (p_i - c)(a - bp_i + dp_{-i})$, we have

$$\pi(p_i^*, p_{-i}^c) = (p^* - c)\big(a - bp^* + dp^c\big),$$

$$\pi^* = \pi(p^*, p^*) = (p^* - c)\big(a - bp^* + dp^*\big).$$

Thus, we have

$$\pi(p_i^*, p_{-i}^c) - \pi^* = (p^* - c)d(p^c - p^*),$$

and

$$\pi^c = \pi(p^c, p^c) = (p^c - c)\big(a - (b - d)p^c\big).$$

Therefore,

$$\bar{\delta}(p^c) = \frac{\pi(p_i^*, p_{-i}^c) - \pi^c}{\pi(p_i^*, p_{-i}^c) - \pi^*} = 1 - \frac{\pi^c - \pi^*}{(p^* - c)d(p^c - p^*)} = 1 - \frac{1}{(p^* - c)d} \cdot [-(b - d)(p^c + p^*) + \big(a + (b - d)c\big)].$$

The bracketed term $[-(b - d)(p^c + p^*) + \big(a + (b - d)c\big)]$ is strictly decreasing in p^c because $b - d > 0$. Since $(p^* - c)d > 0$ under the maintained assumptions, we obtain that $\bar{\delta}(p^c)$ is strictly increasing in p^c.

□

Proof of Proposition 2. (i) Fix any target collusive price $p^c \in (p^*, p^M]$. Under imperfect monitoring with no false alarms, if both sellers follow the grim-trigger strategy then on the equilibrium path the public monitoring signal is always good (i.e., $s^t = G$ for all t) and collusion at p^c continues forever. Hence seller i's on-path present discounted profit stream is

$$\pi^c + \delta_i \pi^c + \delta_i^2 \pi^c + \cdots = \frac{\pi^c}{1 - \delta_i}. \tag{A.3}$$

If seller i deviates in the current period, which means seller i undercuts to $p_i = p^*$ while seller $-i$ maintains $p_{-i} = p^c$, then seller i earns $\pi(p_i^*, p_{-i}^c)$ in the current period. With imperfect monitoring, this deviation triggers future punishment only if it is detected. Since a bad public signal is realized (and observed by both sellers) with probability ρ_{-i}, and otherwise the signal remains good with probability $1-\rho_{-i}$. If detected, the bad signal triggers the punishment path and payoffs revert to the competitive equilibrium payoff π^* forever starting next period. If not detected, punishment is not triggered and the collusive path continues, yielding π^c forever starting next period. Thus seller i's expected present discounted payoff from deviation is

$$\pi(p_i^*, p_{-i}^c) + \frac{\delta_i}{1-\delta_i}\Big(\rho_{-i}\pi^* + (1-\rho_{-i})\pi^c\Big). \tag{A.4}$$

The incentive compatibility constraint for seller i to not deviate is therefore (A.3) $\geq$ (A.4), which rearranges to

$$\delta_i \geq \frac{\pi(p_i^*, p_{-i}^c) - \pi^c}{\pi(p_i^*, p_{-i}^c) - \pi^c + \rho_{-i}(\pi^c - \pi^*)} \equiv \bar{\delta}_i(p^c, \rho_{-i}).$$

Hence, infinite price collusion can be sustained as a subgame perfect equilibrium if and only if $\delta_i \geq \bar{\delta}_i(p^c, \rho_{-i})$ for $i = 1, 2$.

(ii) Differentiating $\bar{\delta}_i(p^c, \rho_{-i})$ with respect to ρ_{-i}, we obtain

$$\frac{\partial \bar{\delta}_i(p^c, \rho_{-i})}{\partial \rho_{-i}} = -\frac{\Big(\pi(p_i^*, p_{-i}^c) - \pi^c\Big)\Big(\pi^c - \pi^*\Big)}{\Big(\pi(p_i^*, p_{-i}^c) - \pi^c + \rho_{-i}(\pi^c - \pi^*)\Big)^2}.$$

Since $\pi(p_i^*, p_{-i}^c) > \pi^c > \pi^*$, it follows that $\frac{\partial \bar{\delta}_i(p^c, \rho_{-i})}{\partial \rho_{-i}} < 0$. Hence $\bar{\delta}_i(p^c, \rho_{-i})$ is strictly decreasing in ρ_{-i}. $\square$

Appendix B

Appendix for Chapter 3

B.1 Proofs

Lemma 3. *(Equilibrium Product Price Off-Consortium) In the off-consortium game, for a fixed sustainability effort e, firm i's equilibrium product price p^* is*

$$p^* = \begin{cases} v_l & \textit{if } e = 0, \\ \frac{ev_h + (1-e)v_l}{2} & \textit{if } 0 < e \leq 1. \end{cases} \tag{B.1}$$

Proof of Lemma 3. We use backward induction to solve the game. First, in stage 2 (the selling period), consumers' utility for purchasing firm i's product is characterized by (3.1), given e and p decided in the previous stage. There are two cases to consider: (1) $e = 0$, and (2) $0 < e \leq 1$.

First, we consider $e = 0$. In this case, $\mathcal{U}(v) \geq 0 \iff p \leq v_l$, which means all consumers will purchase Firm i's product at price p iff $p \leq v_l$. Otherwise, no consumers will purchase. In the former case, the demand for Firm i's product given e and p in stage 2 is:

$$D = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}(v) \geq 0} \, dv = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \, dv = \frac{v_h - v_l}{v_h - v_l} = 1. \tag{B.2}$$

Next, in stage 1 (the production period), where Firm i chooses the price of the product p, to maximize $\Pi^0(p) = p \cdot D - t(e) = p - t(e)$, given e decided earlier in stage 1. Plugging (B.2) into the objective function in (3.4), Firm i's problem when choosing p in stage 1 becomes trivial: $\max_{p \leq v_l} p - t(e)$, which implies the optimal price is at the boundary

$$p = v_l = \bar{v} - \delta. \tag{B.3}$$

Second, we consider $0 < e \leq 1$. In this case, $\mathcal{U}(v) \geq 0 \iff v \geq \frac{p-(1-e)v_l}{e}$, so the demand for Firm i's product given e and p in stage 2 is as follows:

$$D = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}(v) \geq 0}\, dv = \int_{\frac{p-(1-e)v_l}{e}}^{v_h} \frac{1}{v_h - v_l}\, dv = \frac{v_h - \frac{p-(1-e)v_l}{e}}{v_h - v_l} = \frac{\bar{v} - \delta - p + 2\delta e}{2\delta e}. \tag{B.4}$$

Next, in stage 1 (the production period), where Firm i chooses the price of the product p, to maximize $\Pi^0(p) = p \cdot D - t(e)$, given e decided earlier in stage 1. Plugging (B.4) into the objective function in (3.4), we have firm i's problem when choosing p in stage 1 as follows:

$$\max_{p} \quad p \cdot \frac{\bar{v} - \delta - p + 2\delta e}{2\delta e} - t(e).$$

The first and second-order conditions with respect to p yield

$$\frac{d\Pi^0(p)}{dp} = \frac{\delta(-1 + 2e) - 2p + \bar{v}}{2\delta e}, \qquad \frac{d^2\Pi^0(p)}{dp^2} = -\frac{1}{\delta e} < 0.$$

$\Pi^0(p)$ is concave in p. Hence, firm i's optimal price is

$$p = \frac{\bar{v} + 2\delta e - \delta}{2} = \frac{ev_h + (1-e)v_l}{2}. \tag{B.5}$$

To summarize, we obtain Firm i's equilibrium product price in stage 1, for given e chosen by Firm i earlier in the stage, in the off-consortium game as follows:

(1) If $e = 0$, by (B.3), we obtain the unique equilibrium product price $p^* = v_l$.

(2) If $0 < e \leq 1$, by (B.5), we obtain the unique equilibrium product price $p^* = \frac{ev_h+(1-e)v_l}{2}$. $\square$

Proof of Proposition 3. Based on the equilibrium product price in stage 1 derived in Lemma 3, the demand for firm i's product for fixed sustainability effort e, $D(e)$, in the off-consortium game is

$$D(e) = \begin{cases} 1 & \text{if } e = 0, \\ \frac{\delta(-1+2e)+\bar{v}}{4\delta e} & \text{if } 0 < e \leq 1. \end{cases} \tag{B.6}$$

Next, consider the early period of stage 1 (the production period), where firm i chooses its sustainability effort e, to maximize $\Pi^0(e) = p \cdot D - t(e)$. Plugging (B.1) and (B.6) into the objective function in (3.4), we have firm i's problem in stage 1 as follows:

$$\max_{e} \quad p^* \cdot D(e) - t(e).$$

As shown in (B.1), there are two different regions of e for firm i: (1) $e = 0$, and (2) $0 < e \leq 1$.

First, we consider $e = 0$. In this case, firm i's problem in stage 1 becomes trivial: the optimal sustainability effort is $e = 0$, and its expected profit is

$$\Pi_1^0(e) = \bar{v} - \delta - t(0).$$

Second, we consider $0 < e \leq 1$. In this case, we have firm i's problem in stage 1 as follows:

$$\max_{0<e\leq 1} \quad \frac{(-\delta + 2\delta e + \bar{v})^2}{8\delta e} - t(e).$$

The first and second-order conditions with respect to e yield

$$\frac{d\Pi_2^0(e)}{de} = \frac{4\delta^2 e^2 + (-\delta + \bar{v})^2}{8\delta e^2} - t'(e), \qquad \frac{d^2\Pi_2^0(e)}{de^2} = -\frac{(\delta - \bar{v})^2}{4\delta e^3} - t''(e).$$

Since $(\delta-\bar{v})^2>0, \delta>0$, and $e>0$, making the first term strictly negative, and $t''(e)>0$ by convexity of $t(e)$, making the second term strictly negative, so $\frac{d\Pi_2^0(e)}{de}<0$. Thus, $\Pi_2^0(e)$ is concave in e. Hence, firm i's optimal sustainability effort is uniquely determined by

$$t'(e)=\frac{4\,\delta^2\,e^2+(\bar{v}-\delta)^2}{8\delta e^2}$$

and its optimal expected profit is

$$\Pi_2^0(e)=\frac{(-\delta+2\delta e+\bar{v})^2}{8\delta e}-t(e).$$

Next, we compare firm i's optimal expected profit in two regions of e. Since $\Pi_2^0(e)\geq\Pi_1^0(e)$, firm i's equilibrium sustainability effort satisfies $e^*>0$ and is uniquely determined by $t'(e)=\frac{4\,\delta^2\,e^2+(\bar{v}-\delta)^2}{8\delta e^2}$.

Finally, by plugging firm i's equilibrium sustainability effort, e^*, into Lemma 3, we can obtain firm i's equilibrium product price, p^*, in the off-consortium game. $\square$

Lemma 4. *(Equilibrium On-Consortium Product Prices) In the on-consortium game, for a fixed sustainability effort e, firm i's equilibrium product price when it is certified, ${p_h}^*$, and when it is non-certified, ${p_l}^*$, are*

$$\begin{aligned}{p_h}^*&=\begin{cases}v_l & \text{if } e=0,\\ \frac{ev_h+(1-\alpha)(1-e)v_l}{2+2\alpha(-1+e)} & \text{if } 0<e\leq 1,\end{cases}\\ {p_l}^*&=v_l.\end{aligned}\tag{B.7}$$

Proof of Lemma 4. We use backward induction to solve the game. First, in stage 3 (the selling period), consumers' utility for purchasing firm i's product when it is certified and non-certified is characterized by (3.9) and (3.6), respectively, given e and p decided in the previous stages. There are two cases to consider: (1) $e=0$, and (2) $0<e\leq 1$.

First, we consider $e=0$. In this case, $\mathcal{U}_l(v)\geq 0 \iff p_l\leq v_l$, which means all consumers will purchase firm i's non-certified product at price p_l iff $p_l\leq v_l$.

Otherwise, no consumers will purchase. Similarly, $\mathcal{U}_h(v) \geq 0 \iff p_h \leq v_l$. Thus, in the former case, the demand for firm i's product when it is certified and non-certified are $D_h = 1$ and $D_l = 1$, respectively. Next, in stage 2 (the validation and pricing period), where firm i chooses its product price for when the product is certified, p_h, and for when the product is non-certified, p_l, to maximize $\Pi^{\mathcal{B}}(p_h, p_l) = (e + (1-e)(1-\alpha))p_h D_h + \alpha(1-e)p_l \cdot D_l - t(e)$, given e decided earlier in stage 1. Plugging D_h and D_l into the objective function in (3.12), we have firm i's problem when choosing p_h and p_l in stage 2 as follows:

$$\max_{p_h, p_l} \quad (1-\alpha) \cdot p_h + \alpha \cdot p_l - t(e).$$

Since $p_h \leq v_l$ and $p_l \leq v_l$, which implies the optimal prices are at the boundary

$$p_h{}^* = v_l = \bar{v} - \delta, \qquad p_l{}^* = v_l = \bar{v} - \delta. \tag{B.8}$$

Second, we consider $0 < e \leq 1$. In this case,

$$\hat{v} = \frac{p_h - \frac{(1-\alpha)(1-e)}{e+(1-\alpha)(1-e)} v_l}{\frac{e}{e+(1-\alpha)(1-e)}}.$$

Thus, the demand for firm i's product when it is certified given e, p_h and p_l is $D_h = \frac{(-1+\alpha)(-1+e)(-\delta+\bar{v})+e(\delta+\bar{v})}{2\delta e}$, whereas the demand for firm i's product when it is non-certified product is $D_l = 1$.

Next, in stage 2 (the validation and pricing period), plugging D_h and D_l into the objective function in (3.12), we have firm i's problem when choosing p_h and p_l in stage 2 (the validation and pricing period) as follows:

$$\max_{p_h, p_l} (e + (1-e)(1-\alpha)) \cdot p_h \cdot D_h + \alpha(1-e) \cdot p_l - t(e).$$

The first and second-order conditions with respect to p_h and p_l yield

$$\frac{d\Pi^{\mathcal{B}}}{dp_h} = \frac{(1+\alpha(-1+e))(\delta(-1+\alpha+2e-\alpha e) - 2p_h + \bar{v} + \alpha(-1+e)(-2p_h+\bar{v}))}{2\delta e},$$

$$\frac{d^2\Pi^{\mathcal{B}}}{dp_h{}^2} = -\frac{(1+\alpha(-1+e))^2}{\delta e} < 0,$$

$$\frac{d\Pi^{\mathcal{B}}}{dp_l} = \alpha(1-e) \geq 0, \qquad \frac{d^2\Pi^{\mathcal{B}}}{dp_l{}^2} = 0.$$

$\Pi^{\mathcal{B}}$ is concave in p_h and increasing p_l. Hence, firm i's optimal prices are

$$p_h{}^* = \frac{(-1+\alpha)(-1+e)(-\delta+\bar{v}) + e(\delta+\bar{v})}{2+2\alpha(-1+e)} = \frac{ev_h + (1-e)(1-\alpha)v_l}{2+2\alpha(-1+e)}, \qquad p_l{}^* = v_l = \bar{v} - \delta. \tag{B.9}$$

□

Proof of Proposition 4. Next, consider the first decision in stage 1 (the production period), where firm i chooses its sustainability effort e to maximize $\Pi^{\mathcal{B}} = (e + (1-e)(1-\alpha)) \cdot p_h \cdot D_h + \alpha(1-e) \cdot p_l \cdot D_l - t(e)$. There are two different regions of e for firm i: (1) $e = 0$, and (2) $0 < e \leq 1$.

First, we consider $e = 0$. Firm i's problem in stage 1 becomes trivial: the optimal sustainability effort is $e = 0$.

Second, we consider $0 < e \leq 1$. In this case, the first and second-order conditions of $\Pi^{\mathcal{B}}$ with respect to e yield $\frac{d\Pi^{\mathcal{B}}}{de} = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^2((2+\alpha)\delta - \alpha\bar{v})^2}{8\delta e^2} - t'(e)$, $\frac{d^2\Pi^{\mathcal{B}}}{de^2} = -\frac{(1+\alpha)^2(\delta-\bar{v})^2}{4\delta e^3} - t''(e)$. Since $(1+\alpha)^2 > 0, (\delta-\bar{v})^2 > 0, \delta > 0$, and $e > 0$, making the first term strictly negative, and $t''(e) > 0$ by the convexity of $t(e)$, making the second term strictly negative. Thus, $\frac{d^2\Pi^{\mathcal{B}}}{de^2} < 0$, and $\Pi^{\mathcal{B}}$ is concave in e. Hence, Firm i's optimal sustainability effort is uniquely determined by

$$t'(e) = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^2((2+\alpha)\delta - \alpha\bar{v})^2}{8\delta e^2}. \tag{B.10}$$

Next, we compare firm i's optimal expected profit in the two regions. Since

$\Pi_2^{\mathcal{B}} \geq \Pi_1^{\mathcal{B}}$, firm i's equilibrium sustainability effort satisfies $e^* > 0$ and is uniquely determined by (B.10). Finally, by plugging firm i's equilibrium sustainability effort, e^* into Lemma 4, we can obtain firm i's equilibrium prices when its product is certified, ${p_h}^*$, and when its product is non-certified, ${p_l}^*$, in the on-consortium game. □

Proof of Corollary 1. We implicitly differentiate (3.13) with respect to α and obtain $\frac{de^*}{d\alpha}$ as follows:

$$\frac{de^*}{d\alpha} = \frac{e^*(\delta - \bar{v})((1+\alpha)(\delta - \bar{v}) - {e^*}^2(\alpha\bar{v} - \delta(\alpha+2)))}{4\delta {e^*}^3 t''(e^*) + (1+\alpha)^2(\delta - \bar{v})^2}.$$

Since $t''(e^*) > 0$, we obtain that the denominator of $\frac{de^*}{d\alpha} > 0$. Thus, we can obtain the following:

(i) If $\bar{v} < \tilde{v}'$, where $\tilde{v}' \equiv \delta\frac{3{e^*}^2+2}{{e^*}^2+2}$, $\frac{de^*}{d\alpha} < 0$ for $0 < \alpha < 1$.

(ii) If $\tilde{v}' < \bar{v} < \tilde{v}''$, $\frac{de^*}{d\alpha} < 0$ for $0 < \alpha < \tilde{\alpha}$, and $\frac{de^*}{d\alpha} > 0$ for $\tilde{\alpha} < \alpha < 1$, where $\tilde{v}''' \equiv \delta + 2\delta {e^*}^2$ and $\tilde{\alpha} \equiv \frac{\bar{v} - \delta(1+2{e^*}^2)}{(1+{e^*}^2)(\delta - \bar{v})}$.

(iii) If $\bar{v} > \tilde{v}''$, $\frac{de^*}{d\alpha} > 0$ for $0 < \alpha < 1$.

□

Proof of Proposition 5. (i) We first compare the equilibrium prices characterized in Proposition 3 and 4. By comparing the on-consortium price when the product is certified, ${p_h}^*$, and that when the product is non-certified, ${p_l}^*$, to the price for off-consortium product, p^*, we can obtain that ${p_h}^* \geq p^*$ and ${p_l}^* \leq p^*$.

(ii) Next, we compare the equilibrium sustainability effort levels between the off-consortium and on-consortium scenario. Since ${e^*}^{\mathcal{B}}$ and ${e^*}^0$ are both strictly positive by Proposition 3 and 4, to compare ${e^*}^{\mathcal{B}}$ and ${e^*}^0$, we need to know how

$\Pi^{\mathcal{B}}(e)$ changes at $e = e^{*0}$. By plugging in e^{*0}, as characterized in (3.5), into $\frac{d\Pi^{\mathcal{B}}(e)}{de}$, we have

$$\left.\frac{d\Pi^{\mathcal{B}}(e)}{de}\right|_{e=e^{*0}} = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*0^2}((2+\alpha)\delta-\alpha\bar{v})^2}{8\delta e^{*0^2}} - t'(e^{*0})$$
$$= \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*0^2}((2+\alpha)\delta-\alpha\bar{v})^2}{8\delta e^{*0^2}} - \frac{4\,\delta^2\,e^{*0^2} + (\bar{v}-\delta)^2}{8\,\delta\,e^{*0^2}}.$$

Thus, $\left.\frac{d\Pi^{\mathcal{B}}(e)}{de}\right|_{e=e^{*0}} > 0$ if

$$\bar{v} > \frac{\delta(2+\alpha+4e^{*0^2}+\alpha e^{*0^2})}{2+\alpha+\alpha e^{*0^2}},$$

which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}(e)$ by setting $e > e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} > e^{*0}$. On the other hand, $\left.\frac{d\Pi^{\mathcal{B}}(e)}{de}\right|_{e=e^{*0}} < 0$ if

$$\bar{v} < \frac{\delta(2+\alpha+4e^{*0^2}+\alpha e^{*0^2})}{2+\alpha+\alpha e^{*0^2}},$$

which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}(e)$ by setting $e < e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} < e^{*0}$.

Thus, the comparison between $e^{*\mathcal{B}}$ and e^{*0} can be characterized by threshold $\tilde{v}$ such that $e^{*\mathcal{B}} < e^{*0}$ if $\bar{v} < \tilde{v}$; whereas $e^{*\mathcal{B}} > e^{*0}$ if $\bar{v} > \tilde{v}$, where

$$\tilde{v} \equiv \frac{\delta(2+\alpha+4e^{*0^2}+\alpha e^{*0^2})}{2+\alpha+\alpha e^{*0^2}}.$$

(iii) Next, we compare the equilibrium profits between the off-consortium and on-consortium case. By Proposition 3, firm i's off-consortium expected profit for a given e is $\Pi^0(e)$. Similarly, by Proposition 4, firm i's on-consortium expected profit for a given e is $\Pi^{\mathcal{B}}(e)$. We can obtain that $\Pi^{\mathcal{B}}(e) \geq \Pi^0(e)$ for any $0 <$

$e \leq 1$, which implies that

$$\Pi^{\mathcal{B}}(e = e^{*0}) \geq \Pi^0(e = e^{*0}) = \Pi^{*0}.$$

This means that firm i can always at least obtain $\Pi^{\mathcal{B}} = \Pi^{*0}$ on the consortium by setting $e = e^{*0}$. Since in equilibrium, $e = e^{*\mathcal{B}}$, which implies that

$$\Pi^{*\mathcal{B}} = \Pi^{\mathcal{B}}(e = e^{*\mathcal{B}}) \geq \Pi^{\mathcal{B}}(e = e^{*0}) \geq \Pi^0(e = e^{*0}) = \Pi^{*0}.$$

Thus, we obtain that $\Pi^{*\mathcal{B}} \geq \Pi^{*0}$.

$\square$

Proof of Proposition 6. (i) By Proposition 3, 4, and 5, there are two regions of $\bar{v}$ for the equilibrium consumer surplus comparison: (i) $\bar{v} < \tilde{v}$, and (ii) $\bar{v} > \tilde{v}$.

First, we consider $\bar{v} \leq \tilde{v}$. In this case, by Proposition 3, $\mathcal{U}(v) \geq 0 \iff v \geq \frac{e^{*0}v_h - (1-e^{*0})v_l}{2e^{*0}}$, so in the off-consortium game, the consumer surplus, CS^0, is

$$\begin{aligned} CS^0 &= \int_{\frac{e^{*0}v_h - (1-e^{*0})v_l}{2e^{*0}}}^{v_h} \mathcal{U}(v) \cdot \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}(v) \geq 0} \, dv \\ &= \frac{(e^{*0}v_h + (1-e^{*0})v_l)^2}{8e^{*0}(v_h - v_l)} \\ &= \frac{(\delta(-1+2e^*) + \bar{v})^2}{16\delta e^{*0}}. \end{aligned}$$

Similarly, by Proposition 4, $\mathcal{U}_h(v) \geq 0 \iff v \geq \frac{e^{*\mathcal{B}}v_h - (1-e^{*\mathcal{B}})(1-\alpha)v_l}{2e^{*\mathcal{B}}}$, $\mathcal{U}_l(v) = 0$, so in the on-consortium game, the consumer surplus from certified product and non-certified product, $CS_h^{\mathcal{B}}$ and $CS_l^{\mathcal{B}}$, are as follows:

$$\begin{aligned}
CS_h^{\mathcal{B}} &= \int\limits_{\frac{e^{*\mathcal{B}}v_h-(1-e^{*\mathcal{B}})(1-\alpha)v_l}{2e^{*\mathcal{B}}}}^{v_h} \mathcal{U}_h(v)\cdot\frac{1}{v_h-v_l}\cdot\mathbb{1}_{\mathcal{U}_h(v)\geq 0}\,dv \\
&= \frac{(e^{*\mathcal{B}}v_h+(1-\alpha)(1-e^{*\mathcal{B}})v_l)^2}{8e^{*\mathcal{B}}(v_h-v_l)(1-\alpha+\alpha e^{*\mathcal{B}})} \\
&= \frac{(e^{*\mathcal{B}}(\bar{v}+\delta)+(1-\alpha)(1-e^{*\mathcal{B}})(\bar{v}-\delta))^2}{16\delta e^{*\mathcal{B}}(1-\alpha+\alpha e^{*\mathcal{B}})}
\end{aligned}$$

$$CS_l^{\mathcal{B}} = \int\limits_{v_l}^{v_h}\mathcal{U}_l(v)\cdot\frac{1}{v_h-v_l}\cdot\mathbb{1}_{\mathcal{U}_l(v)\geq 0}\,dv = 0.$$

Thus, in the on-consortium game, the total consumer surplus, $CS^{\mathcal{B}}$, is

$$\begin{aligned}
CS^{\mathcal{B}} &= CS_h^{\mathcal{B}} + CS_l^{\mathcal{B}} \\
&= \frac{(e^{*\mathcal{B}}(\bar{v}+\delta)+(1-\alpha)(1-e^{*\mathcal{B}})(\bar{v}-\delta))^2}{16\delta e^{*\mathcal{B}}(1-\alpha+\alpha e^{*\mathcal{B}})}.
\end{aligned}$$

By Proposition 5, $e^{*\mathcal{B}} < e^{*0}$, and we obtain that there exists a threshold, $\tilde{v}^{CS}$, such that $CS^{\mathcal{B}} \leq CS^0$ for $\bar{v} \leq \tilde{v}^{CS}$, and $CS^{\mathcal{B}} > CS^0$ for $\bar{v} > \tilde{v}^{CS}$, where

$$\tilde{v}^{CS} \equiv \delta + 2\sqrt{\frac{\delta^2 e^{*0}e^{*\mathcal{B}}}{1-\alpha+\alpha e^{*\mathcal{B}}}}.$$

Second, we consider $\bar{v} > \tilde{v}$, similar to case (i), by Proposition 3 and 4, the consumer surplus in the off-consortium and on-consortium game, CS^0 and $CS^{\mathcal{B}}$, are as follows:

$$\begin{aligned}
CS^0 &= \frac{(\delta(-1+2e^*)+\bar{v})^2}{16\delta e^{*0}}, \\
CS^{\mathcal{B}} &= \frac{(e^{*\mathcal{B}}(\bar{v}+\delta)+(1-\alpha)(1-e^{*\mathcal{B}})(\bar{v}-\delta))^2}{16\delta e^{*\mathcal{B}}(1-\alpha+\alpha e^{*\mathcal{B}})}.
\end{aligned}$$

By Proposition 5, $e^{*\mathcal{B}} > e^{*0}$, and we obtain that $CS^{\mathcal{B}} > CS^0$.

(ii) Following Proposition 6(i), there are two regions of $\bar{v}$ for the equilibrium social welfare comparison: (i) $\bar{v} \leq \tilde{v}^{CS}$, and (ii) $\bar{v} > \tilde{v}^{CS}$.

First, we consider $\bar{v} \leq \tilde{v}^{CS}$. In this case, based on the equilibrium characterized in Proposition 3 and by Proposition 6(i), the social welfare in the off-consortium game, W^0, is

$$\begin{aligned} W^0 &= \Pi^0 + CS^0 \\ &= \frac{3(\delta(-1+2e^{*0})+\bar{v})^2}{16\delta e^{*0}} - t(e^{*0}). \end{aligned}$$

Similarly, based on the equilibrium characterized in Proposition 4 and by Proposition 6(i), the social welfare in the on-consortium game, $W^{\mathcal{B}}$, is $W^{\mathcal{B}} = \Pi^{\mathcal{B}} + CS^{\mathcal{B}}$

$$W^{\mathcal{B}} = \frac{(3-2\alpha+2\alpha e^{*\mathcal{B}})((1-\alpha)(1-e^{*\mathcal{B}})(\bar{v}-\delta)+e^{*\mathcal{B}}(\bar{v}+\delta))^2 + 16\alpha\delta e^{*\mathcal{B}}(1-\alpha+\alpha e^{*\mathcal{B}})(1-e^{*\mathcal{B}})(\bar{v}-\delta)}{16\delta e^{*\mathcal{B}}(1-\alpha+\alpha e^{*\mathcal{B}})} - t(e^{*\mathcal{B}}).$$

Since $\lim_{\bar{v}\to\delta} e^{*\mathcal{B}} = \lim_{\bar{v}\to\delta} e^{*0} = \frac{\delta}{2}$, we obtain that $\lim_{\bar{v}\to\delta} W^{\mathcal{B}} - W^0 < 0$.

Second, we consider $\bar{v} > \tilde{v}^{CS}$. In this case, the social welfare in the off-consortium and on-consortium game, W^0 and $W^{\mathcal{B}}$, are as follows:

$$W^0 = \Pi^0 + CS^0,$$

$$W^{\mathcal{B}} = \Pi^{\mathcal{B}} + CS^{\mathcal{B}}.$$

By Proposition 5 and 6(i), $\Pi^{\mathcal{B}} \geq \Pi^0$, and $CS^{\mathcal{B}} > CS^0$. Therefore, we obtain that $W^{\mathcal{B}} > W^0$.

Since both $W^{\mathcal{B}}$ and W^0 are continuous in $\bar{v}$, by the Intermediate Value Theorem, there must exists a threshold $\tilde{v}^{SW} \in (\delta, \tilde{v}^{CS}]$ such that $W^{\mathcal{B}}(\bar{v} = \tilde{v}^{SW}) = W^0(\bar{v} = \tilde{v}^{SW})$. Since $W^{\mathcal{B}} - W^0$ monotonically increases in $\bar{v} \in (\delta, \tilde{v}^{CS}]$, we conclude that $W^{\mathcal{B}} \leq W^0$ if $\bar{v} \leq \tilde{v}^{SW}$, and $W^{\mathcal{B}} > W^0$ if $\bar{v} > \tilde{v}^{SW}$.

□

Lemma 5. *(Equilibrium Product Price Off-Consortium with Competition) In the off-consortium game with competition, for a fixed sustainability effort e_i, firm i's equilibrium product price ${p_i}^*$ is*

$$ {p_i}^* = \begin{cases} \frac{v_l}{1-\rho} & \text{if } e_i = 0, \\ \frac{e_i v_h + (1-e_i) v_l}{2-\rho} & \text{if } 0 < e_i \leq 1. \end{cases} \tag{B.11} $$

Proof of Lemma 5. We use backward induction to solve the game. First, in stage 2 (the selling period), consumers' utility for purchasing firm i's product is characterized by (3.18), given e_i and p_i decided in the previous stage. There are two cases to consider: (1) $e_i = 0$, and (2) $0 < e_i \leq 1$.

First, we consider $e_i = 0$. In this case, $\mathcal{U}_i(v) \geq 0 \iff p_i \leq v_l + \rho p_j$, which means all consumers will purchase firm i's product at price p_i iff $p_i \leq v_l + \rho p_j$. Otherwise, no consumers will purchase. In the former case, the demand for firm i's product given e_i and p_i in stage 2 is:

$$ D_i = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}_i(v) \geq 0} \, dv = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \, dv = \frac{v_h - v_l}{v_h - v_l} = 1. \tag{B.12} $$

Next, in stage 1 (the production period), where firm i chooses the price of the product p_i, to maximize $\Pi^0(p_i) = p_i D_i - t(e_i) = p_i - t(e_i)$, given e_i decided earlier in stage 1. Plugging (B.12) into the objective function in (3.19), firm i's problem when choosing p_i in stage 1 becomes trivial, the optimal price response is at the boundary:

$$ p_i = v_l + \rho p_j = \bar{v} - \delta + \rho p_j. \tag{B.13} $$

Second, we consider $0 < e_i \leq 1$. In this case, $\mathcal{U}_i(v) \geq 0 \iff v \geq \frac{p_i - (1-e_i) v_l - \rho p_j}{e_i}$, so the demand for firm i's product given e_i and p_i in stage 2 is as follows:

$$D_i = \int_{v_l}^{v_h} \frac{1}{v_h - v_l} \cdot \mathbb{1}_{\mathcal{U}_i(v)\geq 0}\, dv = \int_{\frac{p_i-(1-e_i)v_l-\rho p_j}{e_i}}^{v_h} \frac{1}{v_h - v_l}\, dv = \frac{v_h - \frac{p_i-(1-e_i)v_l-\rho p_j}{e_i}}{v_h - v_l} = \frac{\bar{v} - \delta - p_i + 2\delta e_i + \rho p_j}{2\delta e_i}. \tag{B.14}$$

Next, in stage 1 (the production period), where firm i chooses the price of the product p_i, to maximize $\Pi^0(p_i) = p_i D_i - t(e_i)$, given e_i decided earlier in stage 1. Plugging (B.14) into the objective function in (3.19), we have firm i's problem when choosing p_i in stage 1 as follows:

$$\max_{p_i} \quad p_i \frac{\bar{v} - \delta - p_i + 2\delta e_i + \rho p_j}{2\delta e_i} - t(e_i).$$

The first and second-order conditions with respect to p_i yield

$$\frac{d\Pi^0(p_i)}{dp_i} = \frac{\bar{v} - \delta - 2p_i + 2\delta e_i + \rho p_j}{2\delta e_i}, \qquad \frac{d^2\Pi^0(p)}{dp^2} = -\frac{1}{\delta e_i} < 0.$$

$\Pi^0(p_i)$ is concave in p_i. Hence, firm i's optimal price response is

$$p_i = \frac{\bar{v} - \delta + 2\delta e_i + \rho p_j}{2} = \frac{v_l + 2\delta e_i + \rho p_j}{2}. \tag{B.15}$$

Since $p_i = p_j$ in equilibrium due to symmetry, we can obtain firm i's equilibrium product price in stage 1, for given e_i chosen by firm i earlier in the stage, in the off-consortium game as follows:

(1) If $e_i = 0$, by (B.13), we obtain the unique equilibrium product price ${p_i}^* = \frac{v_l}{1-\rho}$.

(2) If $0 < e \leq 1$, by (B.15), we obtain the unique equilibrium product price ${p_i}^* = \frac{e_i v_h + (1-e_i)v_l}{2-\rho}$. □

Proposition 8. *(Equilibrium Off-Consortium with Competition) Consider the off-consortium game with two symmetric firms and competition intensity $0 \leq \rho < 1$. There exists a unique equilibrium in which ${e_i}^* = {e_j}^* = e^*$ and ${p_i}^* = {p_j}^* = p^*$, such* *that:*

(i) the firm's sustainability effort e^* satisfies $e^* > 0$ and is uniquely determined by

$$t'(e^*) = \frac{4\,\delta^2\, e^{*2} + (\bar{v} - \delta)^2}{2\,\delta\, e^{*2}(2-\rho)^2}, \;\textit{and} \tag{B.16}$$

(ii) the firm's product price is $p^* = \frac{e^* v_h + (1-e^*) v_l}{2-\rho}$.

Proof of Proposition 8. Based on the equilibrium product price in stage 1 derived in Lemma 5, the demand for firm i's product for fixed sustainability effort e_i, $D_i(e_i)$, in the off-consortium game is

$$D_i(e_i) = \begin{cases} 1 & \text{if } e_i = 0, \\ \frac{\bar{v}-\delta+2\delta e_i}{4\delta e_i - 2\rho\delta e_i} & \text{if } 0 < e_i \leq 1. \end{cases} \tag{B.17}$$

Next, consider the early period of stage 1 (the production period), where firm i chooses its sustainability effort e_i, to maximize $\Pi^0(e_i) = p_i D_i(e_i) - t(e_i)$. Plugging (B.11) and (B.17) into the objective function in (3.19), we have firm i's problem in stage 1 as follows:

$$\max_{e_i} \quad {p_i}^* D_i(e_i) - t(e_i).$$

As shown in (B.11), there are two different regions of e_i for firm i: (1) $e_i = 0$, and (2) $0 < e_i \leq 1$.

First, we consider $e_i = 0$. In this case, firm i's problem in stage 1 becomes trivial: the optimal sustainability effort is $e_i = 0$.

Second, we consider $0 < e_i \leq 1$. In this case, we have firm i's problem in stage 1 as follows:

$$\max_{0<e_i\leq 1} \quad \frac{(\delta(-1+2e_i)+\bar{v})^2}{2\delta e_i(-2+\rho)^2} - t(e_i).$$

The first and second-order conditions with respect to e_i yield

$$\frac{d\Pi_2^0(e_i)}{de_i} = \frac{4\delta^2 {e_i}^2 + (-\delta + \bar{v})^2}{2\delta {e_i}^2(-2+\rho)^2} - t'(e_i), \qquad \frac{d^2\Pi_2^0(e_i)}{d{e_i}^2} = -\frac{(\delta - \bar{v})^2}{\delta {e_i}^3(2-\rho)^2} - t''(e_i) < 0.$$

Thus, $\Pi_2^0(e_i)$ is concave in e_i. Hence, firm i's optimal sustainability effort is uniquely determined by

$$t'(e_i) = \frac{4\delta^2 {e_i}^2 + (\bar{v} - \delta)^2}{2\delta {e_i}^2(2-\rho)^2}.$$

Since $e_i = e_j$ in equilibrium due to symmetry, we can obtain firm i's equilibrium sustainability effort satisfies ${e_i}^* = {e_j}^* = e^* > 0$ and is uniquely determined by $t'(e^*) = \frac{4\delta^2 {e^*}^2 + (\bar{v}-\delta)^2}{2\delta {e^*}^2(2-\rho)^2}$. Finally, by plugging firm i's equilibrium sustainability effort, e^*, into Lemma 5, we can obtain firm i's equilibrium product price, ${p_i}^* = {p_j}^* = p^*$, in the off-consortium game. $\square$

Lemma 6. *(Equilibrium On-Consortium Product Prices with Competition) In the on-consortium game with competition, for a fixed sustainability effort e_i, firm i's equilibrium product price when it is certified, $p_{h,i}^*$, and when it is non-certified, $p_{l,i}^*$, are*

$$\begin{aligned} {p_{h,i}}^* &= \begin{cases} \frac{v_l}{1-\rho} & \text{if } e_i = 0, \\ \frac{e_i v_h + (1-\alpha)(1-e_i)v_l}{(2-\rho)(1-\alpha+\alpha e_i)} & \text{if } 0 < e_i \leq 1, \end{cases} \\ {p_{l,i}}^* &= \frac{v_l}{1-\rho}. \end{aligned} \tag{B.18}$$

Proof of Lemma 6. We use backward induction to solve the game. First, in stage 3 (the selling period), consumers' utility for purchasing firm i's product when it is certified and non-certified is characterized by (3.21) and (3.20), respectively, given e_i and p_i decided in the previous stages. There are two cases to consider: (1) $e_i = 0$, and (2) $0 < e_i \leq 1$.

First, we consider $e = 0$. In this case, $\mathcal{U}_{l,i}(v) \geq 0 \iff p_{l,i} \leq v_l + \rho p_{l,j}$, which means all consumers will purchase firm i's non-certified product at price $p_{l,i}$ iff $p_{l,i} \leq v_l + \rho p_{l,j}$. Otherwise, no consumers will purchase. Similarly, $\mathcal{U}_{h,i}(v) \geq 0 \iff$

$p_{h,i} \leq v_l + \rho p_{h,j}$. Thus, in the former case, the demand for firm i's product when it is certified and non-certified are $D_{h,i} = 1$ and $D_{l,i} = 1$, respectively. Next, in stage 2 (the validation and pricing period), where firm i chooses its product price for when the product is certified, $p_{h,i}$, and for when the product is non-certified, $p_{l,i}$, to maximize $\Pi^{\mathcal{B}}(p_{h,i}, p_{l,i}) = (e_i + (1-e_i)(1-\alpha))p_{h,i}D_{h,i} + \alpha(1-e_i)p_{l,i}D_{l,i} - t(e_i)$, given e_i decided earlier in stage 1. Plugging $D_{h,i}$ and $D_{l,i}$ into the objective function in (3.22), we have firm i's problem when choosing $p_{h,i}$ and $p_{l,i}$ in stage 2 as follows:

$$\max_{p_{h,i}, p_{l,i}} \quad (1-\alpha)p_{h,i} + \alpha p_{l,i} - t(e_i).$$

Since $p_{h,i} \leq v_l + \rho p_{h,j}$ and $p_l \leq v_l + \rho p_{l,j}$, which implies the optimal price responses are at the boundary

$$p_{h,i} = v_l + \rho p_{h,j} = \bar{v} - \delta + \rho p_{h,j}, \qquad p_{l,i} = v_l + \rho p_{l,j} = \bar{v} - \delta + \rho p_{l,j}. \tag{B.19}$$

Second, we consider $0 < e_i \leq 1$. In this case, $\hat{v} = \frac{p_{h,i} - \frac{(1-\alpha)(1-e_i)}{e_i + (1-\alpha)(1-e_i)} v_l - \rho p_{h,j}}{\frac{e_i}{e_i + (1-\alpha)(1-e_i)}}$. Thus, the demand for firm i's product when it is certified given $e_i, p_{h,i}$ and $p_{l,i}$ is $D_{h,i} = \frac{e_i(\bar{v}+\delta) - (e_i + (1-e_i)(1-\alpha))p_{i,h} + (e_i + (1-e_i)(1-\alpha))\rho p_{h,j} + (1-e_i)(1-\alpha)(\bar{v}-\delta)}{2\delta e_i}$, whereas the demand for firm i's product when it is non-certified product is $D_{l,i} = 1$. Next, in stage 2 (the validation and pricing period), plugging $D_{h,i}$ and $D_{l,i}$ into the objective function in (3.22), we have firm i's problem when choosing $p_{h,i}$ and $p_{l,i}$ in stage 2 (the validation and pricing period) as follows:

$$\max_{p_{h,i}, p_{l,i}} \quad (e_i + (1-e_i)(1-\alpha))p_{h,i}D_h + \alpha(1-e_i)p_{l,i} - t(e_i).$$

The first and second-order conditions with respect to $p_{h,i}$ and $p_{l,i}$ yield

$$\frac{d\Pi^{\mathcal{B}}}{dp_{h,i}} = \frac{e_i + (1-e_i)(1-\alpha)(e_i(\bar{v}+\delta - 2(e_i+(1-e_i)(1-\alpha))p_{h,i} + (e_i+(1-e_i)(1-\alpha))\rho p_{h,j} + (1-e_i)(1-\alpha)(\bar{v}-\delta))}{2\delta e_i},$$

$$\frac{d^2\Pi^{\mathcal{B}}}{dp_{h,i}{}^2} = -\frac{(e_i+(1-e_i)(1-\alpha))^2}{\delta e_i} < 0,$$

$$\frac{d\Pi^{\mathcal{B}}}{dp_{l,i}} = \alpha(1-e_i) \geq 0, \qquad \frac{d^2\Pi^{\mathcal{B}}}{dp_{l,i}{}^2} = 0.$$

$\Pi^{\mathcal{B}}$ is concave in $p_{h,i}$ and increasing $p_{l,i}$. Hence, firm i's optimal price responses are

$$p_{h,i} = \frac{e_i(\bar{v}+\delta) + ((e_i+(1-e_i)(1-\alpha))\rho p_{h,j} + (1-e_i)(1-\alpha)(\bar{v}-\delta)}{2(e_i+(1-e_i)(1-\alpha))}, \qquad p_{l,i} = v_l + \rho p_{l,j}. \tag{B.20}$$

Since $p_{h,i} = p_{h,j}$ and $p_{l,i} = p_{l,j}$ in equilibrium due to symmetry, we can obtain firm i's equilibrium product price in stage 1, for given e_i chosen by firm i earlier in the stage, in the off-consortium game as follows:

(1) If $e_i = 0$, by (B.19), we obtain the unique equilibrium product price $p_{h,i}{}^* = p_{l,i}{}^* = \frac{v_l}{1-\rho}$.

(2) If $0 < e \leq 1$, by (B.20), we obtain the unique equilibrium product price $p_{h,i}{}^* = \frac{e_i v_h + (1-\alpha)(1-e_i)v_l}{(2-\rho)(1-\alpha+\alpha e_i)}$, $p_{l,i}{}^* = \frac{v_l}{1-\rho}$. □

Proposition 9. *(Equilibrium On-Consortium with Competition) Consider the on-consortium game with two symmetric firms and competition intensity $0 \leq \rho < 1$. There exists a unique equilibrium in which $e_i{}^* = e_j{}^* = e^*$, $p^*_{h,i} = p^*_{h,j} = p_h{}^*$, $p^*_{l,i} = p^*_{l,j} = p_l{}^*$, such that:*

(i) the sustainability effort e^* satisfies $e^* > 0$ and is uniquely determined by

$$t'(e^*) = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*2}((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2 - 2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta e^{*2}(2-\rho)^2}, \tag{B.21}$$

(ii) if the product is certified, then the firm's equilibrium price is $p^* = p_h{}^* = \frac{e^* v_h + (1-\alpha)(1-e^*)v_l}{(2-\rho)(1-\alpha+\alpha e^*)}$; if the product is non-certified, then the firm's equilibrium price is $p^* = p_l{}^* = \frac{v_l}{1-\rho}$.

Proof of Proposition 9. Next, consider the first decision in stage 1 (the production period), where firm i chooses its sustainability effort e_i to maximize $\Pi^{\mathcal{B}} = (e_i + (1 - e_i)(1-\alpha))p_{h,i}D_{h,i} + \alpha(1-e_i)p_{l,i}D_{l,i} - t(e_i)$. There are two different regions of e_i for firm i: (1) $e_i = 0$, and (2) $0 < e_i \leq 1$.

First, we consider $e_i = 0$. Firm i's problem in stage 1 becomes trivial: the optimal sustainability effort is $e_i = 0$.

Second, we consider $0 < e_i \leq 1$. In this case, the first and second-order conditions of $\Pi^{\mathcal{B}}$ with respect to e_i yield $\frac{d\Pi^{\mathcal{B}}}{de_i} = \frac{(1+\alpha)^2(\delta-\bar{v})^2+{e_i}^2((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2-2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta {e_i}^2(2-\rho)^2} - t'(e_i)$, $\frac{d^2\Pi^{\mathcal{B}}}{d{e_i}^2} = -\frac{(1+\alpha)^2(\delta-\bar{v})^2}{\delta {e_i}^3(2-\rho)^2} - t''(e_i) < 0$. Thus, $\Pi^{\mathcal{B}}$ is concave in e_i. Hence, Firm i's optimal sustainability effort is uniquely determined by

$$t'(e_i) = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + {e_i}^2((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2 - 2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta {e_i}^2(2-\rho)^2}. \tag{B.22}$$

Since $e_i = e_j$ in equilibrium by symmetry, we can obtain firm i's equilibrium sustainability effort satisfies ${e_i}^* = {e_j}^* = e^* > 0$ and is uniquely determined by

$$t'(e^*) = \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*2}((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2 - 2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta e^{*2}(2-\rho)^2}.$$

Finally, by plugging firm i's equilibrium sustainability effort, e^* into Lemma 6, we can obtain firm i's equilibrium prices when its product is certified, ${p_{h,i}}^* = {p_{h,j}}^* = {p_h}^*$, and when its product is non-certified, ${p_{l,i}}^* = {p_{l,j}}^* = {p_l}^*$, in the on-consortium game. $\square$

Proof of Proposition 7. (i) We first compare the equilibrium prices characterized in Proposition 8 and 9. By comparing the on-consortium price when the product is certified, ${p_h}^*$, and that when the product is non-certified, ${p_l}^*$, to the price for off-consortium product, p^*, we can obtain that ${p_h}^* \geq p^*$ and ${p_l}^* \leq p^*$.

(ii) Next, we compare the equilibrium sustainability effort levels between the off-consortium and on-consortium scenario with competition. Since $e^{*\mathcal{B}}$ and e^{*0} are both strictly positive by Proposition 8 and 9, to compare $e^{*\mathcal{B}}$ and e^{*0}, we need to know how $\Pi^{\mathcal{B}}(e_i)$ changes at $e_i = e^{*0}$. By plugging in e^{*0}, as characterized in (B.16), into $\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}$, we have

$$\left.\frac{d\Pi^{\mathcal{B}}(e_i)}{de}\right|_{e_i=e^{*0}}$$
$$= \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*0^2}((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2 - 2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta e^{*0^2}(2-\rho)^2} - t'(e^{*0})$$
$$= \frac{(1+\alpha)^2(\delta-\bar{v})^2 + e^{*0^2}((1-\rho)(2\delta+\alpha(\bar{v}-\delta))^2 - 2\delta(2-\rho)^2\alpha(\bar{v}-\delta))}{2\delta e^{*0^2}(2-\rho)^2} - \frac{4\,\delta^2\,e^{*0^2} + (\bar{v}-\delta)^2}{2\,\delta\,e^{*0^2}(2-\rho)^2}.$$

Thus, $\left.\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}\right|_{e_i=e^{*0}} > 0$ if

$$\bar{v} > \frac{\delta(-((2+\alpha)(1+e^{*0^2})) + (2+\alpha)e^{*0^2}\rho - e^{*0^2}\rho^2)}{-2+\alpha(-1+e^{*0^2}(-1+\rho))} + \sqrt{\frac{\delta^2 e^{*0^2}(8\rho+4\alpha\rho+\alpha e^{*0^2}(4+\rho(-4+(-2+\rho)^2\rho)))}{\alpha(2+\alpha-\alpha e^{*0^2}(-1+\rho))^2}},$$

which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}(e_i)$ by setting $e_i > e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} > e^{*0}$. On the other hand, $\left.\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}\right|_{e_i=e^{*0}} < 0$ if

$$\bar{v} < \frac{\delta(-((2+\alpha)(1+e^{*0^2})) + (2+\alpha)e^{*0^2}\rho - e^{*0^2}\rho^2)}{-2+\alpha(-1+e^{*0^2}(-1+\rho))} + \sqrt{\frac{\delta^2 e^{*0^2}(8\rho+4\alpha\rho+\alpha e^{*0^2}(4+\rho(-4+(-2+\rho)^2\rho)))}{\alpha(2+\alpha-\alpha e^{*0^2}(-1+\rho))^2}},$$

which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}(e_i)$ by setting $e_i < e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} < e^{*0}$. Thus, the comparison between $e^{*\mathcal{B}}$ and e^{*0} can be characterized by threshold $\tilde{v}''''$ such that $e^{*\mathcal{B}} < e^{*0}$ if $\bar{v} < \tilde{v}''''$; whereas $e^{*\mathcal{B}} > e^{*0}$ if $\bar{v} > \tilde{v}''''$, where

$$\tilde{v}'''' \equiv \frac{\delta(-((2+\alpha)(1+e^{*0^2})) + (2+\alpha)e^{*0^2}\rho - e^{*0^2}\rho^2)}{-2+\alpha(-1+e^{*0^2}(-1+\rho))} + \sqrt{\frac{\delta^2 e^{*0^2}(8\rho+4\alpha\rho+\alpha e^{*0^2}(4+\rho(-4+(-2+\rho)^2\rho)))}{\alpha(2+\alpha-\alpha e^{*0^2}(-1+\rho))^2}}.$$

(iii) Next, we compare the equilibrium sustainability effort levels between the on-consortium game with competition, $e^{*\mathcal{B}}$, and on-consortium game without competition, $e^{*\mathcal{BN}}$. Since $e^{*\mathcal{B}}$ and $e^{*\mathcal{BN}}$ are both strictly positive by Proposition 9 and 4, to compare $e^{*\mathcal{B}}$ and $e^{*\mathcal{BN}}$, we need to know how $\Pi^{\mathcal{B}}(e_i)$ changes at $e_i = e^{*\mathcal{BN}}$. Similar to (ii), by plugging in $e^{*\mathcal{BN}}$, as characterized in (3.13), into $\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}$, we have $\left.\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}\right|_{e_i=e^{*\mathcal{BN}}} > 0$ if

$$\bar{v} > \delta + 2\sqrt{\frac{(1+\alpha)^2\delta^2 e^{*\mathcal{BN}^2}(4-\rho)\rho}{(-4(1+\alpha)^2+\rho+\alpha(2+\alpha+\alpha e^{*\mathcal{BN}^2})\rho)^2}} - \frac{2\alpha\delta e^{*\mathcal{BN}^2}\rho}{-4(1+\alpha)^2+\rho+\alpha(2+\alpha+\alpha e^{*\mathcal{BN}^2})\rho},$$

which means firm i can increase its on-consortium (with competition) profit $\Pi^{\mathcal{B}}(e_i)$ by setting $e_i > e^{*\mathcal{BN}}$. Hence, we obtain that $e^{*\mathcal{B}} > e^{*\mathcal{BN}}$ if $\bar{v}$ is sufficiently high. On the other hand, $\left.\frac{d\Pi^{\mathcal{B}}(e_i)}{de_i}\right|_{e_i=e^{*\mathcal{BN}}} < 0$ if

$$\bar{v} < \delta + 2\sqrt{\frac{(1+\alpha)^2\delta^2 e^{*\mathcal{BN}^2}(4-\rho)\rho}{(-4(1+\alpha)^2+\rho+\alpha(2+\alpha+\alpha e^{*\mathcal{BN}^2})\rho)^2}} - \frac{2\alpha\delta e^{*\mathcal{BN}^2}\rho}{-4(1+\alpha)^2+\rho+\alpha(2+\alpha+\alpha e^{*\mathcal{BN}^2})\rho},$$

which means firm i can increase its on-consortium (with competition) profit $\Pi^{\mathcal{B}}(e_i)$ by setting $e_i < e^{*\mathcal{BN}}$. Hence, we obtain that $e^{*\mathcal{B}} < e^{*\mathcal{BN}}$ if $\bar{v}$ is sufficiently low.

□

Proof of Corollary 2. (i) Since the equilibrium sustainability effort e^* in the off-consortium game is uniquely determined by (B.16), we implicitly differentiate (B.16) with respect to ρ and obtain $\frac{de^*}{d\rho}$ as follows:

$$\frac{de^*}{d\rho} = \frac{4\delta^2 e^{*2} + (\delta - \bar{v})^2}{\delta e^{*2}(2-\rho)^3 t''(e^*)}.$$

Since $t''(e^*) > 0$, we obtain that the denominator of $\frac{de^*}{d\rho} > 0$. Thus, we obtain that $\frac{de^*}{d\rho} > 0$.

(ii) Since the equilibrium sustainability effort e^* in the on-consortium game is uniquely determined by (B.21), we implicitly differentiate (B.21) with respect to ρ and obtain $\frac{de^*}{d\rho}$ as follows:

$$\frac{de^*}{d\rho} = \frac{\delta^2(-2(1+\alpha)^2 + (-2+\alpha)^2 e^{*2}\rho) + 4(1+\alpha)^2\delta\bar{v}}{2\delta e^{*2}(2-\rho)^3 t''(e^*)} + \frac{-2(-2+\alpha)\alpha\delta e^{*2}\rho\bar{v} + (-2(1+\alpha)^2 + \alpha^2 e^{*2}\rho)\bar{v}^2}{2\delta e^{*2}(2-\rho)^3 t''(e^*)}.$$

Since $t''(e^*) > 0$, we obtain that the denominator of $\frac{de^*}{d\rho} > 0$. Thus, we can obtain that if $\bar{v} < \tilde{v}'''$, where

$$\tilde{v}''' \equiv \delta + 2\sqrt{2}\sqrt{\frac{(1+\alpha)^2\delta^2 e^{*2}\rho}{(-2(1+\alpha)^2 + \alpha^2 e^{*2}\rho)^2}} - \sqrt{\frac{2\alpha\delta e^{*2}\rho}{-2(1+\alpha)^2 + \alpha^2 e^{*2}\rho}},$$

then $\frac{de^*}{d\rho} < 0$. If $\bar{v} > \tilde{v}'''$, $\frac{de^*}{d\rho} > 0$.

□

Proposition 10. *(Comparison of Sustainability Effort). When the consortium members' verification is fully accurate ($\alpha = 1$), the sustainability effort levels may be higher or lower: $e^{\mathcal{B}*} < e^{0*}$ if $\bar{v} < \tilde{v}'''''$, and $e^{\mathcal{B}*} > e^{0*}$ if $\bar{v} > \tilde{v}'''''$.*

Proof of Proposition 10. In the on-consortium game, when $\alpha = 1$, firm i's problem in stage 1 becomes $\max_e \Pi^{\mathcal{B}} = ep_h D_h + (1-e)p_l D_l - t(e)$

In this case, the first and second-order conditions of $\Pi^{\mathcal{B}}$ with respect to e yield

$$\frac{d\Pi^{\mathcal{B}}}{de} = \frac{4(\delta-\bar{v})^2 + e^2(-3\delta+\bar{v})^2}{8\delta e^2} - t'(e),$$
$$\frac{d^2\Pi^{\mathcal{B}}}{de^2} = -\frac{(\delta-\bar{v})^2}{\delta e^3} - t''(e) < 0.$$

Thus, $\Pi^{\mathcal{B}}$ is concave in e. Hence, firm i's optimal sustainability effort on-consortium approaches a value uniquely determined by

$$t'(e) = \frac{4(\delta - \bar{v})^2 + e^2(-3\delta + \bar{v})^2}{8\delta e^2}. \tag{B.23}$$

Next, we compare the equilibrium sustainability effort levels between the off-consortium and on-consortium scenario when the consortium members' verification ability is $\alpha = 1$. Since $e^{*\mathcal{B}}$ and e^{*0} are both strictly positive, to compare $e^{*\mathcal{B}}$ and e^{*0}, we need to know how $\Pi^{\mathcal{B}}$ changes at $e = e^{*0}$. By plugging in e^{*0}, as characterized in (3.5), into $\frac{d\Pi^{\mathcal{B}}}{de}$, we have

$$\begin{aligned}
\left.\frac{d\Pi^{\mathcal{B}}}{de}\right|_{e=e^{*0}} &= \frac{4(\delta - \bar{v})^2 + e^{*0^2}(-3\delta + \bar{v})^2}{8\delta e^2} - t'(e^{*0}) \\
&= \frac{4(\delta - \bar{v})^2 + e^{*0^2}(-3\delta + \bar{v})^2}{8\delta e^2} - \frac{4\,\delta^2\,e^{*0^2} + (\bar{v} - \delta)^2}{8\,\delta\,e^{*0^2}},
\end{aligned}$$

Thus, $\left.\frac{d\Pi^{\mathcal{B}}(e,c)}{de}\right|_{e=e^{*0}} > 0$ if $\bar{v} > \frac{3\delta + 5\delta e^{*0^2}}{3 + e^{*0^2}}$, which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}$ by setting $e > e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} > e^{*0}$. On the other hand, $\left.\frac{d\Pi^{\mathcal{B}}}{de}\right|_{e=e^{*0}} < 0$ if $\bar{v} < \frac{3\delta + 5\delta e^{*0^2}}{3 + e^{*0^2}}$, which means firm i can increase its on-consortium profit $\Pi^{\mathcal{B}}$ by setting $e < e^{*0}$. Hence, we obtain that $e^{*\mathcal{B}} < e^{*0}$. Thus, the comparison between $e^{*\mathcal{B}}$ and e^{*0} can be characterized by threshold $\tilde{v}''''$ such that $e^{*\mathcal{B}} < e^{*0}$ if $\bar{v} < \tilde{v}''''$; whereas $e^{*\mathcal{B}} > e^{*0}$ if $\bar{v} > \tilde{v}''''$, where

$$\tilde{v}'''' \equiv \frac{3\delta + 5\delta e^{*0^2}}{3 + e^{*0^2}}.$$

$\square$

Appendix C

Appendix for Chapter 4

C.1 Supplementary Figures and Tables

Table C.1: Variable Definitions

Variable	Unit	Definition
Loan Indicators		
Loan approval	–	Dummy that takes one if the applied loan amount is fully disbursed by Kiva guarantors and raised by Kiva MFI otherwise zero
Loan length	Month	Number of months that the loan is disbursed to the borrower to the point when the last repayment is due
Guarantee amount	USD	The amount of the loan that must be funded by Kiva guarantors before the loan can be disbursed
Number of guarantors	–	Number of contributing guarantors to the loan
Per-guarantee amount	USD	The average amount contributed by Kiva guarantors to the loan
Individual Characteristics		
Year of business operated	Year	Years of business operated or working experiences
Number of income sources	–	Number of income streams
Repeat borrower	–	Dummy that takes one if the loan applicant is tagged as "repeat borrower" on Kiva otherwise zero
Gender	–	Dummy that takes one if the loan applicant is a male otherwise zero
Children	–	Number of children for Kiva loan applicant
Marital status	–	Dummy that takes one if the loan applicant is married otherwise zero
Age	Year	Age of Kiva loan applicant
Lending Portfolio		
Interest rate	–	MFI's lending portfolio yield divided by its outstanding portfolio loan

Table C.1 (continued)

Variable	Unit	Definition
Approval rate	–	The loan approval rate in MFI's lending portfolio
Loan ended	Thousand USD	Annual amount of loans no longer in the process of being paid back
Loan extended	Thousand USD	Annual amount of loans raised and disbursed by MFIs
Loan defaulted	Thousand USD	Annual amount of loan that the repaid amount is less than expected as of 360 days prior and no repayments reported to Kiva MFIs
Default rate	–	Annual amount of loans defaulted scaled by the amount of extended loans for Kiva MFIs
Guarantee ratio	–	The proportion of the loan amount funded by Kiva guarantors in MFI's lending portfolio
Loan application	–	Annual number of loan applications received by MFIs
Loan outstanding	–	Annual number of loans currently in the repayment process
Service revenue	Thousand USD	Lending service revenue reported by MFI annual financial statement
Operational expense	Thousand USD	Lending service expense reported by MFI annual financial statement
MFI Characteristics		
Average years of business	Year	Average year of business operated of Kiva borrowers in MFI's lending portfolio
Average income sources	–	Average number of income streams of Kiva borrowers in MFI's lending portfolio
Proportion of repeat borrower	–	The percentage of repeat borrowers in MFI's lending portfolio
Loans to women	–	The percentage of total loans for women borrowers
Time to fund loans	Day	Average length of time that it takes for MFI portfolio loans to be fully funded
Time on Kiva	Month	Number of months an MFI has been posting loans to Kiva for funding
Macroeconomic Factors		
FID	–	Financial Institution Depth index of each country derived from IMF website
FIA	–	Financial Institution Access index of each country derived from IMF website
FIE	–	Financial Institution Efficiency index of each country derived from IMF website
FMD	–	Financial Market Depth index of each country derived from IMF website
FMA	–	Financial Market Access index of each country derived from IMF website

Table C.1 (continued)

Variable	Unit	Definition
FME	–	Financial Market Efficiency index of each country derived from IMF website
Annual income	USD	The average annual income per person in each country
Poverty line	–	The percentage of the population living below the poverty line
Life expectancy	Year	The average length of survival for members of country's population
Literacy rate	–	The percentage of the country's population with the ability to read and write

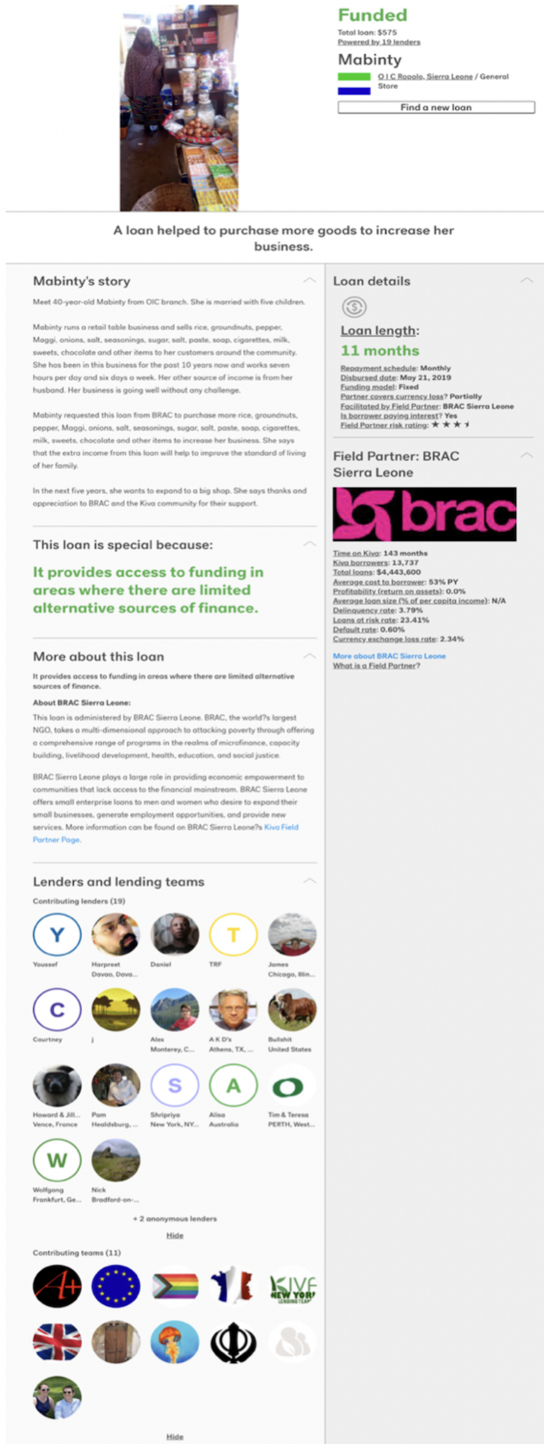


Figure C·1: An Example of a Kiva Borrower Profile

References

Abada, I. and Lambin, X. (2023). Artificial intelligence: Can seemingly collusive outcomes be avoided? *Management Science*, 69(9):4382–4399.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138(1):1–35.

Abreu, D., Pearce, D., and Stacchetti, E. (1986). Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–269.

Abreu, D., Pearce, D., and Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5):1041–1063.

Agrawal, V. and Lee, D. (2019). The Effect of Sourcing Policies on Suppliers' Sustainable Practices. *Production and Operations Management*, 28(4):767–787.

Allon, G., Drakopoulos, K., and Manshadi, V. (2019). Information inundation on platforms and implications. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 555–556.

Anthropic (2025). Project vend: Can Claude run a small shop? (and why does that matter?). https://www.anthropic.com/research/project-vend-1. Accessed: 2026-01-07.

Antitrust Division, Department of Justice (2026). United states v. realpage, inc.; proposed final judgment and competitive impact statement. *Federal Register*, 91:5628–5642.

Arora, N. and Henderson, T. (2007). Embedded premium promotion: Why it works and how to make it more effective. *Marketing Science*, 26(4):514–531.

Asker, J., Fershtman, C., and Pakes, A. (2022). Artificial intelligence, algorithm design, and pricing. *AEA Papers and Proceedings*, 112:452–456.

Assad, S., Clark, R., Ershov, D., and Xu, L. (2024). Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market. *Journal of Political Economy*, 132(3):723–771.

Athey, S. and Bagwell, K. (2001). Optimal collusion with private information. *RAND Journal of Economics*, 32(3):428–465.

Athey, S. and Bagwell, K. (2008). Collusion with persistent cost shocks. *Econometrica*, 76(3):493–540.

Aura Blockchain Consortium (2021). Aura blockchain consortium. United Nations Department of Economic and Social Affairs, Partnerships for SDGs. Accessed: 2025-01-24.

Babich, V. and Hilary, G. (2020). Distributed ledgers and operations: What operations management researchers should know about blockchain technology. *Manufacturing & Service Operations Management*, 22(2):223–240.

Baksi, S. and Bose, P. (2007). Credence goods, efficient labelling policies, and regulatory enforcement. *Environmental & Resource Economics*, 37(2):411–430.

Banchio, M. and Mantegazza, G. (2022). Artificial intelligence and spontaneous collusion. *arXiv preprint arXiv:2202.05946*.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics*, 143(1):30–56.

Bastani, H., Bastani, O., and Sinchaisri, W. P. (2026). Improving human sequential decision making with reinforcement learning. *Management Science*, 72(1):733–755.

Begley, T. A. and Purnanandam, A. (2021). Color and credit: Race, regulation, and the quality of financial services. *Journal of Financial Economics*, 141(1):48–65.

Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *Review of Financial Studies*, 33(7):2845–2897.

Bimpikis, K., Morgenstern, I., and Saban, D. (2024). Data tracking under competition. *Operations Research*, 72(2):514–532.

Borri, N., Liu, Y., and Tsyvinski, A. (2022). The economics of non-fungible tokens. *Available at SSRN*, (4052045).

Brero, G., Mibuari, E., Lepore, N., and Parkes, D. C. (2022). Learning to mitigate ai collusion on economic platforms. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pages 37892–37904. Curran Associates, Inc.

Bricken, T., Templeton, A., Batson, J., Chen, B., Jermyn, A., Conerly, T., Turner, N., Anil, C., Denison, C., Askell, A., Lasenby, R., Wu, Y., Kravec, S., Schiefer, N., Maxwell, T., Joseph, N., Tamkin, A., Nguyen, K., McLean, B., Burke, J. E., Hume, T., Carter, S., Henighan, T., and Olah, C. (2023). Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*. https://transformer-circuits.pub/2023/monosemantic-features/.

Brock, W. A. and Scheinkman, J. A. (1985). Price setting supergames with capacity constraints. *The Review of Economic Studies*, 52(3):371–382.

Brown, M., Guin, B., and Kirschenmann, K. (2016). Microfinance banks and financial inclusion. *Review of Finance*, 20(3):907–946.

Brown, S., Goetzmann, W., Liang, B., and Schwarz, C. (2012). Trust and delegation. *Journal of Financial Economics*, 103(2):221–234.

Brown, Z. and MacKay, A. (2025). Algorithmic coercion with faster pricing. *Available at SSRN 4380895*.

Brown, Z. Y. and MacKay, A. (2023). Competition in pricing algorithms. *American Economic Journal: Microeconomics*, 15(2):109–156.

Bu, D. and Liao, Y. (2022). Shaming microloan delinquents: Evidence from a field experiment in China. *Management Science*, 68(5):3768–3790.

Buell, R. W. and Kalkanci, B. (2021). How Transparency into Internal and External Responsibility Initiatives Influences Consumer Choice. *Management Science*, 67(2):932–950.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2021). The macroeconomics of microfinance. *Review of Financial Studies*, 88(1):126–161.

Burtch, G., Ghose, A., and Wattal, S. (2014). Cultural differences and geography as determinants of online prosocial lending. *MIS Quarterly*, 38(3):773–794.

Calvano, E., Calzolari, G., Denicolò, V., and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Campbell, J. Y. and Cocco, J. F. (2015). A model of mortgage default. *Journal of Finance*, 70(4):1495–1554.

Canales, R. and Greenberg, J. (2016). A matter of (relational) style: Loan officer consistency and exchange continuity in microfinance. *Management Science*, 62(4):1202–1224.

Castka, P. and Corbett, C. J. (2016). Governance of eco-labels: Expert opinion and media coverage. *Journal of Business Ethics*, 135(2):309–326.

CDP (2019). Top fmcgs in race to keep up with conscious consumers. https://www.prnewswire.com/news-releases/cdp-research-top-fmcgs-in-race-to-keep-up-with-conscious-consumers-300800028.html.

CGTN (2018). Manager of Alibaba's Hema Fresh retailer sacked over expiry date label scandal. November 18, 2018. https://news.cgtn.com/news/3d3d674e77677a4e30457a6333566d54/share_p.html.

Chen, J., Qi, A., and Dawande, M. (2020). Supplier Centrality and Auditing Priority in Socially Responsible Supply Chains. *Manufacturing & Service Operations Management*, 22(6):1199–1214.

Chen, S., Doerr, S., Frost, J., Gambacorta, L., and Shin, H. S. (2023). The fintech gender gap. *Journal of Financial Intermediation*, 54:101026.

Chen, T., Huang, Y., Lin, C., and Sheng, Z. (2022). Finance and firm volatility: Evidence from small business lending in China. *Management Science*, 68(3):2226–2249.

Chod, J., Trichakis, N., Tsoukalas, G., Aspegren, H., and Weber, M. (2020). On the financing benefits of supply chain transparency and blockchain adoption. *Management Science*, 66(10):4378–4396.

Chung, S., Kim, K., Lee, C. H., and Oh, W. (2023). Interdependence between online peer-to-peer lending and cryptocurrency markets and its effects on financial inclusion. *Production and Operations Management*, 32(6):1939–1957.

Cole, S., Kanz, M., and Klapper, L. (2015). Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers. *Journal of Finance*, 70(2):537–575.

Compte, O., Jenny, F., and Rey, P. (2002). Capacity constraints, mergers and collusion. *European Economic Review*, 46(1):1–29.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., Tang, K., Wang, Y., and Zhao, X. (2023). Inclusion and democratization through Web3 and DeFi? Initial evidence from the Ethereum ecosystem. *National Bureau of Economic Research*, (w30949).

Conley, T. G. and Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *Review of Economics and Statistics*, 93(1):113–125.

Corbett, C. J. and DeCroix, G. A. (2001). Shared-Savings Contracts for Indirect Materials in Supply Chains: Channel Profits and Environmental Impacts. *Management Science*, 47(7):881–893.

Cui, Y., Gaur, V., and Liu, J. (2024). Supply chain transparency and blockchain design. *Management Science*, 70(5):3245–3263.

Cui, Y., Hu, M., and Liu, J. (2023). Value and design of traceability-driven blockchains. *Manufacturing & Service Operations Management*, 25(3):1066–1084.

Cull, R., Demirgüç-Kunt, A., and Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *Economic Journal*, 117(517):F107–F133.

Davydiuk, T., Gupta, D., and Rosen, S. (2023). De-crypto-ing signals in initial coin offerings: Evidence of rational token retention. *Management Science*, 69(11):6584–6624.

Dell'Ariccia, G. and Marquez, R. (2006). Lending booms and lending standards. *Journal of Finance*, 61(5):2511–2546.

Delmas, M. A. and Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1):64–87.

Demirgüç-Kunt, A., Klapper, L., and Singer, D. (2017). Financial inclusion and inclusive growth: A review of recent empirical evidence. Technical report, World Bank. Policy Research Working Paper 8040. https://doi.org/10.1596/1813-9450-8040.

den Boer, A. V., Meylahn, J. M., and Schinkel, M. P. (2024). Artificial collusion: Examining supracompetitive pricing by q-learning algorithms. *Amsterdam Law School Research Paper No. 2022-25*. Amsterdam Center for Law & Economics Working Paper No. 2022-06. https://ssrn.com/abstract=4213600.

Deng, A. (2024). What do we know about algorithmic collusion now? new insights from the latest academic research. *SSRN Electronic Journal*. https://ssrn.com/abstract=4521959.

Di Maggio, M. and Yao, V. (2021). FinTech borrowers: Lax screening or cream-skimming? *Review of Financial Studies*, 34(10):4565–4618.

Diederich, J. and Goeschl, T. (2014). Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence. *Environmental and Resource Economics*, 57(3):405–429.

Dou, W. W., Goldstein, I., and Ji, Y. (2025). Ai-powered trading, algorithmic collusion, and price efficiency. *NBER Working Paper No. 34054*. https://ssrn.com/abstract=4452704.

Dutta, P. K., Matros, A., and Weibull, J. W. (2007). Long-run price competition. *RAND Journal of Economics*, 38(2):291–313.

Erel, I. and Liebersohn, J. (2022). Can FinTech reduce disparities in access to finance? Evidence from the Paycheck Protection Program. *Journal of Financial Economics*, 146(1):90–118.

Farhi, E., Lerner, J., and Tirole, J. (2013). Fear of rejection? Tiered certification and transparency. *RAND Journal of Economics*, 44(4):610–631.

Fish, S., Gonczarowski, Y. A., and Shorrer, R. I. (2025). Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806*.

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

Fudenberg, D. and Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *Review of Financial Studies*, 32(5):1854–1899.

Gal, M. S. and Rubinfeld, D. L. (2024). Algorithms, ai, and mergers. *Antitrust Law Journal*, 85(3):684–738.

Galak, J., Small, D., and Stephen, A. (2011). Micro-finance decision making: A field study of prosocial lending. *Journal of Marketing Research*, 48:S130–S137.

Gans, J. (2023). Price collusion using artificial intelligence. In *Antitrust Economics for Lawyers*. LexisNexis.

Gao, F. and Souza, G. C. (2022). Carbon offsetting with eco-conscious consumers. *Management Science*, 68(11):7879–7897.

Garmaise, M. J. and Natividad, G. (2013). Cheap credit, lending operations, and international politics: The case of global microfinance. *Journal of Finance*, 68(4):1551–1576.

Gaur, V. and Gaiha, A. (2020). Building a transparent supply chain: Blockchain can enhance trust, efficiency, and speed. *Harvard Business Review*, 98(3):80–88.

Gemini Team (2025). Gemini: A family of highly capable multimodal models. https://arxiv.org/abs/2312.11805.

Genesove, D. and Mullin, W. P. (2001). Rules, communication, and collusion: Narrative evidence from the sugar institute case. *American Economic Review*, 91(3):379–398.

Golob, U. and Kronegger, L. (2019). Environmental consciousness of european consumers: A segmentation-based study. *Journal of Cleaner Production*, 221:1–9.

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

GreenPrint (2022). Greenprint business of sustainability index 2022. https://pditechnologies.com/news/environmental-sustainability-index/.

Griffin, J. M. and Maturana, G. (2016). Did dubious mortgage origination practices distort house prices? *Review of Financial Studies*, 29(7):1671–1708.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Guo, R., Lee, H. L., and Swinney, R. (2016). Responsible sourcing in supply chains. *Management Science*, 62(9):2722–2744.

Gustafson, M. T., Ivanov, I. T., and Meisenzahl, R. R. (2021). Bank monitoring: Evidence from syndicated loans. *Journal of Financial Economics*, 139(2):452–477.

Hansen, K. T., Misra, K., and Pai, M. M. (2021). Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science*, 40(1):1–12.

Harbaugh, R., Maxwell, J. W., and Roussillon, B. (2011). Label confusion: The Groucho effect of uncertain standards. *Management Science*, 57(9):1512–1527.

Harrington, J. E. and Skrzypacz, A. (2011). Private monitoring and communication in cartels: Explaining recent collusive practices. *American Economic Review*, 101(6):2425–2449.

Harvey, C. R. and Rabetti, D. (2024). International business and decentralized finance. *Journal of International Business Studies*, 55(7):840–863.

Helliar, C. V., Crawford, L., Rocca, L., Teodori, C., and Veneziani, M. (2020). Permissionless and permissioned blockchain diffusion. *International Journal of Information Management*, 54:102136.

Houde, S. (2022). Bunching with the stars: How firms respond to environmental certification. *Management Science*, 68(8):5569–5590.

Howell, S. T., Niessner, M., and Yermack, D. (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. *Review of Financial Studies*, 33(9):3925–3974.

Hu, X., Lin, Y., and Liu, X. (2023). The role of information quality in corporate loan securitization evidence from collateralized loan obligations. Technical report. Available at SSRN 4401595.

Huang, K. and Zhu, F. (2023). Chatgpt and gig economy. In Huang, K., Wang, Y., Zhu, F., Chen, X., and Xing, C., editors, *Beyond AI: ChatGPT, Web3, and the Business Landscape of Tomorrow*, Future of Business and Finance, pages 129–158. Springer, Cham.

Huang, M.-H. and Rust, R. T. (2011). Sustainability and consumption. *Journal of the Academy of Marketing Science*, 39:40–54.

Huang, X. N., Atasu, A., and Toktay, L. B. (2019). Design implications of extended producer responsibility for durable products. *Management Science*, 65(6):2573–2590.

Huck, S., Normann, H.-T., and Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4):435–446.

Ivashina, V. (2009). Asymmetric information effects on loan spreads. *Journal of Financial Economics*, 92(2):300–319.

Iyengar, G., Saleh, F., Sethuraman, J., and Wang, W. (2023). Economics of permissioned blockchain adoption. *Management Science*, 69(6):3415–3436.

Jin, S., Wang, W., and Zhang, Z. (2023). The real effects of implicit government guarantee: Evidence from Chinese state-owned enterprise defaults. *Management Science*, 69(6):3650–3674.

John, K., Kogan, L., and Saleh, F. (2022). Smart contracts and decentralized finance. *Annual Review of Financial Economics*, 15(1):523–542.

Johnson, J. P., Rhodes, A., and Wildenbeest, M. (2023). Platform design when sellers use pricing algorithms. *Econometrica*, 91(5):1841–1879.

Kalkanci, B. and Plambeck, E. L. (2020a). Managing supplier social and environmental impacts with voluntary versus mandatory disclosure to investors. *Management Science*, 66(8):3311–3328.

Kalkanci, B. and Plambeck, E. L. (2020b). Reveal the supplier list? a trade-off in capacity vs. responsibility. *Manufacturing & Service Operations Management*, 22(6):1251–1267.

Karaer, O., Kraft, T., and Khawam, J. (2017). Buyer and nonprofit levers to improve supplier environmental performance. *Production and Operations Management*, 26(6):1163–1190.

Kavanagh, J. J. and Magloughlin, A. (2025). DOJ doubles down on algorithmic collusion. StepAhead: Antitrust & Competition Insights, Steptoe & Johnson LLP. https://www.steptoe.com/en/news-publications/stepahead-antitrust-and-competition-insights/doj-doubles-down-on-algorithmic-collusion.html.

Keys, B. J., Pope, D. G., and Pope, J. C. (2016). Failure to refinance. *Journal of Financial Economics*, 122(3):482–499.

Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, 52(3):538–558.

Kraft, T., Valdés, L., and Zheng, Y. (2018). Supply chain visibility and social responsibility: Investigating consumers' behaviors and motives. *Manufacturing & Service Operations Management*, 20(4):617–636.

Laffont, J. and Martimort, D. (1997). Collusion under asymmetric information. *Econometrica*, 65(4):875–912.

Lanz, B., Wurlod, J.-D., Panzone, L., and Swanson, T. (2018). The behavioral effect of pigovian regulation: Evidence from a field experiment. *Journal of Environmental Economics and Management*, 87:190–205.

Lee, Y. S., Soroushian, J., Bullock, J., Carroll, M., Dokko, J., Dominski, J., Hinds, R., Holzer, H., Horrigan, M., Munyikwa, Z., Oschinski, M., Radsch, C., Rock, D., Rossi, M., Seamans, R., Towns, A., and Zhang, B. (2025). Proactively developing and assisting the workforce in the age of AI. White paper, Americans for Responsible Innovation; Keough School of Global Affairs, University of Notre Dame. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5366876.

Leonidou, C. N. and Skarmeas, D. (2017). Gray Shades of Green: Causes and Consequences of Green Skepticism. *Journal of Business Ethics*, 144(2):401–415.

Lin, W., Ortega, D. L., Ufer, D., Caputo, V., and Awokuse, T. (2022). Blockchain-based traceability and demand for u.s. beef in china. *Applied Economic Perspectives and Policy*, 44(1):253–272.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *RAND Journal of Economics*, 30(2):214–231.

Luo, X., Ge, L., and Wang, C. A. (2022). Crowdfunding for microfinance institutions: The new hope? *MIS Quarterly*, 46(1).

Lyandres, E., Palazzo, B., and Rabetti, D. (2022). Initial coin offering (ICO) success and post-ICO performance. *Management Science*, 68(12):8658–8679.

Lyon, T. P. and Maxwell, J. W. (2011). Greenwash: Corporate environmental disclosure under threat of audit. *Journal of Economics & Management Strategy*, 20(1):3–41.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

MacKinnon, J. G. and Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *Econometrics Journal*, 21(2):114–135.

Marquis, C., Toffel, M. W., and Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2):483–504.

Maxwell, J. W., Lyon, T. P., and Hackett, S. C. (2000). Self-regulation and social welfare: The political economy of corporate environmentalism. *Journal of Law and Economics*, 43(2):583–618.

McCarthy, K. (2025). How Delta is using AI for ticket pricing and what it means for air travel. ABC News. August 7, 2025. https://abcnews.com/GMA/Travel/delta-ai-ticket-pricing-means-air-travel/story?id=124343088.

McKinsey & Company (2020). The state of fashion 2020. https://www.mckinsey.com/~/media/mckinsey/industries/retail/our%20insights/the%20state%20of%20fashion%202020%20navigating%20uncertainty/the-state-of-fashion-2020-final.pdf.

Mian, A. and Sufi, A. (2017). Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. *Review of Financial Studies*, 30(6):1832–1864.

Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *Accounting Review*, 87(4):1385–1413.

Miller, A. P., Hosanagar, K., and Ramayya, K. (2024). Harnessing ai for business insight: Key considerations for deploying llms in summarization pipelines. The Wharton School Research Paper. https://ssrn.com/abstract=4935842.

Miller, N. H. and Weinberg, M. C. (2017). Understanding the price effects of the millercoors joint venture. *Econometrica*, 85(6):1763–1791.

Morone, P., Caferra, R., D'Adamo, I., Falcone, P. M., Imbert, E., and Morone, A. (2021). Consumer willingness to pay for bio-based products: Do certifications matter? *International Journal of Production Economics*, 240:108248.

Murali, K., Lim, M. K., and Petruzzi, N. C. (2019). The Effects of Ecolabels and Environmental Regulation on Green Product Development. *Manufacturing & Service Operations Management*, 21(3):519–535.

Musolff, L. (2025). Algorithmic pricing, price wars and tacit collusion: Evidence from e-commerce. *Working paper, Wharton School, University of Pennsylvania.* https://lmusolff.com/papers/Algorithmic_Pricing.pdf.

Narad, R., Boussioux, L., and Wagner, M. (2025). Mechanistic interpretability for neural tsp solvers. In *NeurIPS 2025 Workshop MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making*.

New York Times (2021). Lvmh, richemont and prada join forces on blockchain. April 20, 2021. https://www.nytimes.com/2021/04/20/business/lvmh-richemont-prada-blockchain.html.

NielsenIQ (2023). Consumers care about sustainability—and back it up with their wallets. https://nielseniq.com/global/en/insights/report/2023/consumers-care-about-sustainability-and-back-it-up-with-their-wallets/.

Nygaard, A. and Silkoset, R. (2023). Sustainable development and greenwashing: How blockchain technology information can empower green consumers. *Business Strategy and the Environment*, 32(6):3801–3813.

Olsen, T. L. and Tomlin, B. (2020). Industry 4.0: Opportunities and challenges for operations management. *Manufacturing & Service Operations Management*, 22(1):113–122.

O'Malley, T. (2021). The impact of repossession risk on mortgage default. *Journal of Finance*, 76(2):623–650.

OpenAI (2023). GPT-4 technical report.

Orsdemir, A., Hu, B., and Deshpande, V. (2019). Ensuring corporate social and environmental responsibility through vertical integration and horizontal sourcing. *Manufacturing & Service Operations Management*, 21(2):417–434.

Philippon, T. (2015). Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *American Economic Review*, 105(4):1408–1438.

Philippon, T. and Reshef, A. (2012). Wages and human capital in the US finance industry: 1909–2006. *Quarterly Journal of Economics*, 127(4):1551–1609.

Pigors, M. and Rockenbach, B. (2016). Consumer social responsibility. *Management Science*, 62(11):3123–3137.

Pinto, B., Sokol, D. D., and Zhu, F. (2024). The antitrust and privacy interface: Lessons for regulators from the data. *George Mason Law Review*, 31(4):1019–1041. Also available as USC CLASS Research Paper No. 24-27 (SSRN 4908422).

Plambeck, E. L. and Taylor, T. A. (2016). Supplier evasion of a buyer's audit: Implications for motivating supplier social and environmental responsibility. *Manufacturing & Service Operations Management*, 18(2):184–197.

Plambeck, E. L. and Taylor, T. A. (2019). Testing by Competitors in Enforcement of Product Standards. *Management Science*, 65(4):1735–1751.

Pun, H., Swaminathan, J. M., and Hou, P. (2021). Blockchain adoption for combating deceptive counterfeits. *Production and Operations Management*, 30(4):864–882.

Puri, M. and Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics*, 86(1):71–99.

Qiu, L., Huang, Y., Singh, P. V., and Srinivasan, K. (2023). Personalization, consumer search and algorithmic pricing. *SSRN Electronic Journal*.

Qwen Team (2024). Qwen2.5: A party of foundation models.

Shetty, P., Upadhayaya, A., Shah, P. M., Jagabathula, S., Nayak, S., and Fee, A. J. (2025). Advanced financial reasoning at scale: A comprehensive evaluation of large language models on cfa level iii. *arXiv preprint arXiv:2507.02954*. Accepted at FinLLM @ IJCAI 2025.

Simon-Kucher & Partners (2021). The global sustainability study 2021. https://www.simon-kucher.com/sites/default/files/studies/Simon-Kucher_Global_Sustainability_Study_2021.pdf.

Stigler, G. J. (1964). A theory of oligopoly. *Journal of Political Economy*, 72(1):44–61.

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *Journal of Finance*, 62(2):629–668.

Sumkin, D., Hasija, S., and Netessine, S. (2021). Does blockchain facilitate responsible sourcing? an application to the diamond supply chain. Insead working paper. https://ssrn.com/abstract=3802294.

Svirydzenka, K. (2016). Introducing a new broad-based index of financial development. Technical report, International Monetary Fund. https://www.imf.org/external/pubs/ft/wp/2016/wp1605.pdf.

Tambe, P. B. (2026). Reskilling the workforce for AI: Domain expertise and algorithmic literacy. *Management Science*, 72(1):515–537.

Wang, H. and Overby, E. (2023). Do political differences inhibit market transactions? An investigation in the context of online lending. *Management Science*, 69(8):4685–4706.

Wang, Q., Huang, Y., Singh, P. V., and Srinivasan, K. (2023). Algorithms, artificial intelligence and simple rule-based pricing. *SSRN Electronic Journal*.

Wei, Z. and Lin, M. (2017). Market mechanisms in online peer-to-peer lending. *Management Science*, 63(12):4236–4257.

Whitaker, A. and Kräussl, R. (2020). Fractional equity, blockchain, and the future of creative work. *Management Science*, 66(10):4594–4611.

Yang, S. A. and Zhang, A. H. (2024). Generative ai and copyright: A dynamic perspective. Written February 4 2024; Posted February 28 2024; Revised March 27 2024. https://arxiv.org/abs/2402.17801.

Zhao, E. Y. and Wry, T. (2016). Not all inequality is equal: Deconstructing the societal logic of patriarchy to understand microfinance lending to women. *Academy of Management Journal*, 59(6):1994–2020.

Zhao, H. and Berman, R. (2025). Algorithmic collusion of pricing and advertising on e-commerce platforms. *arXiv preprint arXiv:2508.08325*.

CURRICULUM VITAE

[redacted]

[redacted]