Eva Lestant

September 2026

Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072

elestant@stanford.edu
elestant.github.io
(650) 709-6756

EDUCATION

Ph.D. in Economics, Stanford University, 2021–present.
Expected Completion: June 2027
DISSERTATION: "*Essays on Marketplace, Entrepreneurship, and Youth Employment in West Africa*"

M.Sc. in Economics, ENSAE / École Polytechnique (France), 2017–2019.

B.A. in Economics and Business, Université Catholique de Lille (France), 2014–2017.

DISSERTATION COMMITTEE

Prof. Melanie Morten
Department of Economics, Stanford University
melanie.morten@stanford.edu

Prof. Arun Chandrasekhar
Department of Economics, Stanford University
arungc@stanford.edu

Prof. Matthew Gentzkow
Department of Economics, Stanford University
gentzkow@stanford.edu

Prof. Alessandra Voena
Department of Economics, Stanford University
avoena@stanford.edu

RESEARCH AND TEACHING FIELDS

Primary field: Development Economics.
Secondary field: Industrial Organization.

PUBLICATIONS

DUPAS, PASCALINE, MARCEL FAFCHAMPS AND EVA LESTANT [2023], "Panel Data Evidence on the Effects of the COVID-19 Pandemic on Livelihoods in Urban Côte d'Ivoire," *PLoS ONE*, **18**(2), e0277559.

RESEARCH PAPERS

Sourcing (Together): Fostering Economies of Scale through Retail Cooperatives (**Job Market Paper**)

From Training to Earning: The 7-Year Impact of Dual Apprenticeships on Youth Employment (with Bruno Crépon and Patrick Premand). World Bank Policy Research Working Paper.

Estimating Variances for Causal Panel Data Estimators (with Alexander Almeida, Susan Athey, Guido

Imbens and Alexia Olaizola). arXiv:2510.11841.

RESEARCH IN PROGRESS

Consumer Search in a Centralized Market: Informational Friction and Market Segmentation.

Demand and Supply Determinants of a Formal Apprenticeship Program: Scale-up Analysis (with B. Crépon and P. Premand).

SCHOLARSHIPS, HONORS AND AWARDS

2026	Graduate Fellowship, King Center on Global Development, Stanford University.
2026	Graduate Fellowship, Stanford Institute for Economic Policy Research (SIEPR).
2026	Graduate Fellowship, Stanford Impact Labs.
2024	Community Impact Award, Stanford Alumni Association.
2021–2024	Albion Walter Hewlett Fellowship, Stanford University.
2019	Best Master's Thesis Award, Master in Economics, ENSAE.
2018	Best Research Internship, Department of Economics, École Polytechnique.

RESEARCH GRANTS

2026	Private Enterprise Development in Low-Income Countries (PEDL)	£29,975.
2026	Weiss Fund – Implementation & Policy Grants	$35,225.
2025	George P. Shultz Dissertation Fund	$9,000.
2025	King Center on Global Development	$10,000.
2025	Researcher Diversification Strategy, IPA West Africa	$10,000.
2024	Structural Transformation and Economic Growth (STEG)	$14,000.
2023	King Center on Global Development	$8,000.
2023	Weiss Fund	$11,000.
2020	Weiss Fund for Research in Development Economics	$30,000.

TEACHING EXPERIENCE

2025, 2026	Economic Development and Social Networks with Prof. A. Chandrasekhar, Stanford University.
2026	Junior Economics Honors Course with Prof. M. Clerici-Arias, Stanford University.

PROFESSIONAL ACTIVITIES

Organizer, Stanford Development Seminar and DevPec, 2024.
Mentor, King Center Predoctoral Fellows, 2023 and 2024.
President, WERISE (Women and Underrepresented Groups in Economics), Stanford University.

RELEVANT POSITIONS

2022	Research Assistant for Prof. M. Morten and Prof. B. Magaloni, Stanford University.
2019–2021	Research Manager for Prof. P. Dupas and Prof. M. Fafchamps, Stanford Institute for Economic Policy Research (SIEPR).
2018	Research Assistant for Prof. B. Crépon and P. Premand, Centre de Recherche en Économie et Statistique (CREST) / World Bank, Paris.

PHD COURSEWORK

Microeconomic Theory and Networks

- ECON 284 — Simplicity and Complexity in Economic Theory (Akbarpour)
- ECON 291 — Social and Economic Networks (Jackson)

Development Economics

- ECON 215 — Development Economics II (Morten)

Industrial Organization

- ECON 257 — Industrial Organization I (Cuesta, Einav, Somaini)
- ECON 260 — Industrial Organization III (Einav, Vasserman)

Market Design

- ECON 285 — Matching and Market Design (Roth, Ostrovsky)
- ECON 287 — Topics in Market Design (Ashlagi)

Machine Learning / Computer Science

- CS 229 — Machine Learning (Ng, Guestrin, Charikar)
- CS 234 — Reinforcement Learning (Brunskill)

OTHER

Languages: French (native), English (fluent), Spanish (intermediate).
Software: Python, R, Stata, LaTeX, QGIS, SurveyCTO.
Citizenship: France.

Last updated: September 2026

Eva Lestant
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
+1 (650) 709-6756
elestant@stanford.edu
https://elestant.github.io

Recruiting Committee
NBER Workshop on Development Economics and Networks for Early-Career Researchers

September 15, 2026

Dear Organizers,

I am writing to apply for the **NBER Workshop on Development Economics and Networks for Early-Career Researchers**. I am a development economist working at the intersection of development economics and industrial organization, with research focused on marketplaces, entrepreneurship, and how economic relationships shape firm growth in developing countries. I expect to graduate with a Ph.D. in Economics from Stanford University in June 2027.

This application package contains the requested materials: this cover letter, my curriculum vitae, my job market paper, *Sourcing (Together): Fostering Economies of Scale through Retail Cooperatives*, and a brief statement describing my motivation for attending the workshop and my current research interests.

If I am shortlisted for participation, the following faculty member could provide a recommendation:

Arun Chandrasekhar
Professor of Economics, Stanford University
arungc@stanford.edu

Thank you for considering my application. I would be very grateful for the opportunity to participate in the workshop.

Sincerely,

Eva Lestant

My research lies at the intersection of development economics and industrial organization. I study how the organization of markets and economic relationships shapes the productivity and growth of small informal firms, with a particular focus on entrepreneurship and employment in West Africa. More broadly, I am interested in why potentially valuable forms of cooperation fail to emerge, how economic relationships can overcome market frictions, and how changes in these relationships affect firms, consumers, and market structure.

My job market paper, *Sourcing (Together): Fostering Economies of Scale through Retail Cooperatives*, studies these questions among informal retailers. In the markets I study, retailers operate close to each other and source similar products from the same wholesale market, yet they have thin networks and almost never purchase jointly. I implement a field experiment that separately reduces informational, coordination, and enforcement barriers to forming purchasing groups. The experiment creates an exogenous change in group formation, allowing me to study what could help valuable relationships to form and what happens once they do.

I find that a facilitated meeting generates a large increase in joint sourcing, with most relationships persisting for at least a year and additional adoption beyond the treated group occurring after the intervention ends. These formed groups enable members to source at a lower cost and save on transportation. Retailers also gain access to products sourced by their peers, and cooperation ultimately lowers selling prices and increases profits. Finally, I use the experimental evidence with a model of product entry and quantity competition to quantify how group-mediated access and economies of scale affect market outcomes.

Working on this project has made me particularly interested in how economic networks can substitute for missing market infrastructure. In a follow-up project on fragmented supply chains in sub-Saharan Africa, I show that low producer density can limit economies of scale and, in some settings, make vertical integration preferable to competing with incumbent traders. More broadly, I am interested in how network structure and market organization jointly shape entrepreneurship and firm growth in developing economies.

I would be really excited to attend the NBER Workshop on Development Economics and Networks to develop the network dimension of this research agenda with researchers at the frontier of empirical networks. I would value feedback on how to conceptualize and measure endogenous group formation and diffusion, and how to connect these mechanisms to market-level outcomes. More broadly, I hope the workshop helps me identify network questions and empirical tools for the next stage of my research. I also look forward to learning from other early-career researchers approaching network research questions.

Sourcing (Together): Fostering Economies of Scale through Retail Cooperatives

Eva Lestant *

August 2026

[PRELIMINARY WORK, PLEASE DO NOT DISTRIBUTE!]
Latest Version

Abstract

Programs designed to relax within-firm constraints have had modest effects among single-owner firms. We test an alternative approach that reduces cooperation frictions in a setting where joint sourcing opportunities remain unexploited. In a randomized experiment among clustered fresh-food retailers in urban Côte d'Ivoire, we randomly reduce the cost of cooperation. Our experimental evidence demonstrates that correcting pessimistic second-order beliefs about peers increases participation by 14 p.p. Facilitating a meeting for participants to discuss terms and initiate group operations more than doubles this effect, resulting in a 30 p.p. increase. The introduction of an additional third-party monitor has no additional effect. These effects persist for 12 months, during which collective sourcing also diffuses to the control group. Despite this convergence, early cooperation produces sustained business expansion, with higher wholesale quantities and profits, while greater treatment exposure reduces long-run retail prices by about 19 percent. In a hand-collected retailer × product panel, treated retailers experience a 12.6 percent reduction in wholesale unit costs and a 5.3 percent decrease in retail prices. Finally, a two-stage structural model with endogenous activation decomposes the gains from group formation into lower marginal procurement costs and improved product access. Consumers benefit mainly through lower prices, while retailers benefit primarily from sourcing more goods at a cheaper cost.

JEL Codes: O12, L14, D22, C93, L81

Keywords: cooperation, collective procurement, informal retail, market structure, variety, field experiment, Africa

*Eva Lestant, elestant@stanford.edu, Department of Economics, Stanford University; I am particularly thankful to the teams of the municipality of Bouaké. The experiment was registered in the AEA RCT Registry AEARCTR-0014073 and received approval from Stanford IRB (IRB-70352). I am extremely grateful to Jackson Hien for outstanding field coordination throughout the study. This work was partially supported by King Center, Weiss Fund, STEG and IPA West Africa. The authors acknowledge the use of AI tools for editorial assistance and proofreading to improve language clarity and readability.

1 Introduction

In developing countries, most people are subsistence entrepreneurs operating small firms that remain small. Common policy responses to stimulate growth have been to relax within-firm constraints by providing capital (de Mel et al., 2008; Fafchamps et al., 2014), managerial know-how (Bloom et al., 2013; Karlan and Valdivia, 2011), or information (Hjort et al., 2020). These programs are costly to deliver at scale, poorly tailored to small, single-owner firms, and have had only modest effects (McKenzie and Woodruff, 2014). More fundamentally, they assume that interventions must be delivered and evaluated at the firm level.

An alternative approach is to help small firms achieve economies of scale by pooling their input purchases. In the fresh-food retail sector, the conditions for such cooperation are ideal. Fresh-food marketplaces bring together many retailers who source the same goods at the same time from the same wholesale market and sell them side by side every day. Joint sourcing could lower transportation costs and provide access to bulk discounts. Yet in practice, retailers do not source jointly, suggesting that barriers to cooperation may be substantial.

This paper asks three questions: Can an intervention induce cooperation and joint sourcing? What prevents it from emerging on its own? And what are the welfare consequences for firms and consumers? To answer these questions, we conduct a randomized experiment testing three nested interventions with a hand-collected, high-frequency retailer-by-product panel in Côte d'Ivoire. The treatments address three frictions to cooperation: (i) uncertainty about others' willingness to cooperate, which we correct with an information treatment on second-order beliefs; (ii) the absence of a forum to agree on terms, which we supply through a one-time meeting; and (iii) limited trust in the group, which we restore by adding a monitor to oversee activities in the wholesale market on top of a one-time meeting. Finally, we use a structural model of Cournot competition with CES demand to test how much retailers internalize their group members' quantity decisions and to quantify the welfare impacts of group formation.

We developed a novel survey instrument to collect quantitative data on informal market businesses, due to the lack of administrative records and bookkeeping among retailers. The instrument provides information on individual retailers' sourcing and selling decisions over time. We start with an exhaus-

tive census of retailers operating in one of the twelve marketplaces in Bouaké, a secondary city in northern Côte d'Ivoire. The survey tracks 762 retailers across four waves over the course of a year. At every visit, enumerators catalog all the items displayed on each stall, identify its unit of sale (such as pile, kilogram, or basket), and record its visible quality, retail price, and stock. Each product is then linked to its sourcing transaction, including the date, quantity and unit cost. We also document the nature of the relationship between retailers and their suppliers at baseline. The result is a hand-built inventory for every stall at each visit.

Three descriptive facts from the survey data inform the experimental design. First, retailers in our sample have thin networks: they only have about two friends in the market and know just 10% of their neighbors by name.[1] Second, like many female entrepreneurs elsewhere (Ashraf et al., 2025), they show low trust: only 32% say they trust their partners, and 43% say lack of trust is the main reason they do not cooperate. Finally, 80% think joining a group would help their business. In other words, there is demand for groups, but retailers remain socially isolated.

The experimental design then tests how to alleviate cooperation frictions. The three arms are nested and increasing in intensity. The lightest is an *Information* nudge: at the end of the baseline survey, retailers are told how many others in their marketplace declared they were willing to cooperate. This is meant to correct pessimistic second-order beliefs. The *Meeting* arm is an invitation to a single facilitated meeting where retailers can voluntarily form a group on the spot. The *Monitoring* arm adds monitoring on top of the meeting to further restore trust and improve enforcement. A monitor accompanies the member who sources for the group to the wholesale market and checks prices and quantities. Across arms, including the control arm, newly formed groups receive two weeks of free transport (a reimbursed expense).[2] During this incentivized period, we collect daily surveys to track sourcing and selling decisions.

The first finding is that a light-touch measure leads to substantial cooperation among retailers. In the short run, interventions increased group participation by 26 pp. These groups consist of three or more retailers, one of whom sources goods for the others. These treatments had lasting effects: Twelve

[1] This is very similar to other measures of network density collected in rural areas (Banerjee et al., 2024).

[2] The goal of this experiment is not to test whether financial incentives nudge people to form groups; we therefore keep the incentive the same for all participants, including the control group, conditional on forming a group of 3 or more retailers.

months later, about 30% still actively participate, and an additional 20% have been part of a group at some point in time. This is in part because group formation diffused to the control group, which began participating in groups over time. Meeting-based treatments (*Meeting* and *Monitoring* arms) are most effective at encouraging group formation (+28 pp) but are indistinguishable from each other. In contrast, the *Information* arm achieves only half the effect. These results suggest that organizing collective procurement requires only a single facilitated meeting that gives retailers the opportunity to discuss the terms of a new practice, identify partners, and demonstrate public interest.

The groups formed through the intervention produce economically meaningful gains, reducing input costs and supporting greater sales volume and product scope. In the short run, treated retailers' wholesale unit costs decrease by 12.6%, indicating that they were able to access better prices for all goods in the wholesale market. Some of those benefits translate into lower retail prices. In particular, the selling price of regularly stocked goods decreases by 5.3%. Additionally, retailers within the same group have more similar product assortments at baseline, and this overlap increases during the first weeks of group formation. This pattern suggests both initial sorting and further convergence in product offerings. These changes align with efficiency gains that are partly passed on to consumers.

In the long run, as treatment diffuses, comparing treated and control retailers becomes a comparison between early and recent adopters, so individual-level specifications capture only cumulative effects. Consistent with this, early adopters demonstrate clear signs of business expansion: profits are 18% higher, wholesale purchases from the market are 0.29 log points higher, and the average number of clients served per hour increases by 17%. These results suggest that early participation helps retailers to progressively scale operations compared to late joiners. Therefore, the intervention generates lasting improvements in business performance, demonstrating that early participation allows firms to transform initial group benefits into sustained market growth.

To assess whether change in some retailers' sourcing practices also permanently altered pricing dynamics, the analysis examines whether 12-month equilibrium selling prices diverged from the counterfactual scenario without the intervention. Employing a shift-share specification at the good-market level, where the proportion of treated retailers in each good-market cell serves as the shift variable, the results indicate a 0.21 log-point reduction in selling prices. This substantial decrease demonstrates that short-term cost savings from group-sourcing retailers propagated throughout the market, reducing

average prices most significantly for products with greater initial exposure to treated sellers. These findings confirm that the intervention generated broader economic spillovers beyond individual retailers, shifting the market equilibrium to sustain lower prices for consumers over the long term.

The experiment shows that cooperation lowers procurement costs and retail prices, but the reduced-form estimates do not reveal why prices fall. In particular, cooperation can lower marginal costs, change which retailers sell a good, or alter how group members compete. These mechanisms have different implications for pass-through and surplus. We therefore build a structural model to separate the volume, access, and conduct channels behind the observed price decline.

The model takes purchasing groups as given and has two stages. Retailers first decide which goods to stock, with a lower setup cost when a group member already sources the good; they then compete in quantities, while group members pool wholesale orders and obtain lower input prices through bulk discounts. We also allow group members to internalize one another's profits when choosing quantities. But the incentives for within-group coordination are small, and the data reject full internalization of group members' profits. We therefore focus on quantifying the volume and access channels.

We quantify these channels through separate counterfactuals. The volume counterfactual lowers marginal procurement costs through pooled orders while holding fixed the set of retailers selling each good; the access counterfactual lowers setup costs, allows retailers to re-optimize which goods they stock, and then re-solves retail competition. The welfare incidence differs sharply across the two mechanisms. Consumers benefit primarily from the volume channel. Retailers, by contrast, gain primarily from the access channel, which allows them to profitably add goods they would otherwise not stock. Thus, cooperation raises consumer welfare mainly by making existing products cheaper, while it raises retailer welfare mainly by making more good more affordable to source.

This paper contributes to several strands of the literature on firm growth, cooperation, and private-sector development in low-income settings. In particular, we relate to work on interventions targeting firm constraints, the formation of productive relationships among firms, and experimental approaches to private-sector development.

First, we contribute to the literature studying interventions designed to stimulate firm growth in developing countries. Prior work attempted to address within-firm constraints and studied the effects

of inputs of capital (de Mel et al., 2008; Fafchamps et al., 2014) and labor (de Mel et al., 2009), or business-practice training (Bloom et al., 2013; McKenzie and Woodruff, 2014, 2017) on productivity. These types of programs have shown limited effectiveness among very small firms. In this paper, rather than directly supplying new resources, our intervention eases cooperation frictions through self-organized groups. These groups are persistent to the extent that they remain active after 12 months. We also show how cooperative behavior extends to control retailers.

To our knowledge, our nested design is the first to experimentally disentangle why firm owners fail to form mutually beneficial groups with direct competitors. We document this failure and relate it to three less visible frictions: thin networks, pessimistic second-order beliefs, and a lack of trust (Fleischman, 2026). A single facilitated meeting alleviates these frictions and increases cooperation, benefiting both retailers and consumers. These results are relevant to economic development, where the prevalence of small firms remains a central challenge: shared procurement offers one step toward a more structured sector.

Second, we contribute to the empirical literature on group formation among competing private firms. Much of this literature creates exogenous links between randomly matched firms to identify the value of business relationships (Fafchamps and Quinn, 2018; Cai and Szeidl, 2018). These studies find that such links improve performance by facilitating learning, the diffusion of management practices, and supplier-client matching. Yet when linked firms operate in the same sector, competitive concerns can limit the transmission of information that would benefit rivals (Hardy and McCasland, 2023; Cai and Szeidl, 2018). In these designs, the relationship is the treatment, and production remains an individual outcome. A separate strand studies firm-to-firm interactions sustained by mutually beneficial transactions, including trading access to idle resources (Bassi et al., 2022) or using an intermediary to aggregate purchases and provide shared services (Iacovone and McKenzie, 2022; Macchiavello and Morjaria, 2022). These arrangements are typically sustained through fee-based contracts. Our paper connects information and production sharing margins. Retailers choose their own group members, jointly organize procurement, and share information and access to supplier relationships. We therefore show that group formation among direct competitors is feasible and profitable once cooperation frictions are removed.

Finally, our methodology and policy findings advance experimental research on private-sector develop-

ment. We conduct, to our knowledge, one of the few population-scale firm experiments in a secondary city (Hardy and McCasland, 2021), covering the full population of regular retailers identified through an exhaustive census. Dynamic secondary cities provide a more inclusive path to poverty reduction when capital cities are crowded and hard to navigate (Henderson and Kriticos, 2018; Christiaensen et al., 2013). Finally, our low-cost intervention can also be incorporated into existing programs for female entrepreneurs, complementing evidence that peer-based support can strengthen such programs (Field et al., 2016).

The remainder of the paper proceeds as follows. Section 2 describes the setting, the data infrastructure, and the motivating facts on procurement scale and barriers to cooperation. Section 3 reports the experimental design and the take-up and downstream-market results. Section 5 develops the two-stage model. Section 6 estimates the wholesale-discount elasticity and the activation parameters and reports out-of-sample validation. Section 7 asks whether the downstream price decline reflects competitive cost pass-through or output-side collusion, testing three candidate coalitions against the data. Section 8 quantifies the welfare decomposition and counterfactuals. Section 9 concludes.

2 Setting, Data, and Stylized Facts

2.1 Setting and Main Puzzle

We study informal food retail in Bouaké, a secondary city in the north of Côte d'Ivoire. The empirical setting includes retailers selling fresh fruits and vegetables (F&V) who operate from stalls or tables in informal, clustered marketplaces. Retail is spatially dense (Fig. A.1): within each marketplace, many sellers offer overlapping bundles of F&V goods and operate side by side, often from a single selling point and with very limited capital or labor inputs (Fig. A.2). The typical retailer in our data is female and has long been established; she has been in the market for 13.77 years on average, sells 4.64 goods on a stall she runs alone (99.6% report no hired labor), and works nine hours a day, twenty-seven days a month (Table 1).

The organization of procurement is centralized, repetitive, and common across neighbors. Of these retailers, 87% source from Bromakoté, the main wholesale market; they travel there individually 2.87 times per week on average and spend 1.23 hours per trip. Each retailer transports her own inventory

back to her selling point by shared taxi, paying a fare proportional to distance and seats taken. Transport is the single largest fixed cost a retailer bears, 35% of her weekly fixed expenses (Fig. A.3), which roughly represents two days of profit. The median retailer's wholesale run is 29 kg, a fraction of a shared-taxi load, so the procurement problem each retailer solves is lumpy, frequently incurred, and nearly identical to the problem her neighbor is solving on the same morning. Procurement transportation cost is therefore lumpy, frequently incurred, and nearly identical across retailers.

Table 1: Retailer Business Features

	Mean	SD	p10	p25	Median	p75	p90
Panel A: Operating Model							
Tenure in this market	13.77	9.76	3	7	10	20	30
Respo's no of selling loc	1.11	0.41	1	1	1	1	1
No of good sold	4.64	2.51	2	3	4	6	8
Distinct goods on stall (baseline, by re7)	3.68	2.28	1	2	3	5	7
Can stock everything on site	0.91	0.28	1	1	1	1	1
Hours worked per day	8.77	1.41	7	8	9	10	10
Days worked in the last month	27.09	4.12	26	26	27	30	30
Months worked in the last year	11.55	1.63	11	12	12	12	12
Clients served per hour	5.08	4.05	1	2	4	7	10
Panel B: Procurement							
Hours spent buying inputs at Bromakoté	1.23	1.08	0	1	1	2	2
Buys from Bromakoté	0.87	0.34	0	1	1	1	1
# of trip to wholesale market/week	2.87	2.26	0	1	2	4	7
Buy all goods on credit	0.12	0.33	0	0	0	0	1
Share of weekly profit spent on transportation	0.24	0.41	0	0	0	0	0
Panel C: Return to scale							
No of regular wholesalers	2.22	2.09	0	1	2	3	5
Having at least one regular wholesalers	0.79	0.41	0	1	1	1	1
Always knows what to buy from wholesaler	0.56	0.50	0	0	1	1	1
Always know qtt to buy from wholesaler	0.41	0.49	0	0	0	1	1
Observations	762						

Source: Retailer baseline survey.
Notes: The table reports descriptive statistics for the full sample of retailers at baseline. Binary variables are reported as proportions. Tenure in this market is expressed in years. Hours spent buying inputs at Bromakoté and average number of clients per hour are computed among all retailers, including those reporting zero. Share of weekly profit spent on transportation is set to zero for retailers who do not incur transportation costs. The sample consists of all retailers enrolled in the study at baseline, prior to treatment assignment.

The small scale and repeated nature of individual procurement make coordination potentially valuable. Two or three nearby retailers could share a taxi and pool purchases to move down the wholesale quantity-discount schedule. Using the baseline cross-sectional data, Figure A.4 shows a negative correlation between purchase size and wholesale unit prices.[3] A second potential benefit is access to each

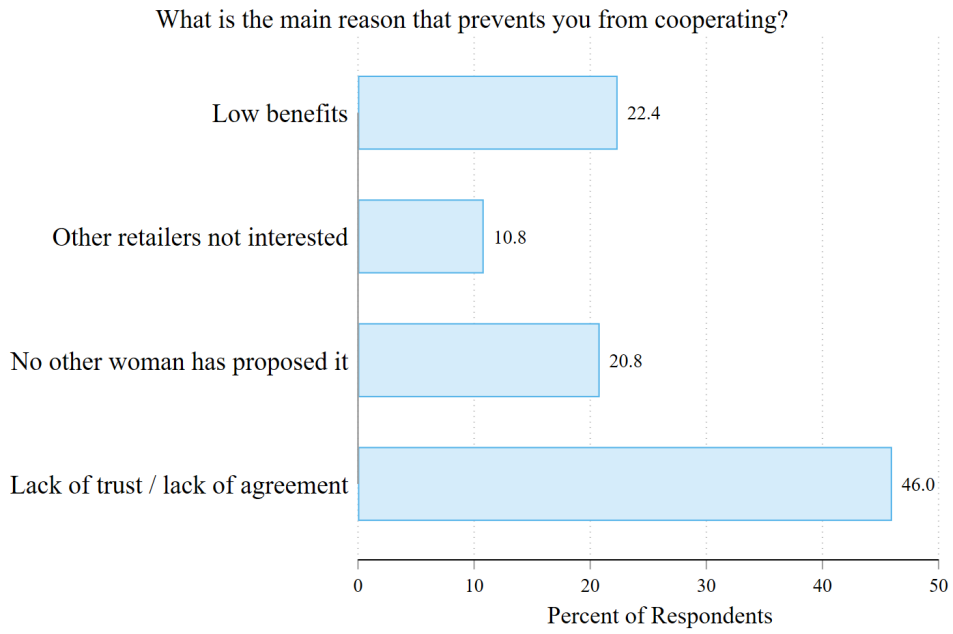
[3]Because larger buyers may differ systematically from smaller ones, this relationship is potentially endogenous and

other's regular supplier networks. Retailers report 2.22 regular wholesalers on average, and 79% have at least one.[4] In our setting, a regular wholesaler relationship exists if a retailer has the wholesaler's phone number registered. Among retailers with regular ties to wholesalers, 56% report that they always know what to buy from their wholesaler and 41% report that they always know the quantity to buy. These patterns suggest that regular supplier relationships reduce search and assortment-planning costs, creating another potential gain from pooling procurement. Pooled procurement can therefore lower costs through two distinct channels: by moving the group down the wholesale price schedule and by giving each member access to her partners' relationships with wholesalers.

Given those benefits, do retailers anticipate them? In the baseline survey, 75% of retailers describe cooperative sourcing as a potentially profitable practice. When asked why they have not yet adopted this new practice, lack of trust (46%) dominates lack of interested partners (11%) and cooperation constraints (21%) (Figure 1). The experimental design maps the resulting barriers to three nested candidate treatments. The first is informational: retailers may not know whether others in the same market are willing to cooperate. The second is relational and organizational: even when gains are obvious and others' preferences are known, finding partners, making the first approach, and agreeing on a trial arrangement may be socially costly. The third is enforcement: once a group forms, the member who travels to the wholesale market may misreport prices or quantities, creating distrust.

should not be interpreted as causal. We return to this relationship below using treatment assignment to isolate exogenous changes in purchase size and estimate the discount schedule slope.

[4]This means that only 20% of retailers exclusively source from the spot market without ever creating a more solid relationship with a wholesaler.



Source: Retailer baseline survey, 2024.

Notes: The figure shows the share of retailers reporting each factor as the main reason for not cooperating in procurement. Here, cooperation refers to an arrangement in which one retailer sources goods on behalf of other retailers.

Figure 1: Self-reported reasons for non-cooperative behavior

Baseline Network Yet at baseline, retailers behave as singleton procurement units.[5] Horizontal ties are sparse: the average retailer reports asking 1.35 other women for business advice, accompanying fewer than one woman on her wholesale trip,[6] and knowing the phone number of only 2.61 neighbors, which is roughly one in 20 neighbors (Table 2). Looking at the density of the network, we find that retailers know, on average, 9.6% of the other retailers' names in the same market.[7] Beyond the survey-based network information, we adapt the dual-incentive coupon scheme of respondent-driven sampling (Heckathorn, 1997; Salganik and Heckathorn, 2004) and validate the survey figures in Table 2.[8] This methodology is useful because retailers go by first names or nicknames, and many share the same name. Alternatives that require respondents to supply more than a name are impractical given the

[5] Anecdotally, retailers help each other with sourcing only when an unexpected event prevents them from going to the wholesale market themselves.

[6] This is different from coordinating; going to the wholesale market together means each retailer still needs to pay an individual transportation cost.

[7] This density measure is similar to that in villages in rural areas in India. See Banerjee et al. (2024).

[8] Our implementation is randomly seeded in a censused population, so the vertex set is known and only edges are unobserved.

anticipated thin network. This exercise measures pre-existing social ties using an incentive-compatible voucher game implemented before the baseline. A randomly selected subset of retailers (about 30% of retailers) received one "root" voucher ticket for themselves worth 1000 FCFA (approximately $1.70) and could request up to 10 additional "leaf" tickets to distribute to other retailers in the same market. Each leaf ticket had a small monetary value (500 FCFA) and could be redeemed during the baseline survey, thereby creating incentives to distribute tickets to friends/colleagues who were also selling fresh produce in the same market. We use the number of leaf tickets requested, the number of unique recipients, and the overlap between giver–receiver pairs to construct measures of network strength at baseline. We repeat this exercise 12 months after the experiment to study changes in social ties.

Table 2: Social Network Module

The respondent …	Nb. of connections			Density (% of market peers)		
	Mean (SD)	p25	p75	Mean (SD)	p25	p75
Asks for business advice	1.35 (2.23)	0	2	1.68 (3.44)	0	2
Goes to the wholesale market with	1.0 (1.7)	0.0	2.0	1.2 (2.2)	0.0	1.9
Cooperates with	0.6 (1.4)	0.0	0.0	0.9 (2.5)	0.0	0.0
Shares clients with	2.0 (2.0)	0.0	3.0	2.4 (2.8)	0.0	3.3
Resells goods to/from	1.5 (2.7)	0.0	3.0	1.6 (2.4)	0.0	2.5
Discusses selling prices with	1.5 (1.9)	0.0	2.0	1.7 (2.2)	0.0	2.7
Asks for personal advice	0.8 (1.1)	0.0	1.0	1.0 (1.6)	0.0	1.4
Knows the name of	7.7 (7.4)	3.0	10.0	9.6 (10.5)	3.0	11.6
Knows the phone number of	2.6 (2.8)	1.0	3.0	3.1 (3.6)	0.5	3.9
Considers as a sister	1.6 (1.9)	0.0	2.0	1.8 (2.2)	0.0	2.3
Observations	762					

Source: Retailer baseline survey. Notes: The table reports descriptive statistics for the full sample of retailers at baseline. The variables capture the structure and intensity of retailers' social interactions with neighboring traders in the market. Variables prefixed *Nb.* measure the number of other retailers with whom a given retailer engages in each type of interaction. The bottom four variables capture the number of neighbors and family members who would cover the retailer's stall in case of a short or prolonged absence, respectively. All count variables are reported in levels. The sample consists of all retailers enrolled in the study at baseline, prior to treatment assignment.

2.2 Data Collection

To study how encouraged cooperation changes procurement, prices, quantities, and assortment in this setting, we constructed a novel high-frequency retailer×product panel. We first conducted a census of 12 informal markets in Bouaké, covering 1,037 retailers in total. Our study focuses on retailers selling F&V inside marketplaces.[9] In July 2024, we surveyed 762 retailers who were actively selling on the survey day and declared that they were *regular* sellers.[10] We track retailers over time using identifiers based on name, phone number, stall location, and enumerator verification. The study follows an active-seller sample through the intervention period and subsequent 4- and 12-month follow-ups.

Our main analysis sample uses the set of retailers observed in the high-frequency panel during the treatment period: 740 retailers tracked multiple times over 10 days. As such, we build a panel structure at the retailer × day level and, within each day, at the retailer × product level. We document sample sizes and attrition by wave (3.5% at endline and 16.3% at 12-month follow-up), and we show that attrition is balanced across treatment arms in Table A.1. Data collection was organized around four main survey waves (as summarized in Table A.2), plus short phone check-ins.

2.2.1 Exhaustive Good Roster

The strength of our data is the information we collected on the exhaustive bundle of items displayed on the retail store's stall at the time of the interview. In practice, enumerators had to list all the goods displayed on the stall and answer a number of questions for each good. In the roster, each good is defined as a unique combination of produce (e.g., tomato, red onion, …) and a unit in which it is sold (e.g., sold by unit, pile, kg, basket).

For each listed good, after reporting the type of produce and the unit for sale, the enumerators record the visible quality[11] and retail price as well as the lowest price the retailer is willing to accept that day.[12]

For quantity estimates, enumerators ask how much was sold since the start of the day, how much

[9] We exclude retailers operating outside of the market, street sellers, and established stores.

[10] Enumerators were instructed to simply ask the question *Are you a regular seller? Like you sell here almost every day?* without imposing a hard constraint.

[11] In our sample, across waves and goods, 78% are of good quality, 21% are of medium quality, and 0.5% are of low quality.

[12] In our sample, both measures coincide 85% of the time.

remains on the stall at the time of the interview, how much was sold yesterday, and how much was wasted since yesterday. Together with the unit size, these measures allow us to construct comparable unit prices and quantities.

Finally, the roster links each retail good to the specific retailer's sourcing decision. We record the date of the retailer's purchase from a wholesaler, the quantity purchased, the unit of measure used at purchase, and the purchase price paid to the wholesaler. We also collect information on the wholesaler relationship: the wholesaler's location, whether the purchase was made on credit, whether the retailer has the wholesaler's phone number, the type of relationship, and how many years the relationship has lasted. In total, less than 2% of goods are sourced from friends or family and the average relationship with any wholesaler is 5 years. During the treatment period, we ask whether the purchase was made as part of a group purchase, the number of other women in the group, and how many of them bought the same product.

2.2.2 Retailer Survey

Baseline (July 2024; 1 week). Enumerators visited markets over one week and surveyed 762 active retailers. The baseline survey has two parts. First, a retailer-level module records business characteristics and background information, including: (i) demographics (age, education, household structure), (ii) business tenure and experience, (iii) labor inputs and work schedule, (iv) selling location and market participation, and (v) sourcing practices (main wholesale market, sourcing frequency, transport mode, and payment methods). Second, we survey characteristics of the sold goods using the good-level roster described above. We also have a module to compute profit in which we collect weekly expenses and costs and ask for a self-declared approximation of profit, reconciling both approaches with the respondent (de Mel et al., 2009). The baseline also includes a social-network module (number of other retailers consulted for advice, joint trips to the wholesale market, cooperators, and phone-number contacts) and a self-declared-barriers module that elicits trust, reservation wage, and the main reasons for not cooperating.

Transportation costs are measured at the trip level. Since prices are fixed and proportional to distance and the number of seats taken in the taxi, we ask retailers about where they source and the cost of their most recent trip to the wholesale source. We construct weekly transport costs based on the sum of

reported transport payments and their frequency.

Midline / treatment period (high-frequency; about 2 weeks). During the treatment period, we collected high-frequency data for 740 retailers. Each retailer was surveyed approximately four times over 10 days.[13] At each visit, enumerators collected the good-level roster information. The midline sample includes retailers who participated in groups, as well as a random daily sample of non-participants.

The midline instrument is designed to generate daily (or near-daily) measures of prices, quantities, and assortment and uses the same good-level roster as the baseline. The midline data are used to measure how cooperation changes procurement, prices, quantities, and assortment during the treatment period.

Endline (shortly after the treatment period; <1 week). We conducted an endline survey of 735 retailers. The endline repeats the baseline modules to measure changes in beliefs, business practices, and social ties after the treatment period.

Phone check-ins (December 2024). We also conducted short phone surveys in December 2024. These check-ins record retailer participation status (still operating or not), current market, contact information updates, and brief measures of cooperation and sourcing behavior. We use these data mainly to track the sample and to measure medium-run persistence of treatment effects.

12-month follow-up (July 2025). One year later, we re-surveyed 638 retailers. This follow-up repeats the baseline survey and includes a good-level roster. Each retailer was interviewed twice, one week apart, to improve measurement of business outcomes and to reduce noise from day-to-day shocks. We use the two interviews to construct more stable measures of assortment and sourcing frequency. The main reasons for non-response were sickness and pregnancy (73% of non-response cases).

2.2.3 Consumer Survey

In 2023, prior to the experiment, we also collected consumer survey data to study shopping behavior and search frictions. The consumer survey records (i) where consumers usually shop, (ii) how many

[13] Those 10 days spanned two weeks, with enumerators working Monday to Friday.

stalls they visit per trip, (iii) whether they compare prices, (iv) the importance of reliability and product availability, and (v) the role of loyalty rewards and repeat relationships. The sample is made up of consumers randomly selected from a subset of marketplaces where retailers were surveyed. Enumerators followed consumers while they shopped and collected basic demographic data at the end, usually while the consumer was waiting for a taxi. At the same time, we also collected retailer data. We are able to match about 51% of the consumer survey sample to the retailer survey, which allows us to link consumer behavior to the characteristics of the retailers from whom they shop. **EL: add the evidence of search frictions and loyalty rewards.**

3 Experimental Design and Participation Results

3.1 Treatments and Implementation

The randomized controlled trial includes a single experiment with three treatment arms designed to test nested nudges. We implement the experiment in 12 marketplaces in Bouaké, and we randomize, a priori comparable, 762 retailers, stratified by marketplace. We provide retailers' characteristics by treatment in Table A.1.

To encourage retailers to work together, newly formed groups that meet the eligibility rule (e.g., at least 3 members) receive short-term facilitation during the first two weeks (e.g., free transportation for two weeks). This is not a cash transfer but a reimbursed expense. To implement these incentives, we station a team member at the taxi station near the wholesale market. When the *Leader* is done with the purchases, she calls the research team staff, who find her and pay the taxi driver directly. Note that the control group would also receive the same facilitation if they were to form a group, but they are not encouraged to do so. The facilitation is removed after two weeks, and we track groups and outcomes over time through medium-run and long-run follow-ups (Figure A.6).

The interventions take place right after the baseline survey, and the implementation of the treatments is as described:

1. **Information (reduce information frictions).** We provide each retailer with information collected at baseline on other retailers' stated willingness to collaborate. After the end of the survey,

we simply provide the proportion of retailers operating in the same marketplace who have expressed interest in forming a CP group. The goal is to reduce uncertainty about whether potential partners would agree to share transport and purchase jointly.

2. **Meeting (reduce social initiation costs).** We invite retailers to a one-time structured meeting held in each marketplace. After a brief, simple introduction[14] to the potential gains of a collective purchasing group, retailers can discuss and voluntarily form CP groups. This treatment is designed to reduce the social cost of starting cooperation and adopting a new practice potentially against the current norm (taking the first step, asking others, coordinating a first trial). The group can be formed with anyone present at the meeting. To include individuals absent from the meeting, retailers have to verify, using the colored cards provided at baseline, that these individuals are in the same treatment arm. Groups formed with individuals outside the treatment arm are ineligible for the financial incentive intended for CP group members.

3. **Monitoring (reduce moral hazard and conflict).** This treatment includes the same meeting as previously described and adds monitors[15] assigned to newly formed groups. Monitors join the *Leader* of the group in the wholesale market to verify that wholesale purchases are made as planned (e.g., quantities and prices paid), and to reduce mistrust, misreporting, and disputes within the group.

Treatment assignment shares are approximately 10% control, 15% Information, 35% Meeting, 40% Monitoring. For each treatment arm, the necessary information is distributed at the end of the baseline survey. For Meeting and Monitoring (T2/T3), we announce the time and place of the meeting, then hold it a day or two later. Each arm meets separately, in the marketplace.[16] For Monitoring (T3), we introduce the monitors right after the meeting and share their phone numbers with retailers.

[14]The introductions were provided by local PhD students from the sociology department of the Université Alassane Ouattara de Bouaké. They were directly trained by the PI and supervised by the Field Coordinator. The meeting lasted 30 to 40 min, but people were free to come and go at any time.

[15]These monitors are also PhD students recruited from the local university in Bouaké.

[16]We coordinate with elders of the market to find a place in the marketplace that could host retailers. We try, when possible, to have a location on the outskirts of the marketplace, but we cannot guarantee that non-invited participants do not overhear what is explained in the meeting.

3.2 First-Stage Results: Group Formation

The key outcome of interest is whether retailers form and sustain *Collective Purchase groups* (CP groups).[17] We estimate intent-to-treat (ITT) effects for take-up outcomes at key horizons (e.g., early participation, active at 6 months, active at 12 months). We present both a pooled specification (any treatment vs. control) and an unpooled specification (each arm separately):

$$Y_{im} = \mu_m + \tau^P \mathbb{1}\{T_i \neq 0\} + \varepsilon_{im}, \tag{1}$$

$$Y_{im} = \mu_m + \tau_1 \mathbb{1}\{T_i = 1\} + \tau_2 \mathbb{1}\{T_i = 2\} + \tau_3 \mathbb{1}\{T_i = 3\} + \varepsilon_{im}. \tag{2}$$

All specifications include marketplace fixed effects, so identification comes from comparing retailers within the same marketplace. We report heteroskedasticity-robust standard errors.

Table 3 shows large short-run increases in collective purchase (CP) group participation. During the two-week treatment period with incentives, the probability of ever participating rises by 0.261 (s.e. 0.040). These impacts persist but attenuate over time: in December 2024, the pooled effect is 0.200 (s.e. 0.046) relative to a control mean of 0.19, and in July 2025, it is 0.096 (s.e. 0.044) relative to a control mean of 0.19.

The unpooled estimates show that treatment arms that include a meeting component drive most of the take-up. The *Information* treatment has small and statistically insignificant effects on early participation (0.049, s.e. 0.045). In contrast, both meeting-based treatments have significantly larger effects than *Information* alone: the p-value for $T1 = T2$ is below 5% for all outcomes.

Finally, note that the control mean is not zero: 9% of control retailers participate in a CP group during the treatment period. This represents 6 people who join groups of treated retailers.[18]

[17] Those groups are defined such that one member, the *Leader*, travels to the wholesale market alone to buy inputs for the group, including herself, and splits transportation costs.

[18] They formed groups with retailers who were assigned to the Meeting treatment arm. At endline, 3 of them declared that they had attended the meeting.

Table 3: Participation in Collective Purchase Group

	Participated Early (Week 1)	Ever Participated in July 24
Panel A. Pooled		
Treatment	0.135	0.259
	(0.035)	(0.040)
	{0.000}	{0.000}
Panel B. Unpooled		
T1: Info	0.049	0.139
	(0.045)	(0.051)
	{0.270}	{0.007}
T2: Meetings	0.148	0.296
	(0.039)	(0.045)
	{0.000}	{0.000}
T3: Monitor	0.156	0.272
	(0.038)	(0.045)
	{0.000}	{0.000}
Observations	757	757
P-Val T1=T2	0.011	0.000
P-Val T2=T3	0.801	0.519
Control Mean	0.07	0.09
Fixed Effects	Mkt	Mkt
Clustering Lvl.	{Robust}	{Robust}

Source: Retailer panel survey, baseline, midline, follow-up and long-term survey.

Notes: The table reports ITT estimates from a cross-sectional regression of each participation outcome on treatment assignment, estimated using OLS with location fixed effects. Panel A reports the pooled treatment effect. Panel B reports treatment effects by arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Robust standard errors are reported (in parentheses); p-values are reported in {braces}. Participated Early (Week 1) equals one if the retailer joined the cooperative arrangement during the first week of the treatment. Ever Participated in July 24 equals one if the retailer participated at any point during the respective reference period. The control mean is computed among control retailers. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding treatment arm coefficients. * p<0.10, ** p<0.05, *** p<0.01.

The main takeaway is that providing a time and space for retailers to meet and discuss cooperation opportunities is an effective strategy to overcome the social costs of initiating cooperation. Providing information alone produces smaller and statistically insignificant effects. Adding in-person monitors does not further increase take-up relative to meetings, suggesting that enforcement frictions are second-order for group formation in this setting.

3.3 Diffusion of Cooperation Over Time

Table 4 describes the change in participation in CP groups over time and partial convergence. Treatment effects after 4 months and 12 months are persistent. In fact, meeting-based treatments show the same strong effect of +23p.p. compared to a control mean of 19% at 4 months, which means that about 40% of treated retailers are now participating in the CP groups. The *Information* treatment effect is weaker but only because the control group has caught up. In the longer run, meeting-based treatments are still 10p.p. higher than controls and the Information treatment.

We also look at treatment effects on the ever-participation outcome, which captures the total number of retailers who have ever participated in a CP group at any point during the study. The pooled effect is 0.165 (s.e. 0.052) relative to a control mean of 0.32 such that, on average, about 50% of retailers have participated in a group at least once by the end of the study period. This is particularly important for understanding the long-term impacts and, in particular, benefits of the groups that cannot be unlearned once experienced. Retailers have learnt that they can trust each other and that the group can be organized without conflict. They might also have learnt from each other about business practice or information about the wholesale market.

Taken together, these results show that the intervention lasts beyond the incentive period. Groups formed during the treatment period survived the removal of the incentive and spread beyond the retailers we treated. That diffusion is itself evidence on the value of the new practice. The intervention created an occasion to coordinate and a low-cost way to test the arrangement. Once treated retailers showed that a group could be formed and a *Leader* could travel for the group without conflict, this organizational template became appealing and started to be imitated locally. Retailers who watched their neighbors run a group then formed one of their own or joined existing groups (Hanna et al., 2014).

Table 4: Long-term Participation in Collective Purchase Group

	Still Participated in Dec. 24	Still Participated in July 25	Ever Participated (Full)
Panel A. Pooled			
Treatment	0.205	0.096	0.165
	(0.046)	(0.044)	(0.052)
	{0.000}	{0.028}	{0.001}
Panel B. Unpooled			
T1: Info	0.085	0.031	0.066
	(0.058)	(0.054)	(0.064)
	{0.145}	{0.567}	{0.302}
T2: Meetings	0.213	0.106	0.190
	(0.050)	(0.048)	(0.055)
	{0.000}	{0.028}	{0.001}
T3: Monitor	0.242	0.112	0.182
	(0.050)	(0.048)	(0.055)
	{0.000}	{0.019}	{0.001}
Observations	757	757	757
P-Val T1=T2	0.007	0.085	0.009
P-Val T2=T3	0.445	0.857	0.811
Control Mean	0.19	0.19	0.32
Fixed Effects	Mkt	Mkt	Mkt
Clustering Lvl.	{Robust}	{Robust}	{Robust}

Source: Retailer panel survey, baseline, midline, follow-up and long-term survey. Notes: The table reports ITT estimates from a cross-sectional regression of each participation outcome on treatment assignment, estimated using OLS with location fixed effects. Panel A reports the pooled treatment effect. Panel B reports treatment effects by arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Robust standard errors are reported (in parentheses); p-values are reported in {braces}. Participated Early (Week 1) equals one if the retailer joined the cooperative arrangement during the first week of the treatment. Ever Participated (Full) equals one if the retailer participated at any point during the respective reference period. Still Participated in Dec. 24 and Still Participated in July 25 equal one if the retailer remained actively participating as of the corresponding survey wave, and thus capture the sustainability of take-up over time. The control mean is computed among control retailers. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding treatment arm coefficients. * p<0.10, ** p<0.05, *** p<0.01.

3.3.1 Effect on Networks and Trust

To write using endline + LT survey.

4 Effects on Market Outcomes

The next question is whether upstream cooperation yields cost savings that translate into downstream changes. First, we measure the effect of the intervention on the quantity purchased by retailers at the wholesale market and their sourcing costs. Second, we evaluate whether these effects translate into changes in selling prices and product variety. These outcomes capture the instrumental return that CP groups can provide.

Empirical Specifications In this section, all treatments are pooled. We will be comparing treated and control retailers selling the same product in the same marketplace at the same time. The analysis uses the high-frequency roster collected during the two-week treatment period, in which outcomes are observed at the retailer $i \times$ product $g \times$ day d level.

Our main outcomes of interest are the effects on the quantity purchased at the wholesale market and the input price paid for these goods. We also report selling prices and product variety. To construct those outcomes, we record selling prices and quantities in the local unit of trade (e.g., kilogram, bowl, pile, piece) for each retailer and good pair. When a good is reported in different units across retailers or days, we harmonize units using a conversion table built by enumerators through calibration exercises.[19]

Throughout, we distinguish the *core bundle* of frequently traded goods, those a retailer carries on at least 80% of her survey days, from the full roster, which keeps every good we see on the stall. Working with a balanced set of products observed at both baseline and midline nets out a composition effect. In itself, this is an interesting subsample of goods that are frequently traded and known to consumers. Thus we report both core and full samples estimates.

The quantity outcome is the quantity bought at the wholesale market, expressed in kilograms and recorded once per wholesale batch; duplicated batch records are dropped. We average it in two ways: at the retailer ($\times$ week) level and at the retailer $\times$ good ($\times$ week) level.

The selling price is the retail price per unit, observed for each retailer $\times$ good $\times$ day in the high-frequency roster. The wholesale unit cost is the corresponding price paid per unit. Because the panel

[19] Enumerators independently measured unit-to-kilogram conversions for each good multiple times across markets; we use the average conversion factor.

is short, we do not deflate prices; instead, we control for time fixed effects.

We also construct simple proxies for unit-level profitability measures using the retail unit price P and its corresponding wholesale unit cost MC. We define unit marginal profit as $P - MC$ and Lerner index as $(P - MC)/P$. To limit the influence of extreme values, we winsorize these measures at the product level (5th percentile).

To analyze this change in product scope, we construct two measures of variety. First, daily variety is the number of distinct goods a retailer sells on a given day. Second, two-week variety is the total number of distinct goods the retailer sold at least once during the two-week window (the union of goods across days). For the first outcome, we can easily compute our DiD estimate using the baseline roster as a pre-treatment benchmark. For the second outcome, we can only compare treated and control retailers at midline because the two-week window does not exist at baseline.

Although treatment is randomized at the retailer level, our outcomes of interest, selling prices, and wholesale unit costs, are observed at the retailer × good level. We adopt the DiD specification as our preferred design because the baseline good-level roster is itself rich enough to serve as a within-cell pre-treatment benchmark. Following McKenzie (2012), conditioning on the baseline good-level roster improves precision when outcomes are sufficiently autocorrelated. We use a DiD specification because price autocorrelation at the good × week level is high (well above the $\rho > 0.5$ threshold at which differencing dominates a post-only design); the more efficient ANCOVA alternative yields very similar estimates (Table A.6) but reduces the sample to only goods that can be observed at both baseline and midline. Finally, we report the cross-sectional ITT (Table A.4); the three specifications point in the same direction, with DiD and ANCOVA yielding tighter standard errors on the price coefficient. For retailer i in marketplace m, selling product g, observed in period $p \in \{0, 1\}$ (baseline vs. midline), we estimate:

$$Y_{igmp} = \alpha + \beta_1 \,\text{Treat}_i + \beta_2 \,\text{Post}_p + \beta_3 \,(\text{Treat}_i \times \text{Post}_p) + \mu_{mg} + \varepsilon_{igmp}, \tag{3}$$

where $\text{Treat}_i = \mathbb{1}\{T_i \neq 0\}$ is an indicator for being assigned to any treatment arm (pooled), and $\text{Post}_p = \mathbb{1}\{p = \text{midline}\}$ is the post-treatment indicator. The fixed effects μ_{mg} absorb all time-invariant differences across marketplace × product cells, so identification comes from comparing

treated and control retailers selling the same product (pooling all goods of the same product together) in the same marketplace. The coefficient of interest, β_3, is the pooled ITT DiD effect on downstream outcomes.

All specifications use probability weights to account for differential observation and attrition between baseline and midline measurements. We cluster standard errors at the retailer level to allow for arbitrary correlation within retailers across products and weeks.

Wholesale Quantity Encouraged cooperation raises the quantity retailers buy at the wholesale market, but the increase sits outside the core bundle. In Table 5, the pooled treatment DiD interaction is 0.068 (s.e. 0.111) on core goods and 0.245 (s.e. 0.110) on the full roster, relative to a control mean of 2.82. The good-level panel reproduces the same contrast, 0.034 (s.e. 0.104) against 0.181 (s.e. 0.084). Treated retailers do not buy more of the staples they already sourced regularly but buy more of the goods they previously carried only occasionally. Although the unpooled estimates suggest that the core-bundle response is concentrated in the Information treatment, equality tests do not reject that the three arms move together. The quantity margin therefore responds to being able to source jointly and experimenting with more non-core goods.

Table 5: Effects on Wholesale Quantity, by Treatment Arm

	Retailer: per line		Good: per line	
	Core	Full	Core	Full
Panel A. Pooled				
Treat × Midline	0.077	0.235	0.038	0.181
	(0.109)	(0.101)	(0.104)	(0.084)
	{0.479}	{0.020}	{0.718}	{0.031}
Panel B. Unpooled				
T1: Info × Midline	0.276	0.242	0.178	0.281
	(0.157)	(0.147)	(0.153)	(0.109)
	{0.079}	{0.102}	{0.246}	{0.010}
T2: Meetings × Midline	0.098	0.265	0.041	0.161
	(0.119)	(0.109)	(0.114)	(0.091)
	{0.412}	{0.015}	{0.717}	{0.077}
T3: Monitor × Midline	-0.021	0.204	-0.016	0.162
	(0.116)	(0.110)	(0.113)	(0.089)
	{0.859}	{0.062}	{0.887}	{0.068}
Observations	1760	2060	3547	8483
P-Val T1=T2	0.192	0.853	0.308	0.166
P-Val T2=T3	0.169	0.452	0.501	0.984
Control Mean	2.97	3.23	2.58	2.53
Fixed Effects	Retailer, Week	Retailer, Week	Mkt x Good, Retailer	Mkt x Good, Retailer
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, baseline and midline.

Notes: The outcome is the log quantity bought at the wholesale market (kg), recorded once per wholesale batch; duplicated batch records are dropped. The four columns differ only in the unit the data are collapsed to and in the set of goods retained. The retailer-level columns average the logged quantity within retailer × wave × week and absorb retailer and week fixed effects; the good-level columns average the raw quantity within retailer × good × wave × week, take logs after the collapse, and absorb market × good and retailer fixed effects. "Core" keeps goods present on at least 80% of the retailer's survey days; "Full" keeps every good on the roster. All columns are estimated by `reghdfe` with inverse-probability-of-censoring weights, and standard errors are clustered at the retailer level, the unit of randomization. Only the difference-in-differences interaction is reported: both collapses absorb the retailer, so the treatment-level term is collinear and carries no information. Standard errors in parentheses and p-values in {curly brackets}. The control mean is the mean of the outcome, in logs, among control retailers at baseline. P-Val T1=T2 and P-Val T2=T3 report F-tests of equality between the corresponding Treat × Midline coefficients. Observations are counted at each column's collapse unit, which is why they differ across columns.

Unit Costs and Prices All treatments lower both what retailers pay at the wholesale market and what consumers pay at the stall, but the cost savings are about twice the size of the price cut. Table 6 reports DiD estimates. In Panel A, the intervention decreases log selling prices by -0.054 ($p = 0.046$), corresponding to about -5.3%. At the same time, it decreases log wholesale unit costs by -0.135 ($p = 0.001$), corresponding to about -12.6%. Both the retail price and the marginal cost of procurement fall in response to the intervention, and both effects are statistically significant. Together, these patterns show that the intervention lowered both the procurement cost faced by retailers and the price observed by consumers.

Table 6 (Panel B) decomposes the DiD effect by treatment arm. Across arms, the wholesale-cost reductions are largest in T1 and T3, while the price response is clearest in T1. Formal tests do not reject equality of arm-specific DiD effects: p-values for T1=T2 are 0.176 (prices), and 0.068 (wholesale costs), and for T2=T3 are 0.479, and 0.279, respectively. Thus, while point estimates suggest larger effects under Information, the data are also consistent with broadly similar impacts across arms. This indicates that groups, regardless of how they were formed, induce the same downstream effects.

Table 6: Effects on Retail Prices, and Costs, by Treatment Arm

	Core bundle		Full bundle	
	Log Wholesale Cost	Log Selling Price	Log Wholesale Cost	Log Selling Price
Panel A. Pooled				
Treat × Midline	-0.135	-0.054	-0.090	-0.038
	(0.042)	(0.027)	(0.034)	(0.022)
	{0.001}	{0.046}	{0.009}	{0.082}
Panel B. Unpooled				
T1: Info × Midline	-0.200	-0.085	-0.150	-0.054
	(0.062)	(0.035)	(0.046)	(0.028)
	{0.001}	{0.015}	{0.001}	{0.057}
T2: Meetings × Midline	-0.097	-0.035	-0.074	-0.045
	(0.046)	(0.034)	(0.039)	(0.024)
	{0.036}	{0.301}	{0.055}	{0.068}
T3: Monitor × Midline	-0.146	-0.059	-0.082	-0.026
	(0.051)	(0.032)	(0.038)	(0.024)
	{0.004}	{0.061}	{0.029}	{0.282}
Observations	5889	7374	12307	15106
P-Val T1=T2	0.068	0.176	0.058	0.708
P-Val T2=T3	0.279	0.479	0.792	0.352
Control Mean	2.72	3.37	2.74	3.38
Fixed Effects	Mkt x Good, Week	Mkt x Good, Week	Mkt x Good, Week	Mkt x Good, Week
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, baseline and midline.

Notes: The table reports difference-in-differences estimates from a regression of each outcome on treatment assignment, a post-period indicator (Midline), and their interaction, estimated with market × good fixed effects. Observations are weighted by IPCW at midline; baseline observations receive a weight of one. Panel A reports the pooled Treat × Midline coefficient. Panel B reports the interaction by treatment arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Standard errors are clustered at the retailer level and reported (in parentheses); p-values are reported in {curly brackets}. The control mean is computed among control retailers at baseline. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding Treat × Midline coefficients across arms. * p<0.10, ** p<0.05, *** p<0.01.

Treated retailers buy goods more cheaply and return part of that saving to their customers, but they focus their efforts on a particular set, their core bundle. This includes the goods customers come to a stall for because they expect the item will be in stock. On the full bundle, wholesale costs still fall significantly, but the selling-price cut is noisier (Table A.5). One explanation is that only repeat buyers, who mostly come for the core bundle and know the price history, can notice a few percent off; a cut on an occasional good is invisible to the people who would reward it, so retailers concentrate their signaling effort where it is best detectable. This suggests that some form of competition exists, but the reduced-form results cannot characterize it, which motivates the need for our structural model.

4.1 Results on Variety and Entry

Because retailers make sourcing decisions every time they go to the wholesale market, the intervention may also change the set of goods they carry. Treated retailers can expand their assortment at low cost by adding goods the group is already sourcing for other members. Assortment is itself a competitive margin: a retailer carrying more goods can attract new customers. We first present experimental results on variety expansion, then descriptive evidence on bundle composition and overlap within groups.

Table 7 shows that treated retailers' daily product scope is not significantly different, but the total number of distinct goods present on their stalls over the two-week period expands: treatment changed rotation rather than the breadth of goods sold. On a daily basis, retailers carry 3.30 goods; if anything, treated retailers have fewer distinct goods. However, over the two-week window, treated retailers' variety expands by 0.955 goods relative to a control mean of 4.73 (about 20%).

These results point to a high rotation effect rather than a permanent addition of new goods. To further test this hypothesis, we define a retailer-level assortment rotation index, (R_i), as the number of distinct goods observed on the retailer's stall over the full midline survey period divided by the average number of distinct goods displayed on a given survey day. The index equals one when a retailer carries exactly the same set of goods every day and increases as the composition of the stall changes across days. In Table A.8, we show that treatment increases the rotation index by 0.128 relative to a control mean of 1.33 ((p=0.005)), implying substantially greater turnover in the set of goods displayed over the survey period. The estimate remains similar when we control for baseline daily variety, at 0.099 ((p=0.024)).

The treatment effect is positive across all three treatment arms and is largest for the monitoring arm. In levels, control retailers hold roughly (1/1.33=75%) of their total midline assortment on a typical day, compared with about (1/1.46=69%) among treated retailers. Thus, the increase in the number of distinct goods observed over the wave appears to reflect more frequent rotation through a broader assortment, rather than retailers permanently expanding the number of goods displayed at any one time.

Table 7: Treatment Effects on Variety Expansion

	Daily variety (ANCOVA)	Two-weeks Variety (ANCOVA)
Panel A. Pooled		
Treatment	-0.061	0.381
	(0.135)	(0.251)
	{0.652}	{0.129}
Panel B. Unpooled		
T1: Info	-0.022	0.318
	(0.167)	(0.333)
	{0.896}	{0.339}
T2: Meetings	-0.146	0.127
	(0.146)	(0.276)
	{0.318}	{0.644}
T3: Monitor	-0.003	0.624
	(0.145)	(0.279)
	{0.982}	{0.025}
Observations	740	740
P-Val T1=T2	0.335	0.503
P-Val T2=T3	0.147	0.021
Control Mean	3.36	4.73
Fixed Effects	Mkt	Mkt
Clustering Lvl.	{Retailer}	{Retailer}

Notes: *Daily variety* is the number of distinct goods observed on a given day, averaged within retailer × wave × week, and exists in both waves. *Two-week variety* counts goods seen on *any* midline day; it is post-treatment only, so column 2 takes its pre-period from daily variety on the single baseline day. Column 3 is the midline cross-section of the same measure. All specifications are retailer-level with market fixed effects, week fixed effects in column 1, IPCW weights at midline, and standard errors clustered at the retailer level. Standard errors in parentheses, p-values in {curly brackets}. The P-Val rows are F-tests of arm equality on the interactions in columns 1–2 and on the levels in column 3. Control means are baseline in columns 1–2 and midline in column 3. Source: Retailer panel survey, midline survey.

Retailers in the same CP group may expand variety by learning from one another where to source new goods. A member who has sold a good for years knows its wholesale market, the wholesalers who carry it, and the prices they charge. This knowledge can translate into a relational contract with a particular wholesaler. At baseline, 32% of core goods come from a wholesaler the retailer has repeatedly interacted with and 20% from one whose phone number she has; 17% satisfy both. When sourcing together, each member needs to disclose the wholesaler relationship she has built and, in exchange, gains access to the relationships the others bring. The group therefore not only pools orders but also sourcing relationships.

To investigate how groups might shape bundle choices, we compare observed groups to placebo groups randomly formed. The placebo benchmark isolates the assortment overlap that would occur by chance among any group of retailers. We first categorize goods according to their sourcing channel; then we define a within-group similarity measure that we track across time.

We categorize each (i, g) pair sold at midline into four mutually exclusive groups: (1) own core bundle good, where the good is already included in the retailer's pre-treatment core or is covered by the retailer's own contract; (2) group-exposed good, when the good is included in a peer's pre-treatment core or contract but not in the retailer's own; (3) good sourced outside observable channel; and (4) goods sold by group members, but which is not part of their core bundle. See Table A.7 in Appendix A.

Among grouped retailers ($N_{(i,g)} = 1{,}753$), 55.0% of midline-sold goods are own entry, 15.6% are group-exposed via a peer's prior wholesale relationship, 20.3% match a good a peer previously sold, and only 9.1% have no observed channel. The mirror image for singletons ($N_{(i,g)} = 2{,}384$) is 64.1% own entry and 35.9% with no observed channel. The two peer-mediated channels are mechanically zero. To check that these peer-mediated channels reflect the actual composition of CP groups, the third column reassigns *all* retailers to random groups within their own marketplace ($N_{(i,g)} = 4{,}137$). Under this placebo benchmark, the two peer channels account for only 10.1% (group-exposed) and 13.1% (group ever sold) of entry, a combined 23.2% that falls well short of the 35.9% observed for real groups. The realized groups exhibit significantly higher levels of peer-mediated entry than those observed among randomly assigned retailers. The structural model presented in Section 5 formalizes the mechanisms underlying peer-mediated entry.

4.2 Long-term changes in market outcomes

Because the practice of cooperation spread widely after the intervention, the long-run intention-to-treat (ITT) estimates are no longer informative except for outcomes for which prolonged exposure to group formation may have persistent effects. This scenario is plausible for private decisions that accumulate over time, such as the quantity purchased in wholesale markets, which may indicate a larger market share secured over time. However, treatment-assignment effects are not expected for prices under the law of one price or for product variety once control retailers also adopt collective purchasing. To address this limitation, we use alternative estimands for these two outcomes. For prices, we estimate at the market–good level and exploit differences in baseline treatment exposure across goods within a market. This is a design-based shift-share strategy (Borusyak et al., 2022): goods more exposed to treated retailers at baseline should settle at lower long-run prices than less-exposed instances of the same good. For variety, we conduct a placebo test using an assortment measure that is comparable across waves.

The absence of detectable changes in log wholesale cost need not imply that collective purchasing had no lasting effect, because cost reductions are a contemporaneous margin that can disappear as the practice diffuses and prices converge. Wholesale quantity, in contrast, captures scale accumulated through earlier advantages: retailers that expanded sales and retained customers continue to purchase more even after unit-cost differences vanish. We therefore interpret the positive long-run quantity estimates, 0.258 log points for the core bundle and 0.298 for the full assortment, as consistent with an increase in market share built up over time, rather than evidence of a persistent wholesale discount.

Wholesale Quantity. The long-run wholesale results point to a persistent expansion in purchasing scale rather than a lasting reduction in unit costs. Retailers assigned to collective purchasing source substantially larger quantities one year later, while wholesale costs are statistically indistinguishable from those of the comparison group. The magnitude of the quantity effect is economically meaningful, around 30% in the reduced form, and is consistent with retailers retaining gains in sales or market share accumulated while purchasing jointly. By contrast, any initial cost advantage may have dissipated as collective purchasing became more common. These results could also reflect the diffusion of information about which wholesalers offer better prices or terms, reducing the persistence of cost differences across retailers.

Table 8: Long-Run Effects on Wholesale Quantity: DiD and ANCOVA

	Log Wholesale Quantity	
	Core	Full
DiD: Treat × Long-Term	0.264	0.290
	(0.148)	(0.109)
	{0.076}	{0.008}
Observations	2368	5778
ANCOVA: Treat, baseline-adjusted	0.146	0.190
	(0.108)	(0.092)
	{0.175}	{0.039}
Observations	1315	1735
Control mean	2.605	2.528
Fixed Effects	Mkt × Good	Mkt × Good
Clustering Lvl.	{Retailer}	{Retailer}

Source: Retailer panel survey, baseline and 12-month follow-up (July 2024).

Notes: The outcome is log wholesale quantity purchased (kg), on the core bundle (Core) and on the full assortment (Full). The unit of observation is retailer × good × survey visit, with the baseline visit as the pre-period. The DiD row is the coefficient on the interaction between treatment assignment and the long-run wave. The ANCOVA row is the coefficient on assignment in the long-run visits alone, conditioning on the baseline value of the same retailer × good cell; the estimated coefficient on that baseline value is 0.35 in both columns. The two rows are not estimated on the same sample: the ANCOVA drops every cell first observed at long term, which is why observation counts are reported under each row. All specifications include market × good and visit fixed effects and are weighted by the per-visit inverse-probability weight. Standard errors are clustered at the retailer level and reported in parentheses; p-values are in {curly brackets}. The control mean is the baseline control mean of the outcome.

Table 8 reports a positive long-run effect on the quantity sourced at wholesale. In the difference-in-differences estimates, log wholesale quantity rises by 0.264 on the core bundle ($p = 0.076$) and by 0.290 on the full assortment ($p = 0.008$), increases of roughly 30% and 34% against a baseline control mean of 2.5 log kilograms. The ANCOVA estimates are smaller and less precise, 0.146 ($p = 0.175$) and 0.190 ($p = 0.039$), or roughly 16% and 21%. The two designs differ because wholesale quantity is only weakly persistent at this horizon: the estimated coefficient on the baseline value is 0.35, far below the unit lag the difference-in-differences imposes, so part of the difference-in-differences estimate reflects mean reversion from a lower treated baseline rather than a treatment effect. Treated retailers did purchase less before the intervention, by 0.22 log points on the full assortment. The ANCOVA is the more conservative reading and it still returns a positive effect on the full assortment, but the two rows bracket the magnitude rather than agreeing on it, and the core-bundle effect is not distinguishable from

zero under either design. Wholesale costs, estimated on the same panel and reported in Table A.9, are a null under both designs: the difference-in-differences estimates are imprecise and change sign across the two samples, while the ANCOVA estimates are −0.000 and −0.005 with standard errors roughly half as large, ruling out cost changes above about five percent in either direction. Taken together, these estimates suggest that any lasting effect of collective purchasing operates through accumulated scale or market share rather than a persistent wholesale discount.

Prices. Markets with greater baseline exposure to treated retailers experience larger price declines twelve months later, consistent with lower prices where cooperative purchasing becomes more prevalent. In Table 9, cells more exposed to treated retailers at baseline have higher take-up saturation in the long-term, with a first-stage estimate of 0.237 ($p = 0.056$). The corresponding reduced-form effect on prices is −0.242 ($p = 0.007$), indicating that greater exposure to treatment is associated with lower long-run prices. To put this magnitude in context, a 10 percentage-point increase in treated exposure is associated with a 0.024 log-point decline in prices, or approximately 2.4%; a one-standard-deviation increase in residualized exposure (0.111) corresponds to a price decline of about 2.7%. Scaling the reduced form by realized take-up saturation gives an IV estimate of −1.020 ($p = 0.062$). Concretely, a 10 percentage-point increase in take-up saturation implies a 0.102 log-point decline in prices, or approximately 9.7%.

Table 9: Long-Run Price Effects: Design-Based Shift-Share Estimates

	Take-up satur.	Δln price	Δln price
Treated exposure (exog. shift)	0.237	-0.214	
	(0.111)	(0.104)	
	{0.056}	{0.065}	
Take-up saturation			-0.902
			(0.489)
			{0.092}
Good × market cells	126	126	126
Exposure resid. sd	0.111		
First-stage F (cl.)	4.57		
Weak-ID F (KP)			4.61
RI p (reduced form)		0.330	

Source: Retailer panel survey, baseline and 12-month follow-up (July 2025).

Notes: The unit is a good × market cell. *Treated exposure* is the baseline share of a cell's retailers assigned to treatment, the exogenous shift in the design-based shift-share strategy. Column 1 is the first stage, with long-run take-up saturation in the cell as the outcome. Column 2 is the reduced form, with the long-run change in the cell's log price as the outcome. Column 3 instruments take-up saturation with treated exposure. *Exposure resid. sd* is the standard deviation of exposure after partialling out the design controls. *First-stage F (cl.)* is the clustered first-stage F statistic and *Weak-ID F (KP)* is the Kleibergen–Paap weak-identification statistic. *RI p* is the randomization-inference p-value for the reduced form, drawn from the assignment mechanism. Standard errors are in parentheses; p-values are in {curly brackets}.

In the context of the experiment, these estimates are directionally consistent with the medium-run evidence that greater cooperation among sellers lowers consumer prices, although the IV magnitude should be interpreted cautiously. Take-up saturation varies substantially across cells: the average cell has a saturation level of 0.350, with an interquartile range from 0.150 to 0.545, while most cells lie between 0.25 and 0.75 saturation and only a small fraction are close to full take-up.

For a cell in which take-up saturation increases from 0.30 to 0.50, a 20 percentage-point increase well within the observed distribution, the IV estimate implies approximately an 18% reduction in prices. This large implied effect underscores the uncertainty surrounding the magnitude: the first stage is weak, with a clustered F-statistic of 4.57 and a Kleibergen–Paap statistic of 4.61, and randomization inference for the reduced form gives $p = 0.283$. We therefore interpret the endline results primarily as suggestive evidence that markets with greater realized cooperation continue to exhibit lower prices, rather than as providing a precise estimate of how much prices fall with additional take-up.

Variety. To unpack the cross-sectional results described above, we use a Jaccard linear index that normalizes bundle size and can be computed accross waves to distinguish between baseline assortment and convergence of bundles. For any two retailers i and j, let G_i be the set of goods they sell. We measure overlap as:

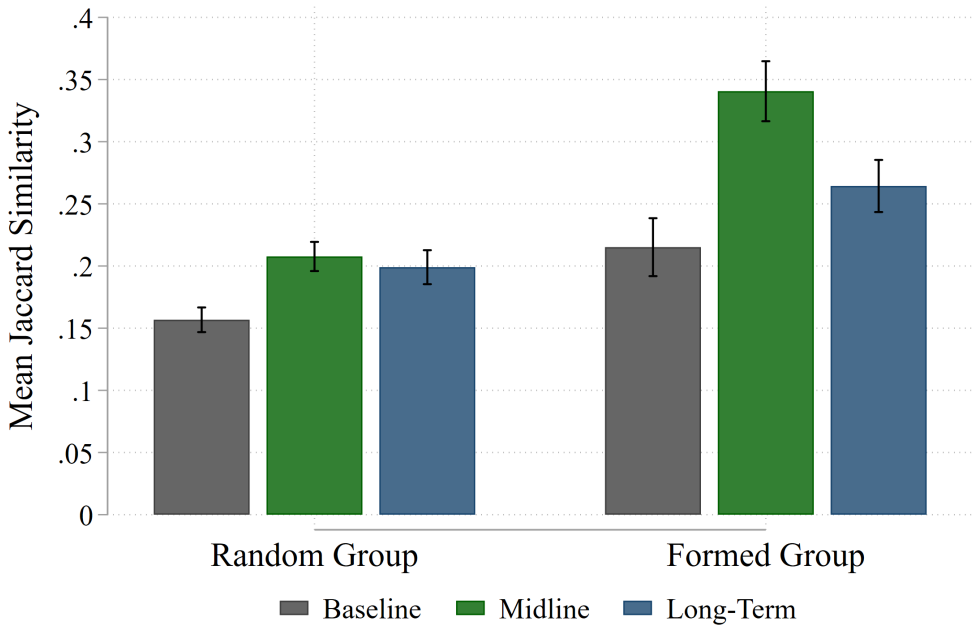
$$\sum^{B} J^{t}(i,j), \quad \forall j \in \{\text{Group of i}\}, \qquad \text{where } J(i,j) = \frac{|G_i \cap G_j|}{|G_i \cup G_j|}. \tag{4}$$

We then compute the mean Jaccard similarity within observed CP groups and compare it to a placebo benchmark. We also study the distribution of changes in similarity from baseline to midline, $J^{\text{mid}} - J^{\text{base}}$, to separate initial sorting from copying/learning.

Fig. 2 provides evidence for initial sorting but also high convergence at midline. First, retailers who later form a CP group sell more similar assortments than random placebo groups (0.21 vs. 0.16 mean Jaccard similarity; Figure 2). This initial sorting is consistent with retailers choosing partners who already buy similar goods, which helps the group capture economies of scale. Second, the similarity among formed groups rises sharply at midline, to 0.34, a 61% increase, while random groups rise to 0.21, a 32% increase.

To isolate the channel behind this rise, Fig. A.5 plots the kernel density of the change in the *numerator* of the Jaccard ratio, the average count of goods shared by within-group pairs. The placebo-group distribution peaks just above zero, at +0.3, with a median of +0.7 shared goods, whereas the formed-group distribution shifts right to a broad mode between +1 and +2.5, with a median of +2.2. The mean Jaccard index rise is therefore driven by within-pair *additions* of common goods: after forming a group, members start selling goods that other members previously sold.

Finally, in the long term, the overlap among groups active in July 2025 remains significantly higher, at 0.26, than that of randomly formed groups, at 0.20. This sustained overlap cannot be explained by higher entry or turnover rates. These results suggest that retailers first overcorrected their bundle decisions during the experiment but continue to coordinate on the goods they source in order to sustain the value of group formation.



Source: Retailer panel survey, baseline, midline, and 12-month follow-up.

Notes: Each bar is the mean pairwise Jaccard similarity of assortments among members of the same group, computed within each group and then averaged. *Formed Group* uses the observed CP groups; *Random Group* uses the placebo partition of retailers into groups of three to four drawn within their own marketplace. Bars are shown for the three survey waves. Vertical lines are 95% confidence intervals.

Figure 2: Mean Jaccard Similarity Within Groups, by Wave and Group Type

In the model, this combination of sorting and copying implies improved access to profitable products. Sorting increases initial overlap and therefore strengthens economies of scale in procurement, while copying allows members to access new goods at lower effective costs, both through pooled quantities (bulk terms) and through access to wholesalers (lower prices from repeated transactions).

5 A Model of Collective Procurement and Retail Competition

The last part of the paper asks what the broader welfare impacts of these cooperatives are for retailers and consumers. The experiment in Section 3 shows that procurement cooperation lowers prices and wholesale costs and broadens assortment, with the newly stocked goods concentrated among products a retailer's group members already sold. But the reduced form does not tell us why we observe those outcomes.

We therefore write a two-stage model of weekly retail competition. In Stage 1, each retailer decides which goods to stock, facing a setup cost that shifts with supplier access. In Stage 2, active retailers compete in quantities, with marginal costs that move along a wholesale quantity schedule with elasticity ψ.

We use an augmented version of Stage 2 to work through two hypotheses: (1) firms in a group choose quantities non-cooperatively; (2) firms in a group choose quantities collusively (i.e., they choose the total quantity that maximizes group profits). Section 7 shows that the data reject the joint maximization assumption.

We then use the full model to unpack the three channels through which group formation changes the cost structure. This distinction is important for understanding whether there is a dominant channel. First, we measure the *volume channel*: retailers who pool purchases unlock bulk discounts. Second, we define the *access channel*: a supplier contact or experience held by one group member becomes available to her co-members, lowering the good-specific fixed cost of stocking a good she did not previously carry. Finally, we compute the *transport channel*: retailers who send only one member can share the transportation cost.

5.1 Environment and timing

Market structure. A retailer is indexed by i, a market by m, a good by g, and a survey period by $t \in \{0, 1, 2\}$, corresponding to baseline and two midline weeks.[20] The model treats these observations as repeated static games. There is a set of goods $\mathcal{G}_m$ sold by retailers who operate from a single market m in period t. Each good $g \in \mathcal{G}_m$ is homogeneous within the same market m. Retailer i is always present in market m, but makes a good-specific stocking decision $a_{imgt} \in \{0, 1\}$. Competition takes place within a market–good–period cell (m, g, t). Conditional on activation, we observe selling price p_{mgt}, wholesale cost p^{wh}_{imgt}, and quantity q_{imgt}. We also observe retailer i's group membership.

Incumbency. Retailer i's observable characteristic is whether or not she is an incumbent in good g. Let $I_{img,0}$ be the binary variable equal to 1 if she regularly stocks it or has a supplier contact. Incumbency is a time-invariant characteristic that is common knowledge. As a result, if a retailer is an

[20] Both midline surveys occur after information was revealed in the *Information* arm and after the meetings were organized in the other two treatment arms.

incumbent in a good g, we assume that she stock this good deterministicly and with probability equal to 1.

Timing. We consider two environments: one without the cooperation program and one with it. In each environment, the stocking decision and quantity competition unfold as a two-stage game in every market–good pair and period. In the first stage, potential entrants decide whether to stock the good based on whether expected variable profit exceeds the setup cost. Firms observe common demand and the predetermined characteristics of all potential entrants. Each retailer also knows the predictable component of her own procurement cost and her private setup cost, but not her rivals' cost realizations. After stocking decisions are made, cost shocks are realized and active retailers compete in quantities.

Demand. A representative consumer in each market has CES preferences over goods and does not differentiate among retailers: units of the same good are perfect substitutes regardless of which retailer supplies them. Demand is therefore isoelastic at the good level,

$$Q_{mgt} = A_{mgt}\, p_{mgt}^{-\eta}, \qquad Q_{mgt} = \sum_{i} q_{imgt}, \tag{5}$$

where $\eta > 1$ is the elasticity of substitution across goods, q_{imgt} is the quantity retailer i sells of good g in market m at time t, p_{mgt} is the common cell price, and A_{mgt} is a market–good demand shifter that absorbs total expenditure and the prices of other goods. Retailers take A_{mgt} as given.

Supply. We model retailers' stocking decision and quantity decisions for each good in each market as a two-stage game. In the first stage, the potential non-incumbent retailers of good g in market m simultaneously decide whether to include the good in their baskets, incurring a setup cost S_{imgt} if they do. All retailers operating in market m can potentially stock any good g. These decisions determine the realized stocking set $\mathcal{R}_{mgt} = \{j : a_{jmgt} = 1\}$. In the second stage, retailers in this set observe their realized wholesale costs and compete in quantities. Each retailer i faces constant marginal cost c_{imgt}; because units of the same good are perfect substitutes, retailers compete only through quantities and costs. Conditional on a realized market structure and realized wholesale costs, the variable profit of

retailer i is

$$\tilde{\pi}_{imgt}(\mathcal{R}_{mgt}) = \max_{q_{imgt}} \left(p_{mgt}(q_{imgt} + Q_{-i,mgt}) - c_{imgt} \right) q_{imgt}, \qquad p_{mgt}(Q) = \left(A_{mgt}/Q \right)^{1/\eta}, \tag{6}$$

where $Q_{-i,mgt}$ is the total quantity supplied by the other active retailers and $p_{mgt}(\cdot)$ is the inverse of the CES demand curve in Eq. (5). Because marginal costs differ across retailers, the Cournot solution can be at a corner: ordering the stocking set from cheapest, only the $K^* = \max\{K : c_{(K)} < p(K)\}$ lowest-cost retailers sell positive quantities, and the remainder are priced out.

Information. When deciding whether to stock a good, a retailer observes common demand, the predictable component of her own procurement cost, and her own setup-cost shock. We write the marginal cost realized before quantity competition as

$$c_{imgt} = c^{reg}_{imgt} \exp(\omega_{imgt}),$$

where c^{reg}_{imgt} is the predictable component and ω_{imgt} is the cost shock realized after baskets are fixed. Each rival's cost belief is centered on a retailer-specific predictable component constructed from dated information, principally that rival's baseline procurement history, the lagged market–good cost mean, and good effects. The retailer does not observe rivals' subsequent cost shocks, but knows their conditional distributions and forms rival-specific stocking probabilities. Expected profits therefore average jointly over which rivals stock and the costs they subsequently realize.

Groups and exposure. A procurement source is either a purchasing group or a singleton retailer. The groups form only after the treatment is administered, $t \in \{1, 2\}$. For now, we assume that groups do not change competition but only the costs that enter the problem. First, groups allow retailers to act as a single source at the wholesale market such that the wholesale cost corresponds to the total quantity that all members purchase. Second, groups provide a discounted setup cost to their group members. A retailer faces a lower setup cost of stocking the good if one of her group members already has experience in sourcing this good.

5.2 Equilibrium

An equilibrium specifies stocking and quantity choices under the information structure above. Each retailer therefore evaluates the profit from stocking using joint beliefs about which rivals will stock and the procurement costs those rivals will face conditional on stocking.

The return to stocking depends on which rivals also stock the good and the procurement costs they face, since both determine the intensity of quantity competition. Before committing to the good, retailer i does not know which non-incumbent rivals will belong to $\mathcal{R}_{mgt}$ or the wholesale costs that members of this set will realize. A rival's identity therefore matters before stocking decisions through her stocking probability and predictable cost. Once the stocking set and cost vector are realized, however, identities have no additional role in the Cournot subgame.

Let $\mathbf{P}_{-i,mgt}$ collect the probabilities that her rivals stock the good. Conditional on the common market–good–period state, retailers' stocking shocks are independent across retailers, so $\mathbf{P}_{-i,mgt}$ determines the distribution of rival stocking configurations. Conditional on the pre-decision information set, the focal and rival post-commitment shocks are independent draws from the same G_ω. Group members therefore share the shock distribution, not the same numerical realization. Let ω_{mgt} denote a joint realization of rival stocking and all post-commitment wholesale-cost shocks, and let $\mathcal{R}_{mgt}(\omega_{mgt})$ and $\mathbf{c}_{mgt}(\omega_{mgt})$ denote the resulting stocking set and cost vector. We write $\tilde{\pi}_{imgt}(\mathcal{R}_{mgt}(\omega_{mgt}), \mathbf{c}_{mgt}(\omega_{mgt}))$ for retailer i's optimized Stage-2 profit in (6), making its dependence on realized costs explicit. Her expected variable profit conditional on stocking is

$$\bar{\pi}_{imgt}(\mathbf{P}_{-i,mgt}) = \mathbb{E}_{\omega_{mgt}}\left[\tilde{\pi}_{imgt}\big(\mathcal{R}_{mgt}(\omega_{mgt}), \mathbf{c}_{mgt}(\omega_{mgt})\big) \mid a_{imgt} = 1\right], \tag{7}$$

where the distribution of ω_{mgt} is induced by $\mathbf{P}_{-i,mgt}$ and the distribution of wholesale-cost shocks realized after commitment.

To construct $\tilde{\pi}_{imgt}(\mathcal{R}_{mgt}, c_{mgt})$ inside this expectation, we solve the Cournot subgame for each realized stocking set and cost vector, subject to $q_{imgt} \geq 0$. A retailer whose marginal profit at zero is non-positive optimally sells nothing, while retailers with positive quantities satisfy the usual Cournot first-order condition. Stocking therefore need not imply positive sales in every realization. The expected profit in (7) averages these constrained Cournot profits before the retailer decides whether to

stock.

Expected variable profit is the benefit from stocking. The setup cost S_{imgt} is the fixed, good-specific cost of making good g available for sale in period t before any units are sold. It captures activities such as finding or contacting a supplier, learning how to procure the good, and arranging the order. Unlike the marginal cost c_{imgt}, it does not vary with the quantity sold. An incumbent has already made these investments for good g, so we normalize her incremental deterministic setup cost to zero; a non-incumbent must incur a positive deterministic setup cost:

$$S_{imgt} = \begin{cases} 0, & \text{if } I_{img,0} = 1, \\ \exp\left(\alpha_{mg} + \gamma_{grp} G_{img} + \gamma_T D_{it} + \gamma_{inter}(D_{it} \times G_{img})\right), & \text{if } I_{img,0} = 0. \end{cases} \tag{8}$$

Here, α_{mg} captures market–good differences in setup costs, G_{img} indicates whether retailer i has a group member who is an incumbent seller of good g, and $D_{it} = T_i \times \text{Post}_t$; the interaction allows treatment to change setup costs specifically for group-exposed goods. Incumbents do not draw an stocking shock and stock with probability one by assumption. A non-incumbent faces the total setup cost $S_{imgt} + \tau\varepsilon_{imgt}$, where ε_{imgt} has the standard logistic distribution, and stocks when expected variable profit covers that cost. An equilibrium is a vector of stocking probabilities $\mathbf{P}^*_{mgt}$ satisfying

$$P^*_{imgt} = \begin{cases} 1, & I_{img,0} = 1, \\ \Lambda\left(\dfrac{\bar{\pi}_{imgt}(\mathbf{P}^*_{-i,mgt}) - S_{imgt}}{\tau}\right), & I_{img,0} = 0, \end{cases} \tag{9}$$

for every retailer in the market–good–period cell. The best-response map is continuous and maps the unit cube into itself, so an equilibrium exists. Strategic substitutability alone does not guarantee uniqueness, although the equilibrium is unique when competitive effects are sufficiently weak relative to the stocking-shock scale τ, which determines how sharply stocking responds to expected-profit differences. Cooperation does not change this decision problem or the form of competition; it can lower setup costs and, through pooled purchasing, the predictable component of marginal costs. These changes make stocking more profitable directly, but they also change competition because larger stock by one retailer lowers her rivals' expected profits. We therefore evaluate cooperation by solving (9) under environments $d \in \{0, 1\}$, without and with cooperation, rather than by changing each retailer's

probability while holding rivals fixed. The difference $\mathbf{P}^{*,1}_{mgt} - \mathbf{P}^{*,0}_{mgt}$ combines the direct cost effects with the indirect effects operating through rivals' equilibrium responses. Transport savings do not alter this equilibrium because retailer travel every period to stock at least one good. It enters the welfare calculation separately.

Pricing. Conditional on a realized stocking set and cost vector, the Cournot subgame determines a single price for the market–good cell. Let $\mathcal{K}_{mgt} \subseteq \mathcal{R}_{mgt}$ denote the retailers who sell a positive quantity and let $N_{mgt} = |\mathcal{K}_{mgt}|$. Summing their first-order conditions gives the pricing rule

$$p_{mgt} = \frac{\eta \sum_{k \in \mathcal{K}_{mgt}} c_{kmgt}}{\eta N_{mgt} - 1}. \tag{10}$$

Thus, the cell price depends on the number and costs of the retailers who sell positive quantities. The relevant set $\mathcal{K}_{mgt}$ is determined endogenously: it retains the low-cost retailers for whom the resulting price exceeds marginal cost.

5.3 Discussion of maintained assumptions

Five maintained assumptions delimit what the model can answer.

First, *incumbency*. Incumbents stock with probability one, so the model explains the addition of goods outside a retailer's established assortment rather than the exit of goods from her core assortment.

Second, *group composition*. The model conditions on the groups that formed and does not explain partner selection. It can evaluate how access and pooled orders operate within the observed group network, but it cannot evaluate a policy that changes who matches with whom.

Third, *downstream conduct*. The canonical model treats a good as homogeneous within a market and group members as separate Cournot competitors. It interprets residual within-cell price dispersion as transaction-level heterogeneity rather than persistent differentiation across stalls. The structural appendix relaxes this restriction with a nested-CES specification, while Section 7 tests whether group members behave as a single downstream firm.

Fourth, *a common stocking-shock scale*. The model uses one τ across retailers, goods, and markets.

The deterministic setup cost S_{imgt} may vary across retailers and market–good cells, but the response of stocking to a given expected-profit difference is common. This restriction makes profits and setup costs comparable in FCFA across cells; Section 6 reports how strongly the data discipline that common scale.

Fifth, *pre-trip commitment and post-commitment cost shocks*. A retailer observes the predictable component of her own procurement cost when deciding whether to stock. After commitment but before Cournot competition, the shock ω_{imgt} is realized and determines her marginal cost c_{imgt}. Her beliefs about each rival are centered on that rival's own dated procurement history, with independent draws from the same post-commitment shock distribution. This timing prevents realized cost shocks from selecting the stocking set while preserving selection on predictable rival costs.

Finally, the Stage-1 and Stage-2 extensive margins are conceptually distinct. The stocking indicator records whether a retailer stocks a good and pays its setup cost. Conditional on stocking, the retailer may choose zero quantity in some realized rival-stocking and cost configurations. This distinction is what allows a retailer to have positive expected profit when stocking even though she is priced out in some configurations.

6 Estimation

This section maps the model's primitives into observables and estimates them in three steps. First, we estimate the demand elasticity η, which converts observed prices and quantities into marginal costs and demand shifters. Second, we estimate the wholesale cost–quantity elasticity ψ, which measures the volume channel and provides cost estimates for retailer–good pairs that are not observed trading. Third, we estimate the activation-shock scale τ and the setup-cost coefficients, including the treatment-by-group-exposure interaction γ_{inter} that measures the access channel. Each subsection introduces its estimating sample and notation before presenting the corresponding equation.

6.1 Demand

Specification. For each cell (m, g, t), let $Q_{mgt} = \sum_i q_{imgt}$ be total retail quantity and let $p_{mgt} = \sum_i p_{imgt} q_{imgt} / Q_{mgt}$ be the cell-level retail price, constructed as the quantity-weighted mean of retailer

prices. Taking logs of the demand curve in (5), we estimate

$$\log Q_{mgt} = \mathrm{FE}_{mg} + \mathrm{FE}_t - \eta \log p_{mgt} + \varepsilon_{mgt}, \tag{11}$$

where FE_{mg} is an interacted market–good fixed effect, FE_t is a survey-period fixed effect, and ε_{mgt} is a cell-level demand shock. The market–good fixed effect absorbs the cell's persistent demand level, while the period effects absorb common shifts. Thus, η is identified from changes in a good's price and total quantity within the same market–good pair over the three survey periods. The estimation sample requires positive price and quantity and at least two usable periods for a market–good pair.

The cell price is endogenous because positive demand shocks can raise both price and quantity. We instrument $\log p_{mgt}$ with the log of the wholesale unit price paid by active retailers in the cell. This is a supply shifter: its exclusion requires wholesale-price movements, conditional on market–good and period fixed effects, to affect total retail quantity only through the retail price.

Why constant elasticity. This specification matches the demand curve used in the Cournot model and allows one across-product elasticity to govern the marginal-cost inversion and demand-shifter recovery below. With only three survey periods, richer patterns of cross-price elasticities among goods are hard to identify.

Results and construction of model inputs. The IV point estimate is $\hat{\eta} = 1.1455$ (market-clustered s.e. 0.4934), using 793 market–good–period cells, 275 market–good pairs, and 12 markets. A one-percent increase in a good's cell price lowers total cell quantity by about 1.15 percent. The first-stage coefficient on the wholesale-price shifter is 0.2135 (s.e. 0.0937), for a cluster-robust F statistic of 5.20. Because the instrument is weak by conventional standards, the Anderson–Rubin 95% confidence set, $[0.55, 7.07]$, is the appropriate uncertainty statement. A market-level wild-cluster bootstrap gives a median of 1.1552 and a 95% bootstrap percentile interval of $[0.5674, 5.1655]$.

We use the full-precision point estimate of η to construct the two inputs required by the supply and activation stages. In each market–good–period cell, define retailer i's quantity share as $s_{imgt} = q_{imgt}/Q_{mgt}$. The homogeneous-good Cournot first-order condition, its marginal-cost inversion, and the implied de-

mand shifter are

$$\frac{p_{mgt} - c_{imgt}}{p_{mgt}} = \frac{s_{imgt}}{\eta}, \tag{12}$$

$$\hat{c}_{imgt} = p_{mgt}\left(1 - \frac{s_{imgt}}{\eta}\right), \tag{13}$$

$$\hat{A}_{mgt} = Q_{mgt} p_{mgt}^{\eta}. \tag{14}$$

Here c_{imgt} is retailer i's marginal cost, $\hat{c}_{imgt}$ is its value recovered from observed prices and quantities, and $\hat{A}_{mgt}$ is the demand shifter recovered from the market demand curve. The same estimate $\hat{\eta} = 1.1455$ is used in all three objects.

Two qualifications matter for interpreting the downstream estimates. First, the Anderson–Rubin set includes values at or below one, even though $\eta > 1$ is required for finite consumer-surplus levels under isoelastic demand; welfare levels are therefore sensitive to demand uncertainty. Second, the downstream estimates condition on the point estimate of η: its sampling uncertainty is not propagated into $\hat{\psi}$, $\hat{\tau}$, or the setup-cost coefficients. Appendix J considers a nested-CES alternative and jointly estimates its across-product elasticity η and across-stall elasticity σ.

6.2 The cost channel: wholesale discount elasticity

Wholesale schedule and estimating equation. Recall that a procurement source s is either a singleton retailer or a purchasing group; we define M_s as the set of its member retailers. If q_{imgt}^{wh} is retailer i's wholesale quantity, the order presented to the wholesaler is

$$Q_{smgt}^{wh} = \sum_{i \in M_s} q_{imgt}^{wh}.$$

We parameterize the source-level marginal cost as a constant-elasticity schedule,

$$c_{smgt}^{s} = c_{smg}^{0}\left(Q_{smgt}^{wh}\right)^{\psi}, \qquad \psi = \frac{\partial \log c_{smgt}^{s}}{\partial \log Q_{smgt}^{wh}}, \tag{15}$$

where c_{smg}^{0} is the source–market–good cost level and a bulk discount implies $\psi < 0$.

To model the interaction in the wholesale market, we estimate ψ on the source×market×good×week

panel.[21] This specification reflects the fact that a singleton retailer continues to act alone, but groups pooling their purchase orders act as one procurement source. Let D_s indicate that source s belongs to a treatment arm and let Post_t indicate a post-treatment survey week. The instrument is their interaction, $D_s \times \text{Post}_t$:

$$\log \hat{c}^{s}_{smgt} = \psi \log Q^{wh}_{smgt} + \phi_{Post}\,\text{Post}_t + \phi_{mg} + \phi_s + u_{smgt}, \tag{16}$$
$$\log Q^{wh}_{smgt} \leftarrow D_s \times \text{Post}_t \qquad \text{(DiD-IV first stage)}.$$

Here $\hat{c}^{s}_{smgt}$ is the source-level counterpart of the retailer-level marginal cost recovered in (13), ϕ_{mg} and ϕ_s are market–good and source fixed effects, ϕ_{Post} absorbs the common post-treatment change, and u_{smgt} is the residual wholesale-cost shock realized after stocking decisions are made. To construct $\hat{c}^{s}_{smgt}$, we first invert marginal cost retailer by retailer, using each retailer's own share, and then average those recovered costs within source s.

Identification. We identify ψ from the post-treatment increase in pooled wholesale quantity induced by treatment. Pooled quantity is otherwise endogenous because lower-cost sources may choose larger orders. We therefore use $D_s \times \text{Post}_t$ as an instrument for Q^{wh}_{smgt}: treated sources begin pooling their orders only after treatment. The sample is restricted to sources whose pre-existing wholesale contract access is unchanged by group formation, preventing the instrument from mechanically switching the relevant supplier contract. The exclusion restriction is that, within this restricted sample, treatment changes source-level marginal cost only by moving the source along the wholesale quantity schedule. Treatment must therefore not change bargaining power or supplier terms at a given pooled quantity or affect any component of retail marginal cost other than the wholesale unit cost generated by that quantity. The resulting $\hat{\psi}$ is a design-based IV elasticity that the structural cost imputation below treats as an input.

[21] The source-level collapse does not drive ψ. As a check, we re-estimate the module on the retailer×good×week panel, holding the pooled-quantity regressor fixed by construction, so that each retailer carries her own randomized treatment assignment, her own attrition weight, and her own purchase price. The elasticity the model uses is unchanged: ψ on inverted marginal cost is −0.052 (s.e. 0.057) at the retailer level compared with −0.0697 at the source level (no-switch sample; 5,483 retailer×good×week rows and 721 retailers vs. 4,838 source cells). The retailer-level estimate is less precise because a retailer's own treatment moves the shared group quantity only partially (first-stage coefficient 0.34, F = 12.5, vs. 0.60, F = 41).

Results. The canonical estimate is $\hat{\psi} = -0.0697$ (s.e. 0.0367). Under the constant-elasticity schedule, doubling pooled wholesale quantity lowers marginal cost by $\simeq 4.7$ percent. Its sign and magnitude are consistent with the experimental wholesale-cost decline. Treatment raises log pooled quantity by 0.603, so $\hat{\psi}$ implies a reduced-form marginal-cost decline of $(-0.0697)(0.603) = -0.042$ log points, or about 4.1 percent. Assuming full pass-through and a fixed market structure, this is close in sign and order of magnitude to the -0.038 retail-price effect in Section 3.

6.3 The stocking stage

At this point, we have estimated demand and the wholesale quantity discount; we now estimate the good-specific setup cost from the activation decisions retailers make in each period. As discussed above, equilibrium activation is described by a fixed point in Eq. (9). Each retailer's probability of stocking a good depends on the equilibrium activation probabilities of her potential rivals. Imposing this restriction for every candidate parameter vector would require a computationally intensive nested-fixed-point likelihood estimator (Seim, 2006). We therefore adopt the two-step logic used to estimate static games of incomplete information (Sweeting, 2009; Bajari et al., 2010).

In the first step, we estimate activation probabilities directly from the data. These probabilities summarize retailers' beliefs about which rivals are likely to stock the same good (P^{CCP}_{jmgt}). We combine them with the estimated distribution of wholesale-cost shocks to construct each retailer's expected variable profit from stocking: possible rivals and their realized costs determine the market price, quantities, and ultimately the focal retailer's profit.

In the second step, we compare these expected profits with observed stocking decisions to estimate the setup cost and the scale of the private activation shock. Substituting our setup-cost specification in Eq. (8) into the model's activation rule gives the probability that a non-incumbent stocks the good:

(17)

$$\Pr\left(a_{imgt} = 1 \mid I_{img,0} = 0\right) = \Lambda\left[\frac{1}{\tau}\left(\widehat{\tilde{\pi}}_{imgt}\left(\widehat{\mathbf{P}}^{CCP}_{-i,mgt}\right) - \underbrace{\exp\left(\alpha_{mg} + \gamma_{grp}G_{img} + \gamma_{T}D_{it} + \gamma_{inter}\left(D_{it} \times G_{img}\right)\right)}_{S_{imgt}}\right)\right].$$

Here, G_{img} indicates whether retailer i has a group member who is an incumbent seller of good g,

and $D_{it} = T_i \times \text{Post}_t$. We normalize incumbents' incremental setup cost to zero and estimate (17) on non-incumbents. The main causal parameter is γ_{inter}, while τ determines how strongly stocking responds to differences between expected profit and setup cost.

6.3.1 Variable Profit

To estimate the activation margin, we construct each retailer's expected variable profit by integrating over possible activation sets and wholesale-cost realizations through simulation. For each simulated realization, the Cournot solver identifies the subset $\mathcal{K}_{mgt}$ that sells a positive quantity and computes p_{mgt} from Eq. (10). We then average the focal retailer's resulting profits across realizations.

To estimate rival activation beliefs, we fit a reduced-form logit separately by period. The logit includes the rival's baseline incumbency status, treatment assignment, baseline assortment breadth, the number of incumbent rival sellers in the market–good cell, and the co-member rival share. Conditional on these covariates, we assume that retailers' private activation shocks are independent, as specified in Eq. (7). Thus, in each simulation draw ω_{mgt}, rival j activates independently with her fitted probability P^{CCP}_{jmgt}. Conditional on activation by the focal retailer i, the resulting set $\mathcal{R}_{mgt}(\omega_{mgt})$ includes i and every rival whose simulated activation indicator equals one.

Results. We estimate an activation-shock scale of $\hat{\tau} = 6{,}536$ FCFA (s.e. 1,410). Given that the median weekly profit is 14,000 FCFA ($\simeq$ \$30), this result indicates that activation responds gradually rather than sharply to small profit differences. For the setup cost, the access estimate is $\hat{\gamma}_{inter} = -0.2267$ (s.e. 0.0650; likelihood-ratio $p = 0.0005$). Because setup cost enters in logs, this estimate implies that treatment lowers the setup cost of a group-exposed good by $1 - e^{-0.2267} \simeq 20$ percent. The median predicted setup cost is 15,180 FCFA, so the estimated access effect corresponds to a reduction of approximately 3,079 FCFA for the median non-incumbent retailer–good pair. Because simulated expected profit already incorporates predictable procurement-cost savings, this reduction captures the additional benefit of easier access to a good through an experienced group member rather than the wholesale-cost channel or the total effect of treatment.

7 Competition or Collusion?

A group that meets to coordinate on sourcing could in principle also coordinate on selling. While the intervention is unlikely to create a market-wide cartel given the number of players, even after group formation, the collusion it could plausibly induce would involve co-members of one group internalizing one another's quantity choices. That is the hypothesis this section tests.

The exercise compares how retailers behave when other members of their sourcing group sell the same good in the same market. If group members continue to compete independently, each retailer should respond to a group peer just as she would to any other rival. If they instead coordinate their sales, they should behave as though they control some or all of their peers' sales as well as their own. We use this contrast to assess where observed behavior lies between independent competition and the strongest group-collusion benchmark, in which the group acts as a single seller.

A conduct parameter for the group. To do so, we apply the standard internalization-parameter test (Miller and Weinberg, 2017) to an experimentally induced coalition and validate it against a pre-treatment formation placebo.

First, we redefine the market share associated with each seller. Let s_{imgt}^{peer} denote the combined within-cell share of retailer i's active sourcing-group peers, and let λ index how much of that share she internalizes when choosing quantity. Under $\lambda = 0$, each retailer is an individual Cournot competitor and treats group peers like any other rivals. Under $\lambda = 1$, the group prices as a single firm owning the combined share, the strongest collusion the intervention could mechanically produce. Intermediate values represent partial internalization.[22]

Under homogeneous CES demand, individual Cournot conduct implies the inverted marginal cost

[22] Following the model-implied-margin comparison in Nevo (2001), one could instead compare conduct models using the level of the observed gross margin. That comparison has little power here: individual Cournot and full within-group internalization imply average gross margins of approximately 4.9 and 7.5 percent, respectively. Because these predictions are too close for the observed margin to separate reliably, we identify λ from how the inferred margin varies with co-members' shares within retailer–good pairs.

$\widehat{c}^{C}_{imgt}$. Allowing retailer i to internalize a fraction λ of her peers' share gives

$$\widehat{c}^{C}_{imgt} \equiv p_{mgt}\left(1 - \frac{s_{imgt}}{\eta}\right), \tag{18}$$

$$c_{imgt}(\lambda) = p_{mgt}\left(1 - \frac{s_{imgt} + \lambda s^{\text{peer}}_{imgt}}{\eta}\right) = \widehat{c}^{C}_{imgt}\left(1 - \lambda x_{imgt}\right), \qquad x_{imgt} \equiv \frac{s^{\text{peer}}_{imgt}}{\eta - s_{imgt}}. \tag{19}$$

We write marginal cost as observed procurement cost times a time-invariant retailer–good wedge, $c_{imgt}(\lambda) = a_{ig} w_{imgt}$. Taking logs of (19) produces the estimating equation

$$\log\left(\frac{\widehat{c}^{C}_{imgt}}{w_{imgt}}\right) = \alpha_{ig} + \delta_t - \log\left(1 - \lambda x_{imgt}\right) + \varepsilon_{imgt}, \tag{20}$$

where $\alpha_{ig} \equiv \log a_{ig}$ and δ_t absorbs common week shocks. We estimate (20) directly by nonlinear least squares.[23] For the alternative demand systems in the appendix, the same derivation replaces x_{imgt} by $\chi_{imgt} = \ell_1 s^{\text{peer}}_{imgt} / (1 - \ell_0 - \ell_1 s_{imgt})$, where the affine Lerner index is $L(s) = \ell_0 + \ell_1 s$.[24]

[23] The estimator uses the exact nonlinear term. Since $-\log(1 - \lambda x) \neq \lambda x$, an OLS slope is not λ, and testing that slope against one is not a test of $\lambda = 1$.

[24] The homogeneous model has $(\ell_0, \ell_1) = (0, 1/\eta)$; the nested model has $(\ell_0, \ell_1) = (1/\sigma, 1/\eta - 1/\sigma)$.

Table 10: Conduct: sourcing groups do not price as one firm

	Full sample	No-switch
Panel A: joint nuisance-adjusted exact NLS		
Conduct parameter $\hat{\lambda}$ (joint adjustment)	−0.075 (0.152)	+0.074 (0.178)
95% confidence interval	[−0.373, 0.223]	[−0.275, 0.422]
p-value, $\lambda = 0$ (individual Cournot)	0.62	0.68
p-value, $\lambda = 1$ (group as one firm)	$< 10^{-11}$	$< 10^{-6}$
Baseline future-group formation placebo $\hat{\phi}_0$	+0.333 (0.191)	+0.452 (0.188)
$\hat{\lambda}$ imposing $\phi_0 = 0$ (robustness)	−0.232 (0.133)	−0.152 (0.157)
Demand elasticity	$\eta = 1.1455$	$\eta = 1.1455$
Lerner slope κ	0.873	0.873
Observations	6,181	4,851
Retailer × good pairs	2,528	1,903
Rows with a co-active group peer	2,154	1,497
Within-pair variance removed by joint fit (%)	0.24	0.28

Notes: Test of individual Cournot ($\lambda = 0$) against group-as-one-firm pricing ($\lambda = 1$) on the retailer×good×week panel of active sellers with observed wholesale prices, estimated jointly by exact nonlinear least squares with retailer×good and week fixed effects and IPCW weights; standard errors are clustered at the retailer, the randomization unit. The baseline formation placebo uses future-group co-membership at baseline, where no group was yet operating, and tests whether future peer exposure predicts the baseline cost wedge; it is a nuisance adjustment, not a baseline conduct estimate. The row imposing $\phi_0 = 0$ reports the strict structural-restriction robustness estimate. The full-sample column uses every active seller with an observed wholesale price. The no-switch column restricts to the retailer×good pairs whose own baseline supplier-contract status already equals the best in their sourcing group, so procurement cannot be re-routed through a co-member's contract. The conduct parameter λ is identified given the demand elasticity η, itself estimated under $\lambda = 0$.

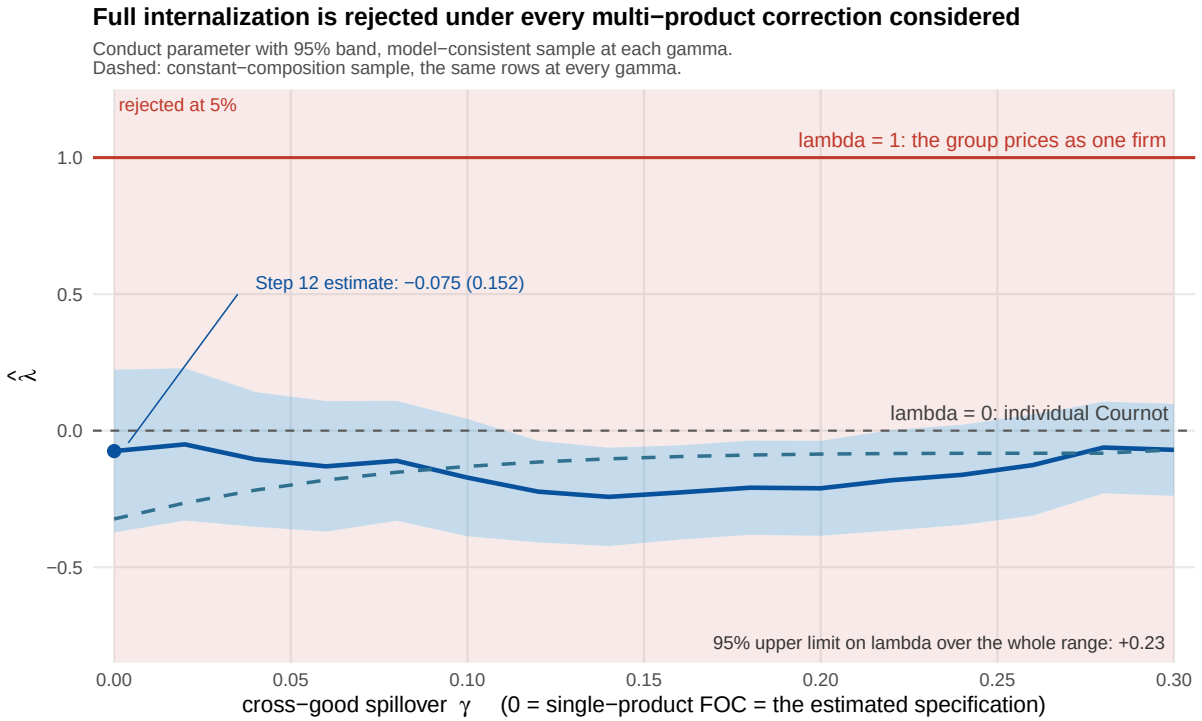
Table 10 shows that the estimate is $\hat{\lambda} = -0.075$ (s.e. 0.152). Its 95% confidence interval, $[-0.373, 0.223]$, includes zero but lies far below one. Individual Cournot therefore cannot be rejected ($p = 0.62$), while group-as-one-firm is rejected decisively ($p < 0.001$). The future-group-formation placebo is 0.333 (s.e. 0.191): it assigns the eventual group roster to the baseline period, when no group yet operated, and enters the joint estimator as a nuisance adjustment rather than as baseline conduct. Under homogeneous demand and multiplicative cost, groups coordinate procurement but do not price as one downstream firm.

The finding is not specific to the headline demand and cost specifications. Table E.1 re-estimates the same conduct model under alternative demand parameterizations. The across-stall specification

closely reproduces the headline estimate. Table E.2 varies the cost wedge, baseline adjustment, fixed effects, sample, clustering, and estimator. Across 23 variants, $\hat{\lambda} \in [-0.284, 0.491]$ and every specification rejects $\lambda = 1$ at 5%; only four reject $\lambda = 0$. The evidence is therefore sharp against full group internalization but less informative about small departures from individual Cournot.

A potential concern is that the wholesale-price gradient reflects changes in procurement cost, either through access to better contracts via co-members or through the bulk discount. We therefore restrict the sample to retailer–good pairs for which no supplier switch is available. The wholesale-price gradient is 0.262 (s.e. 0.116), compared with 0.239 (s.e. 0.123) in the full sample, while the joint conduct estimate remains close to zero, $\hat{\lambda} = 0.074$ (s.e. 0.178). Full internalization remains strongly rejected. Moreover, a wholesale-price reduction associated with greater co-member exposure raises $\log(\widehat{c}^{C}_{imgt}/w_{imgt})$ in the same direction as positive internalization. Procurement discounts therefore work against, rather than toward, rejecting $\lambda = 1$, making that test conservative with respect to this channel.

The baseline first-order condition treats inverse demand as block diagonal across goods, even though retailers sell multiple goods. We therefore introduce a cross-good spillover γ and re-estimate conduct over $\gamma \in [0, 0.30]$; $\gamma = 0$ reproduces the headline model. Positive values allow a retailer's quantity choice for one good to affect margins on her other goods. Because γ and λ are not separately identified, this is a bounded sensitivity exercise rather than an estimate of γ.



Notes: The solid line and band re-estimate λ on the model-admissible sample at each value of γ; the dashed line holds the sample fixed across the sweep. The horizontal lines mark individual Cournot ($\lambda = 0$) and group-as-one-firm pricing ($\lambda = 1$). Appendix E.2 reports the specification ladder and diagnostics behind the sweep.

Figure 3: Full internalization is rejected under multiproduct corrections

Figure 3 shows that the result survives throughout the range: the 95% upper confidence limit for λ never exceeds 0.23, and the largest p-value against $\lambda = 1$ is 6.2×10^{-10}. Holding the sample fixed moves the estimate at $\gamma = 0$ to -0.323 (s.e. 0.214), revealing some leverage sensitivity, but neither conduct conclusion changes. Omitted cross-good multiproduct effects therefore do not account for the rejection of full internalization.

Retailers who source together do not appear to internalize one another's output decisions, and the economically relevant group-as-one-firm benchmark is rejected. The appendix reaches the same conclusion using a cost-side wedge, pass-through under alternative conduct inversions, and simulated coalitions that assess the tests' power. Taken together, the results support the conduct assumption used in the model and welfare analysis while recognizing that the data are less informative about small departures from individual Cournot behavior.

8 Welfare and Counterfactuals

We use the estimated model to translate procurement cooperation into welfare. The reduced form shows that cooperation lowers prices and broadens assortment, but it cannot say how the resulting gains are split between consumers and retailers, or how much each owes to cheaper inventory on existing goods versus the entry of new goods versus shared transportation. The structural model answers both questions because it separates the three channels through which cooperation operates. The first is a *volume* channel: pooling orders moves a source down the wholesale schedule and lowers marginal cost. The second is a *supplier-access* channel: treatment lowers the Stage-1 setup cost S_{imgt} in (8) specifically for goods for which there is an experienced group member. Goods that were not worth stocking can therefore become active. The third is a *transportation* channel: by sending only one retailer to the wholesale market, the taxi fare is shared among group members.

This section quantifies the welfare gain, decomposes it across the three channels, and asks who benefits from it. Throughout the canonical welfare exercise, group members remain separate downstream competitors: we set the conduct parameter of Section 7 to $\lambda = 0$. The welfare results are therefore conditional on individual Cournot conduct.

We can then define the retailer welfare change as $\Delta W^{\text{retailer}} = \left(\mathbf{c}^{s,1}, \mathbf{S}^1, \boldsymbol{\mathcal{T}}^1\right) - \left(\mathbf{c}^{s,0}, \mathbf{S}^0, \boldsymbol{\mathcal{T}}^0\right)$. Given the separability of goods, the change in welfare is the sum of changes across goods for a given retailer; in particular, gains for incumbents and non-incumbents are different. For a retailer selling good g as a core good, who is considered an incumbent, the benefits are concentrated in the *volume* channel, while a non-incumbent can benefit from both the *volume* and *supplier-access* channels. Finally, transportation reductions apply to all retailers uniformly. We decompose the effect of group formation as follows:

$$
\begin{aligned}
\Delta W^{\text{retailer}} &\equiv E_{00} - E_{11} \\
&= \sum_{imgt: I_{img,0}=1} \left(\bar{\pi}^1_{imgt} - \bar{\pi}^0_{imgt}\right) \\
&\quad + \sum_{imgt: I_{img,0}=0} \tau \left\{ \log\left[1 + \exp\left(\frac{\bar{\pi}^1_{imgt} - S^1_{imgt}}{\tau}\right)\right] - \log\left[1 + \exp\left(\frac{\bar{\pi}^0_{imgt} - S^0_{imgt}}{\tau}\right)\right] \right\} \qquad (21) \\
&\quad + \sum_{it} \left(\mathcal{T}^0_{it} - \mathcal{T}^1_{it}\right). \qquad (22)
\end{aligned}
$$

Consumer welfare Under the canonical homogeneous-good demand system (5), the cell-level consumer surplus associated with price p_{mgt} is

$$CS_{mgt} = \frac{A_{mgt}}{\eta - 1} p_{mgt}^{1-\eta} = \frac{Q_{mgt}\, p_{mgt}}{\eta - 1}, \tag{23}$$

where A_{mgt} is the demand shifter and $Q_{mgt} = A_{mgt} p_{mgt}^{-\eta}$ is the cell quantity. Because activation and procurement quotes are stochastic, the counterfactual consumer-surplus object averages (23) over the realized stocking sets and cost draws generated by each equilibrium, rather than evaluating it only at an average price.

8.1 The gains from cooperation

Pooled procurement raises total surplus primarily by increasing retailer welfare, with volume and supplier access both contributing substantially to the overall gain. Table 11 shows that total surplus increases by about 1.338m FCFA, with 40.6% attributable to the volume channel and 59.4% to supplier access. Retailer inclusive value rises by 741.5k FCFA, accounting for 55.4% of the total gain, while consumer surplus increases by 596.9k FCFA. For retailers, 98.4% of the gain comes from supplier access and 1.6% from volume. For consumers, by contrast, 89.0% of the gain comes from the volume channel and 11.0% from supplier access. The extensive-margin response is almost entirely driven by supplier access: cooperation activates about 90.4 additional goods, with a 100.2% supplier-access contribution offset by a −0.2% volume contribution. Excluding saved setup costs, gross total surplus rises by 609.6k FCFA, with 89.1% of this gain coming from volume and 10.9% from supplier access. Shared transport generates an additional 30.7k FCFA in vehicle-fare savings, reported separately from the Shapley decomposition. The large difference between net-of-setup and gross surplus reflects the importance of interpreting the setup-cost shock $\tau\varepsilon$ as genuine heterogeneity in entry costs rather than decision error.

Calibration exercise are still in progress to provide a more meaningful interpretation of these results.

Table 11: Welfare Gain from Pooled Procurement, by Channel (Shapley)

Welfare object	Total gain (FCFA)	% of E_{00}	Volume	Supplier access
Net-of-setup basis (headline)				
Producer inclusive value W_{incl}	+741,500	+2.37%	1.6%	98.4%
Consumer surplus	+596,900	+0.39%	89.0%	11.0%
$CS = \sum Q_{mgt} p_{mgt} / (\eta - 1)$				
Total surplus $W_{\text{incl}} + CS$	**+1,338,400**	**+0.73%**	**40.6%**	**59.4%**
Gross basis (excludes the setup cost entrants save)				
Gross total surplus	+609,600	+0.39%	89.1%	10.9%
Trip-sharing add-on (accounting; outside the Shapley split)				
Shared vehicle fares (treatment-induced)	+30,700	–	–	–
Extensive margin				
Expected new goods activated	+90.4 goods	+2.28%	−0.2%	100.2%

Source: Estimated structural model of Section 6, four counterfactual equilibria. Notes: All figures are changes from the no-cooperation equilibrium E_{00} to full cooperation E_{11}, in FCFA, at the estimated demand elasticity and entry-cost scale. Each row's % of E_{00} is that change divided by the E_{00} level of the same object, so the percentages are not comparable across rows with different bases and do not sum. Volume and Supplier access are each row's Shapley shares and sum to 100 percent; a share outside $[0, 100]$ means the two channels offset. The net-of-setup rows include the setup cost entering retailers save; the gross row excludes it. The trip-sharing row is the treatment-induced saving in shared vehicle fares, restricted to the fare share of the per-trip cost; it is an accounting benchmark outside the Shapley split, has no E_{00} level to be a share of, and is not summed into the totals.

The incumbent margin: a saving competed away. The gross retailer-profit gain is concentrated on new goods (+296k FCFA). On goods they already sold, incumbents' variable profit falls by about 28k FCFA. Table 12 separates this change into three economically distinct lines. The direct procurement saving is large (+289k), but it is almost entirely competed away on existing goods (−284k): most of it reaches consumers as lower prices rather than staying with retailers. On top of that, supplier access brings new sellers into the cell; the expected number of rivals per incumbent cell rises by about 1.4, which steals a further 33k of incumbent business. When compared with baseline profits, this decrease is relatively small, in particular because no retailer sells only core goods that are always present. This diversification means that every retailer pays a small cost for participating in cooperative purchasing but benefits through fringe goods that earn them much more profit.

Table 12: Welfare Changes for Incumbents

	FCFA		
Incumbent variable-profit change	Total	Per retailer-day	% of E_{00}
(1) Direct cost benefit (procurement saving, holding competition fixed)	+289,000	+260	+10.8%
(2) Competition erodes it on existing goods (intensive margin)	−284,000	−255	−10.6%
(3) Entry business-stealing (new competitors, extensive margin)	−33,000	−29	−1.2%
Net change (general equilibrium)	**−28,000**	**−25**	**−1.0%**

Source: Estimated structural model of Section 6. Notes: Change in gross variable profit on incumbent (already-stocked) goods from E_{00} to E_{11}. Line (1) applies the cost saving holding competition fixed; lines (2) and (3) add the price response of existing rivals and the entry of new ones. Lines sum to the net change up to rounding. Basis: an observation is a retailer × good × survey week and carries a per-day profit flow. Total sums these over all 4,127 incumbent observations, so it is a daily rate aggregated over the sample rather than a per-retailer or whole-period amount. Per retailer-day divides that total by the 1,113 retailer × week cells observed after the intervention, which is the domain over which the change is defined (it is zero beforehand by construction). The last column expresses the total as a share of the E_{00} level of the same sum, 2,671,353 FCFA.

Line (2) is the strategic cost-reallocation effect that Bloch (1995) studies for cost-reducing groups, and line (3) is the business-stealing externality of entry in the spirit of Mankiw and Whinston (1986). Neither model is exactly ours, so we use them only as labels; what matters is that the entry channel, absent from a fixed-roster model, carries almost all of the incumbent loss.

8.2 Long run data and extrapolation exercise.

Write this section to show how the model perform in the long run given the diffusion of the treatment.

9 Conclusion

This paper studies a simple puzzle: why do small informal retailers operating side by side, facing common and repeated procurement costs, so rarely cooperate to exploit economies of scale? This paper studies that puzzle among fresh-food retailers in Bouaké, who operate side by side, source overlapping goods from the same wholesale market, and repeatedly incur procurement costs that could be shared.

By combining a nested randomized experiment with a retailer-by-product panel and a structural model, the paper connects the formation of cooperative arrangements to their consequences for costs, product scope, competition, and welfare. We show that this inefficiency reflects not a lack of potential gains, but the frictions of turning proximity into cooperation.

The randomized experiment testing three nested interventions to lower the cost of cooperating shows that reducing the social cost of initiating cooperation increases collective purchasing. A one-time facilitated meeting generates large and persistent increases in participation, while information alone has smaller effects and additional monitoring does not further raise take-up. The effects of the meeting-based interventions remained detectable twelve months later, even as collective sourcing spread to the control group. Cooperation also changed market outcomes: treated retailers' wholesale unit costs fell by 12.6 percent, their retail prices fell by 5.3 percent, and the number of distinct goods sold over the two-week intervention period increased by approximately 20 percent. Collective procurement therefore generated cost savings that were partly passed through to consumers while also allowing retailers to experiment with a broader assortment.

We develop a structural model with CES Cournot competition and a stocking entry game to separate three mechanisms behind these effects and test whether the data support joint maximization. The estimated wholesale schedule implies that doubling the pooled quantity lowers marginal cost by approximately 4.7 percent. For goods previously sold by another group member, treatment reduces the estimated setup cost of entry by approximately 20 percent. At the same time, retailers who source together continue to compete independently in the retail market: the estimated conduct parameter is statistically indistinguishable from individual Cournot competition, while the hypothesis that the group behaves as a single downstream firm is decisively rejected. We find important welfare gains for both consumers and retailers. Supplier access accounts for most of the retailers' gains, whereas consumers benefit primarily from the pass-through of quantity-driven procurement savings; shared transportation contributes a smaller additional gain.

The paper contributes to a broader literature on how productive scale is achieved in settings where firms remain small and informal. Economically relevant scale need not coincide with firm boundaries. Retailers remain separate businesses, but can still share procurement scale through transport, pooled purchases, and access to supplier relationships. More generally, the paper shows that missing scale in

developing-country markets may reflect missing organizational infrastructure.

The results have important policy implications given the resources geared toward single-owner firms operating in retail. In urban food markets where informal retailers play a central role in household consumption, light-touch interventions that help firms test cooperative arrangements may improve market performance without generating the harms typically associated with coordination among competitors. Facilitating cooperation can therefore be a pro-consumer market-design tool: one that lowers prices, expands availability, and improves the functioning of fragmented markets.

References

Ashraf, N., A. Delfino, E. L. Glaeser, and K. Sarnoff (2025, December). Female entrepreneurship and trust in the market. Working paper.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010). Estimating static models of strategic interactions. Journal of Business & Economic Statistics 28(4), 469–482.

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2024). Changes in social network structure in response to exposure to formal credit markets. The Review of Economic Studies 91(3), 1331–1372.

Bassi, V., R. Muoio, T. Porzio, R. Sen, and E. Tugume (2022). Achieving scale collectively. Econometrica 90(6), 2937–2978.

Bloch, F. (1995). Endogenous structures of association in oligopolies. RAND Journal of Economics 26(3), 537–556.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013, February). Does management matter? evidence from india. The Quarterly Journal of Economics 128(1), 1–51.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. The Review of Economic Studies 89(1), 181–213.

Cai, J. and A. Szeidl (2018). Interfirm relationships and business performance. The Quarterly Journal of Economics 133(3), 1229–1282.

Christiaensen, L., J. De Weerdt, and Y. Todo (2013). Urbanization and poverty reduction: The role of rural diversification and secondary towns. Agricultural Economics 44(4–5), 435–447.

de Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: Evidence from a field experiment. The Quarterly Journal of Economics 123(4), 1329–1372.

de Mel, S., D. J. McKenzie, and C. Woodruff (2009). Measuring microenterprise profits: Must we ask how the sausage is made? Journal of Development Economics 88(1), 19–31.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014). Microenterprise growth and

the flypaper effect: Evidence from a randomized experiment in ghana. Journal of Development Economics 106, 211–226.

Fafchamps, M. and S. Quinn (2018). Networks and manufacturing firms in africa: Results from a randomized field experiment. The World Bank Economic Review 32(3), 656–675.

Field, E., S. Jayachandran, R. Pande, and N. Rigol (2016). Friendship at work: Can peer effects catalyze female entrepreneurship? American Economic Journal: Economic Policy 8(2), 125–153.

Fleischman, G. (2026, July). Economic and psychological returns to social relationships: Alleviating constraints to network formation in malawi. Job market paper, University of Pittsburgh.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014, August). Learning through noticing: Theory and evidence from a field experiment. The Quarterly Journal of Economics 129(3), 1311–1353.

Hardy, M. and J. McCasland (2021). It takes two: Experimental evidence on the determinants of technology diffusion. Journal of Development Economics 149, 102600.

Hardy, M. and J. McCasland (2023). Are small firms labor constrained? experimental evidence from ghana. American Economic Journal: Applied Economics 15(2), 253–284.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems 44(2), 174–199.

Henderson, J. V. and S. Kriticos (2018). The development of the african system of cities. Annual Review of Economics 10, 287–314.

Hjort, J., V. Iyer, and G. de Rochambeau (2020, August). Informational barriers to market access: Experimental evidence from liberian firms. Working Paper 27662, National Bureau of Economic Research.

Iacovone, L. and D. McKenzie (2022, October). Shortening supply chains: Experimental evidence from fruit and vegetable vendors in bogota. Economic Development and Cultural Change 71(1), 111–149.

Karlan, D. and M. Valdivia (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. The Review of Economics and Statistics 93(2), 510–527.

Macchiavello, R. and A. Morjaria (2022, September). Empirical approaches to trust and relational contracts. International Journal of Industrial Organization 84, 102849.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. RAND Journal of Economics 17(1), 48–58.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. Journal of Development Economics 99(2), 210–221.

McKenzie, D. and C. Woodruff (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? The World Bank Research Observer 29(1), 48–82.

McKenzie, D. and C. Woodruff (2017, September). Business practices in small firms in developing countries. Management Science 63(9), 2967–2981.

Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the MillerCoors joint venture. Econometrica 85(6), 1763–1791.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica 69(2), 307–342.

Salganik, M. J. and D. D. Heckathorn (2004). Sampling and estimation in hidden populations using respondent-driven sampling. Sociological Methodology 34(1), 193–239.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. RAND Journal of Economics 37(3), 619–640.

Sweeting, A. (2009). The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. RAND Journal of Economics 40(4), 710–742.

A Additional Tables and Figures

A.1 Tables

Table A.1: Balance Table between Treatment and Control Group at Baseline

	Pooled	Control	Treatment	Diff	p-val	q-val
Demographics						
Below 30yo	19.95	21.05	19.83	-1.23	0.78	1.00
Has no schooling	67.85	65.79	68.08	2.29	0.68	1.00
Has some primary educ	22.31	22.37	22.30	-0.07	1.00	1.00
Has more than primary educ	18.77	21.05	18.51	-2.54	0.58	1.00
Don't know how to read or write	72.57	67.11	73.18	6.07	0.22	1.00
Belongs to Dioula ethnic group	19.69	22.37	19.39	-2.98	0.62	1.00
Belongs to Baoule ethnic group	24.02	18.42	24.64	6.21	0.20	1.00
Is a foreigner	12.34	15.79	11.95	-3.84	0.32	1.00
Is the head of household	35.30	35.53	35.28	-0.25	0.78	1.00
Father was a farmer	43.04	44.74	42.86	-1.88	0.65	1.00
Mother was in retail	58.14	57.89	58.16	0.27	0.80	1.00
Market Features						
Respo's no of selling loc	1.11	1.14	1.11	-0.04	0.51	1.00
Respo's no of operating market	1.03	1.01	1.03	0.02	0.35	1.00
Tenure in this market	13.77	14.68	13.67	-1.01	0.42	1.00
Experience as a retailer	16.90	17.28	16.86	-0.41	0.90	1.00
Has been a retailer all life	88.58	90.79	88.34	-2.45	0.51	1.00
Is part of retailers associations	9.97	5.26	10.50	5.23	0.07	1.00
Using current stall for free	21.65	19.74	21.87	2.13	0.64	1.00
Retail is the main and unique activity	97.51	94.74	97.81	3.08	0.08	1.00
Business started to support hh	32.15	31.58	32.22	0.64	0.86	1.00
Business started to be independent and earn money	56.04	56.58	55.98	-0.60	0.89	1.00
No accounting practice	95.67	100.00	95.19	-4.81	0.08	1.00
Can stock everything on site	91.47	94.74	91.11	-3.63	0.32	1.00
Productivity						
Hours worked per day	8.77	8.57	8.79	0.23	0.29	1.00
Hours spent buying inputs at Bromakoté	1.23	1.11	1.24	0.13	0.37	1.00
Days worked in the last month	27.09	26.67	27.14	0.47	0.52	1.00

Continued on next page

(continued from previous page)

	Pooled	Control	Treatment	Diff	p-val	q-val
Months worked in the last year	11.55	11.57	11.54	-0.02	0.82	1.00
Average weekly sales	62480.61	62314.07	62499.06	184.99	0.99	1.00
Average weekly costs	32949.17	34175.06	32813.35	-1361.71	0.84	1.00
Self-declared average weekly profit	23923.22	22550.00	24076.25	1526.25	0.66	1.00
Saving over the last 30 days	25181.45	22763.16	25456.59	2693.43	0.39	1.00
Clients served per hour	5.08	5.20	5.06	-0.14	0.71	1.00
Selling Activity						
No of good sold	4.64	4.21	4.69	0.48	0.08	1.00
Distinct goods on stall (baseline, by re7)	3.68	3.30	3.72	0.42	0.11	1.00
No of wholesalers	6.92	5.69	7.05	1.36	0.34	1.00
Having at least one regular wholesalers	78.74	80.26	78.57	-1.69	0.52	1.00
Sells nothing else but perishable veg/fruits	39.76	46.05	39.07	-6.99	0.23	1.00
Chooses goods according to wholesalers' prices	36.09	31.58	36.59	5.01	0.25	1.00
Chooses goods according to clients' requests	3.94	3.95	3.94	-0.01	0.99	1.00
Self-perceived direct competitors (%)	27.12	28.04	27.01	-1.02	0.66	1.00
Sourcing Activity						
Hours spent buying inputs at Bromakoté	1.23	1.11	1.24	0.13	0.37	1.00
Buys from Bromakoté	86.75	84.21	87.03	2.82	0.35	1.00
# of trip to wholesale market/week	2.87	2.98	2.86	-0.12	0.64	1.00
Buy all goods on credit	12.47	9.21	12.83	3.62	0.33	1.00
Share of weekly profit spent on transportation	0.24	0.26	0.23	-0.03	0.44	1.00
Fixed Costs						
Weekly cost of taxes	721.39	766.45	716.40	-50.05	0.81	1.00
Weekly cost of plastic bags	1755.51	1979.39	1730.71	-248.68	0.21	1.00
Weekly cost of transportation costs	2370.54	2358.55	2371.86	13.31	0.93	1.00
Weekly cost of security services	134.18	144.51	133.03	-11.48	0.68	1.00
Weekly cost of phonecredit	354.80	359.21	354.31	-4.90	1.00	1.00
Weekly cost of financial services	1159.25	878.95	1190.31	311.36	0.43	1.00
Social Network						
Nb of women respo asks for business advice	1.35	1.39	1.35	-0.05	0.74	1.00
Nb of women respo goes to wholesalemkt with	0.98	0.74	1.01	0.27	0.07	1.00
Nb of women respo cooperates with	0.61	0.42	0.63	0.21	0.08	1.00
Nb of women respo shares clients with	2.03	2.20	2.01	-0.19	0.17	1.00

Continued on next page

(continued from previous page)

	Pooled	Control	Treatment	Diff	p-val	q-val
Nb of women respo resells goods	1.54	2.26	1.46	-0.80	0.01	1.00
Nb of women respo discusses selling prices	1.48	1.51	1.48	-0.03	0.67	1.00
Nb of women respo asks for perso advice	0.78	0.84	0.78	-0.06	0.54	1.00
Nb of women respo knows name	7.73	7.26	7.78	0.52	0.35	1.00
Nb of women respo knows phone no	2.61	2.34	2.64	0.30	0.39	1.00
Nb of women considered as sister	1.57	1.95	1.53	-0.42	0.02	1.00
Short Absence: ask neighboor	79.66	77.63	79.88	2.25	0.72	1.00
Short abssence: ask family	18.24	21.05	17.93	-3.12	0.57	1.00
Long Absence: ask neighboor	19.69	19.74	19.68	-0.06	0.93	1.00
Long Absence: ask family	41.60	44.74	41.25	-3.48	0.47	1.00
Job Opportunity						
Has been looking for a job in the last 30 days	5.25	7.89	4.96	-2.94	0.28	1.00
In the future, would prefer to be self-employed	85.83	88.16	85.57	-2.59	0.36	1.00
Is a retailer by choice rather than lack of opportunity	83.60	89.47	82.94	-6.53	0.08	1.00
Observations	762	76	686	762	762	762

Source: Retailer baseline survey.

Notes: The table reports baseline means and treatment-control balance for the full sample of retailers, organized into eight thematic blocks. The Pooled column reports the unconditional mean across all retailers. The Control and Treatment columns report means separately by assignment status. The Diff column reports the coefficient from an OLS regression of each variable on a pooled treatment indicator, capturing the mean difference between treatment and control retailers. The p-val column reports the corresponding p-value from a two-sample t-test of equality of means between treatment and control groups. Binary variables are multiplied by 100 and expressed as percentages. The sample consists of all retailers enrolled in the study at baseline, prior to treatment assignment.

Table A.2: Data Collection Summary

	Baseline *July 2024*	**Midline** *treatment period*	**Endline** *post-treatment*	**12mo Follow-up** *July 2025*
Retailers	762	740	735	638
Sample	Screened from a census of 1,037 retailers across 12 markets: selling F&V inside the marketplace on the survey day and self-declared *regular* sellers	CP participants and a random sample of non-participants, drawn from the baseline sample	Baseline sample re-contacted	Baseline sample re-contacted
Methodology	Single visit per retailer, fielded over 1 week	High-frequency panel: ≈4 visits over 10 days (2 per retailer per week)	Single visit per retailer, fielded in under 1 week	2 visits per retailer, 1 week apart
Attrition	–	2.9%	3.5%	16.3%

Notes: Wave dates and the position of each wave relative to the intervention are shown in Figure A.6. Every wave collects the retailer module and the exhaustive good-level roster described in Section 2.2.1; the endline and the 12-month follow-up repeat the baseline modules. Short phone check-ins fielded in December 2024 are not reported here. Attrition is measured relative to the 762 retailers surveyed at baseline and is balanced across treatment arms (Table A.1).

Table A.3: Effects on Retailer Price, Cost and Markups, Pooled Treatment

	Log Selling Prices	Log Wholesale Price Cost	Mark-up (p-c)/p
Treatment	0.018	0.011	-0.007
	(0.026)	(0.038)	(0.028)
	{0.482}	{0.771}	{0.804}
Midline	0.023	0.060	-0.011
	(0.028)	(0.039)	(0.033)
	{0.406}	{0.123}	{0.736}
Treat × Midline	-0.013	-0.123	0.030
	(0.030)	(0.044)	(0.035)
	{0.653}	{0.005}	{0.402}
Observations	7387	5499	7231
Control Mean	3.46	2.81	0.33
Fix Effects	Mkt x Good	Mkt x Good	Mkt x Good
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, baseline and midline.

Notes: The table reports difference-in-differences estimates from a regression of each outcome on a pooled treatment indicator, a post-period indicator (Midline), and their interaction, estimated using reghdfe with Market × Goods fixed effects. Observations are weighted by IPCW at midline; baseline observations receive a weight of one. The Treatment row captures pre-existing differences across treatment and control retailers. The Midline row captures common time trends. The Treat × Midline row is the coefficient of interest and captures the causal effect of treatment on the outcome at midline. Standard errors are clustered at the retailer level and reported (in parentheses); p-values are reported in {curly brackets}. Log Selling Prices and the Markup (p-c)/p are estimated on the core bundle sample. Log Wholesale Price Cost is additionally restricted to retailers for whom a matched wholesale price is observed. The control mean is computed among control retailers at baseline. * p<0.10, ** p<0.05, *** p<0.01.

Table A.4: ITT Effects on Retailer Price, Cost and Markups

	Log Selling Prices	Log Wholesale Price Cost	Mark-up (p-c)/p
Panel A. Pooled			
Treatment	-0.004	-0.085	0.042
	(0.018)	(0.025)	(0.020)
	{0.822}	{0.001}	{0.033}
Panel B. Unpooled			
T1: Info	0.012	-0.118	0.074
	(0.024)	(0.035)	(0.025)
	{0.604}	{0.001}	{0.004}
T2: Meetings	-0.009	-0.073	0.033
	(0.019)	(0.027)	(0.020)
	{0.637}	{0.007}	{0.104}
T3: Monitor	-0.007	-0.081	0.036
	(0.019)	(0.027)	(0.021)
	{0.708}	{0.003}	{0.081}
Observations	5413	4248	5322
P-Val T1=T2	0.274	0.142	0.021
P-Val T2=T3	0.900	0.676	0.750
Control Mean	3.54	2.79	0.33
Fix Effects	Mkt × Goods	Mkt × Goods	Mkt × Goods
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, midline.

Notes: The table reports ITT estimates from a cross-sectional regression of each outcome on treatment assignment, estimated using reghdfe with Market × Goods fixed effects and inverse probability of censoring weights (IPCW). Panel A reports the pooled treatment effect. Panel B reports treatment effects by arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Standard errors are clustered at the retailer level and reported (in parentheses); p-values are reported in {curly brackets}. Log Selling Prices and the Markup (p-c)/p are estimated on the core bundle sample. Log Wholesale Price Cost is additionally restricted to retailers for whom a matched wholesale price is observed. The control mean is computed among control retailers at midline. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding treatment arm coefficients. * p<0.10, ** p<0.05, *** p<0.01.

Table A.5: DID estimates on Full Sample

	Log Selling Prices	Log Wholesale Price Cost	Mark-up (p-c)/p
Panel A. Pooled			
Treatment	0.029	0.019	-0.009
	(0.017)	(0.030)	(0.020)
	{0.092}	{0.517}	{0.667}
Treat × Midline	-0.035	-0.082	0.036
	(0.019)	(0.032)	(0.024)
	{0.059}	{0.010}	{0.143}
Panel B. Unpooled			
T1: Info × Midline	-0.047	-0.140	0.063
	(0.025)	(0.044)	(0.030)
	{0.064}	{0.001}	{0.036}
T2: Meetings × Midline	-0.049	-0.071	0.023
	(0.021)	(0.035)	(0.027)
	{0.021}	{0.046}	{0.387}
T3: Monitor × Midline	-0.021	-0.072	0.036
	(0.021)	(0.035)	(0.025)
	{0.323}	{0.039}	{0.156}
Observations	14523	11576	14191
P-Val T1=T2	0.936	0.068	0.098
P-Val T2=T3	0.104	0.973	0.479
Control Mean	3.48	2.90	0.30
Fixed Effects	Mkt x Good, Week	Mkt x Good, Week	Mkt x Good, Week
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, midline.

Notes: The table reports DID estimates from a regression of each outcome on a pooled treatment indicator, a post-period indicator (Midline), and their interaction, estimated using reghdfe with Market × Goods fixed effects. Panel A reports the pooled treatment effect. Panel B reports treatment effects by arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Standard errors are clustered at the retailer level and reported (in parentheses); p-values are reported in {curly brackets}. Log Selling Prices and the Markup (p-c)/p are estimated on the core bundle sample. Log Wholesale Price Cost is additionally restricted to retailers for whom a matched wholesale price is observed. The control mean is computed among control retailers at midline. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding treatment arm coefficients. * p<0.10, ** p<0.05, *** p<0.01.

Table A.6: ANCOVA estimates on Core Goods Sample

	Log Selling Prices	Log Wholesale Price Cost	Mark-up (p-c)/p
Panel A. Pooled			
Treatment	-0.052	-0.145	0.057
	(0.027)	(0.044)	(0.029)
	{0.050}	{0.001}	{0.052}
Panel B. Unpooled			
T1: Info	-0.082	-0.230	0.102
	(0.037)	(0.063)	(0.036)
	{0.026}	{0.000}	{0.005}
T2: Meetings	-0.054	-0.116	0.050
	(0.031)	(0.048)	(0.031)
	{0.081}	{0.016}	{0.113}
T3: Monitor	-0.041	-0.143	0.049
	(0.030)	(0.051)	(0.031)
	{0.166}	{0.005}	{0.118}
Observations	4966	3720	4709
P-Val T1=T2	0.403	0.040	0.058
P-Val T2=T3	0.637	0.514	0.961
Control Mean	3.54	2.77	0.33
Fix Effects	Mkt	Mkt	Mkt
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}

Source: Retailer panel survey, midline.

Notes: The table reports ANCOVA estimates from a regression of each outcome on a pooled treatment indicator, a post-period indicator (Midline), and their interaction, estimated using reghdfe with Market × Goods fixed effects. Panel A reports the pooled treatment effect. Panel B reports treatment effects by arm: T1 (Information), T2 (Meeting), and T3 (Monitoring). Standard errors are clustered at the retailer level and reported (in parentheses); p-values are reported in {curly brackets}. Log Selling Prices and the Markup (p-c)/p are estimated on the core bundle sample. Log Wholesale Price Cost is additionally restricted to retailers for whom a matched wholesale price is observed. The control mean is computed among control retailers at midline. P-Val T1=T2 and P-Val T2=T3 report the p-values from F-tests of equality between the corresponding treatment arm coefficients. * p<0.10, ** p<0.05, *** p<0.01.

Table A.7: Decomposition of Midline-Sold Goods: Real Groups vs. Random-Group Placebo

Entry channel for goods sold at midline	Grouped retailer	Singleton retailer	Random group (placebo)
(1) Own (baseline / contract / core)			
Frequency	965	1,527	2,492
Percent	55.0	64.1	60.2
(2) Group-exposed (peer core / peer contract)			
Frequency	274	0	416
Percent	15.6	0.0	10.1
(3) Other (no observed channel)			
Frequency	159	857	689
Percent	9.1	35.9	16.7
(4) Group ever sold (peer sold, no peer core/contract)			
Frequency	355	0	540
Percent	20.3	0.0	13.1
Observations (retailer-good cells)	1,753	2,384	4,137

Source: Retailer panel survey, baseline and midline article roster.

Notes: The unit is a retailer × good (i, g) cell with the good on the stall at midline. Each cell is assigned to exactly one of four channels, in the order listed: (1) the good is already in i's baseline roster, her pre-treatment core bundle, or covered by her own wholesale contract; (2) not own, but covered by a peer's pre-treatment core or contract; (3) neither, and no peer sold the good; (4) not own and not peer core/contract, but some peer previously sold the good. Peer sets are leave-one-out within the retailer's group. Column 1 uses real CP groups (`source_id`, the same grouping as the Jaccard figures); column 2 covers ungrouped retailers, for whom channels (2) and (4) are zero by construction. Column 3 reassigns *all* retailers to random groups of three to four drawn within their own marketplace (seed 12345; the partition used for the Jaccard placebo), so it spans the full sample and its observation count is the sum of columns 1 and 2. Percentages are column shares and sum to 100 within each column.

Table A.8: Treatment Effects on the Variety Ratio and Assortment Churn

	Variety Ratio	+ Baseline Variety	Jaccard Churn	+ Baseline Variety
Panel A. Pooled				
Treatment	0.146	0.117	0.059	0.043
	(0.049)	(0.048)	(0.033)	(0.031)
	{0.003}	{0.015}	{0.070}	{0.162}
Panel B. Unpooled				
T1: Info	0.107	0.092	0.100	0.093
	(0.061)	(0.060)	(0.040)	(0.038)
	{0.081}	{0.126}	{0.013}	{0.014}
T2: Meetings	0.132	0.096	0.058	0.035
	(0.054)	(0.052)	(0.035)	(0.033)
	{0.015}	{0.069}	{0.101}	{0.287}
T3: Monitor	0.173	0.144	0.047	0.033
	(0.054)	(0.052)	(0.035)	(0.032)
	{0.002}	{0.006}	{0.172}	{0.315}
Observations	740	740	651	651
P-Val T1=T2	0.621	0.942	0.159	0.041
P-Val T2=T3	0.320	0.215	0.632	0.897
Control Mean	1.36	1.36	0.38	0.38
Fixed Effects	Mkt	Mkt	Mkt	Mkt
Clustering Lvl.	{Retailer}	{Retailer}	{Retailer}	{Retailer}

Notes: Columns 1–2 report the retailer-level variety ratio R_i, the number of distinct goods seen on the stall over the midline wave divided by the mean number of distinct goods on the stall per day. $R_i = 1$ means the retailer displays the same basket every day; $R_i = 2$ means she cycles through twice as many goods over the wave as she carries on a given day. Columns 3–4 report Jaccard churn C_i, one minus the average Jaccard similarity of the good sets on two survey days, taken over all day pairs: $C_i = 0$ if the basket never changes, $C_i = 1$ if no good is ever repeated on two days. The two measures answer the same question but R_i reads only the marginal presence counts while C_i reads which goods co-occur, and R_i rises mechanically with the number of survey days, which treatment moved, whereas C_i does not. Churn needs at least two midline days, hence the smaller sample in columns 3–4. Estimates are midline ITT levels with market fixed effects, weighted by midline IPCW, clustered at the retailer. The baseline wave is a single survey day, so $R_i \equiv 1$ pre-treatment for every retailer, C_i is undefined, and no difference-in-differences is available. Columns 2 and 4 add the retailer's baseline daily variety, which is imbalanced by +0.485 goods on a control mean of 3.252 ($p = .030$) and enters the denominator of R_i. Panel A pools the arms; Panel B reports them separately. Standard errors in parentheses, p-values in {curly brackets}. * p<0.10, ** p<0.05, *** p<0.01.

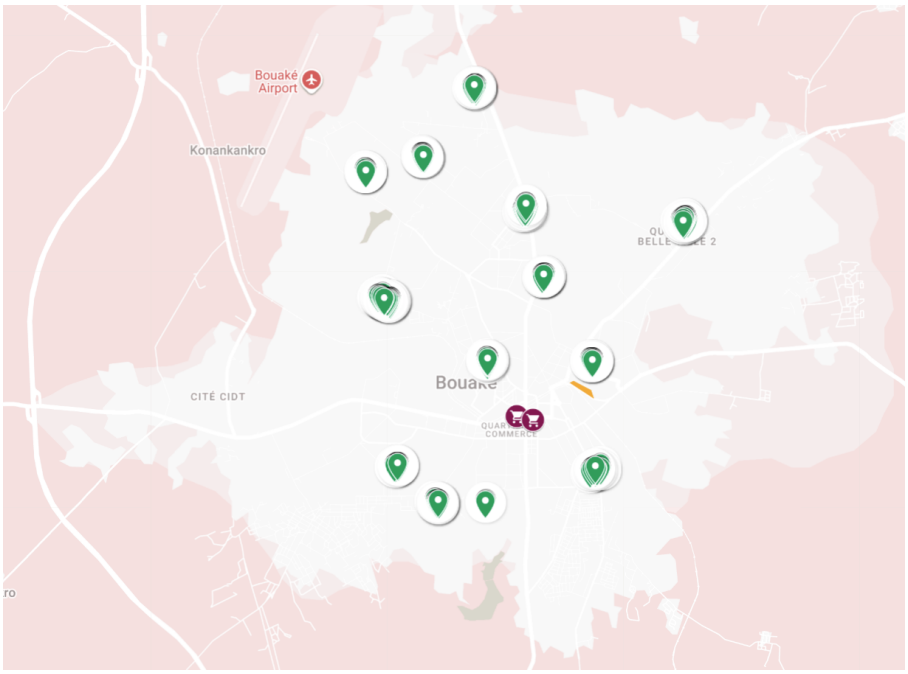
Table A.9: Long-Run Effects on Wholesale Cost: DiD and ANCOVA

	Log Wholesale Cost	
	Core	Full
DiD: Treat × Long-Term	0.051	-0.040
	(0.064)	(0.045)
	{0.430}	{0.378}
Observations	2405	5784
ANCOVA: Treat, baseline-adjusted	-0.000	-0.005
	(0.032)	(0.024)
	{0.988}	{0.825}
Observations	1335	1757
Control mean	2.803	2.755
Fixed Effects	Mkt × Good	Mkt × Good
Clustering Lvl.	{Retailer}	{Retailer}

Source: Retailer panel survey, baseline and 12-month follow-up.

Notes: This table is the wholesale-cost counterpart of Table 8 and is built identically: the outcome is the log wholesale unit cost paid, on the core bundle (Core) and on the full assortment (Full); the unit of observation is retailer × good × survey visit with the baseline visit as the pre-period; all specifications carry market × good and visit fixed effects, are weighted by the per-visit inverse-probability weight, and are clustered at the retailer. The DiD row is the coefficient on the interaction between assignment and the long-run wave. The ANCOVA row is the coefficient on assignment in the long-run visits alone, conditioning on the baseline value of the same retailer × good cell; it is estimated only on cells that have a baseline value, which is why observation counts are reported under each row. Wholesale cost is a null under both designs. The difference-in-differences estimates are imprecise and change sign between the two samples, so on their own they would not distinguish a small effect from none; the ANCOVA estimates are the informative ones, at −0.000 and −0.005 with standard errors roughly half as large, which rules out cost changes above about five percent in either direction. Standard errors in parentheses, p-values in {curly brackets}.

A.2 Figures



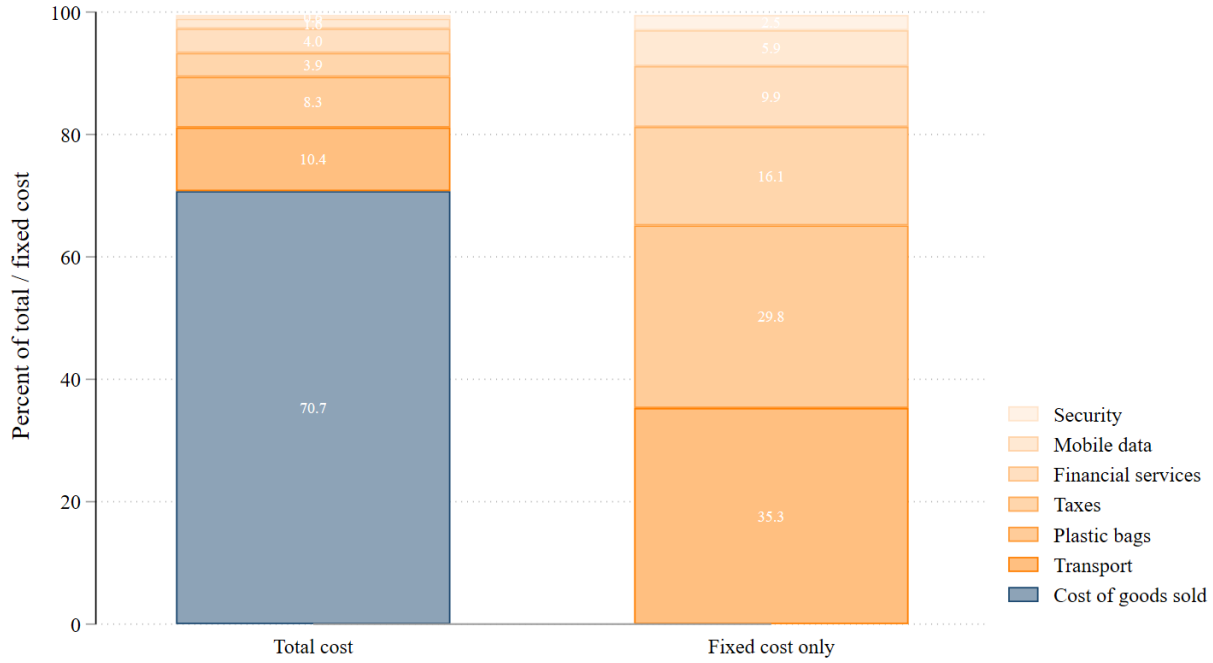
Note: This figure maps the main food retail locations in Bouaké. Green pins represent clusters of informal retailers operating within marketplaces, purple cart symbols identify the two supermarkets in the city, and the orange area marks the central wholesale market.

Figure A.1: Map of Bouaké showing spatial distribution of food marketplaces



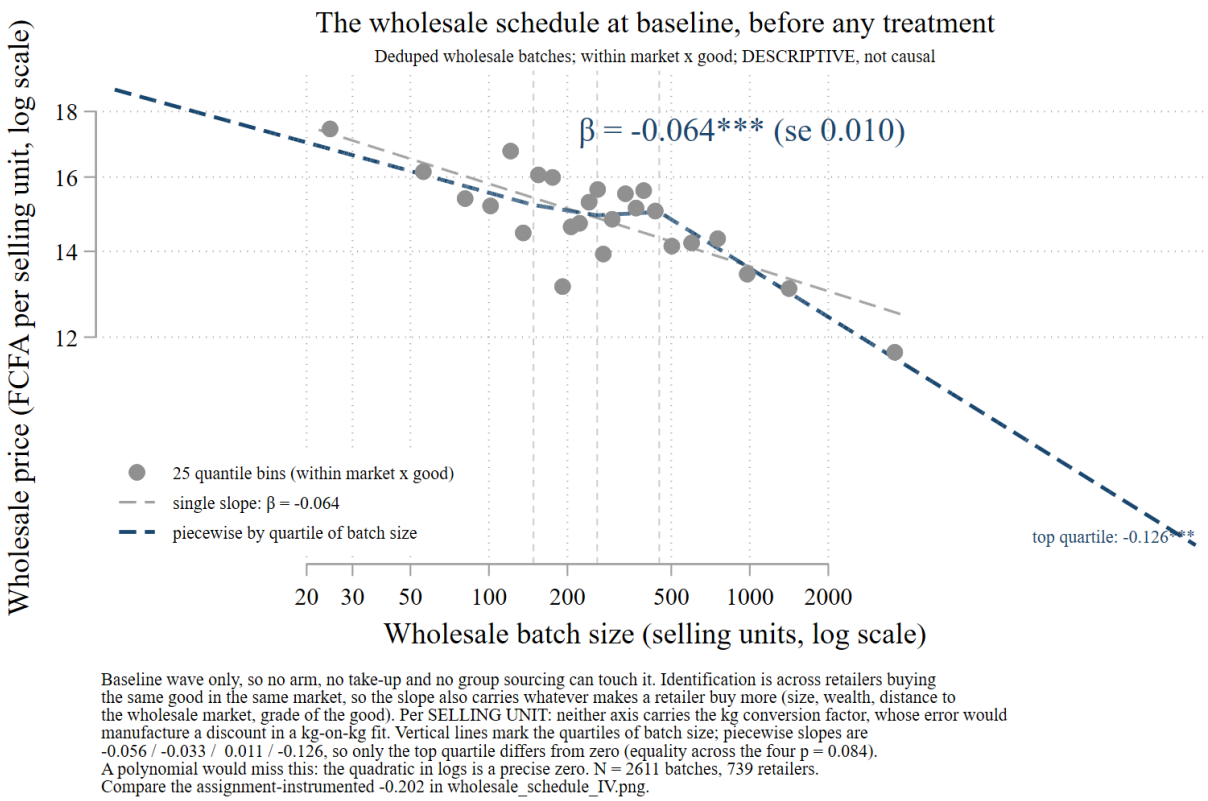
Note: This shows the GPS coordinates of 75% of the retailers surveyed in the Sokoura market. The attrition is due to tin roofs, which sometimes prevent precise GPS coordinates from being recorded.

Figure A.2: Retailer's Clustering Example



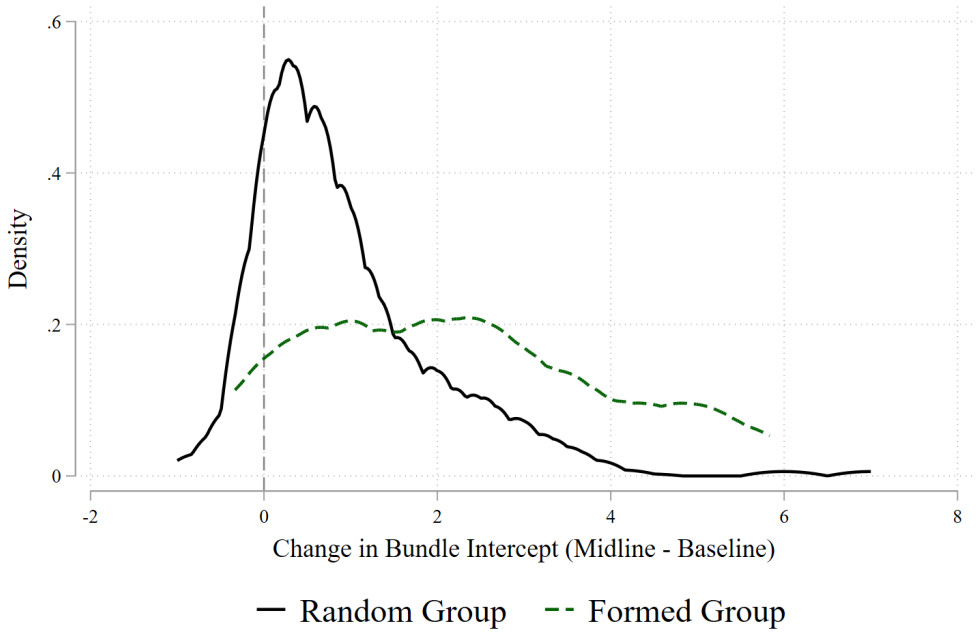
Note: TODO add notes

Figure A.3: Cost Decomposition



Note: TODO add notes

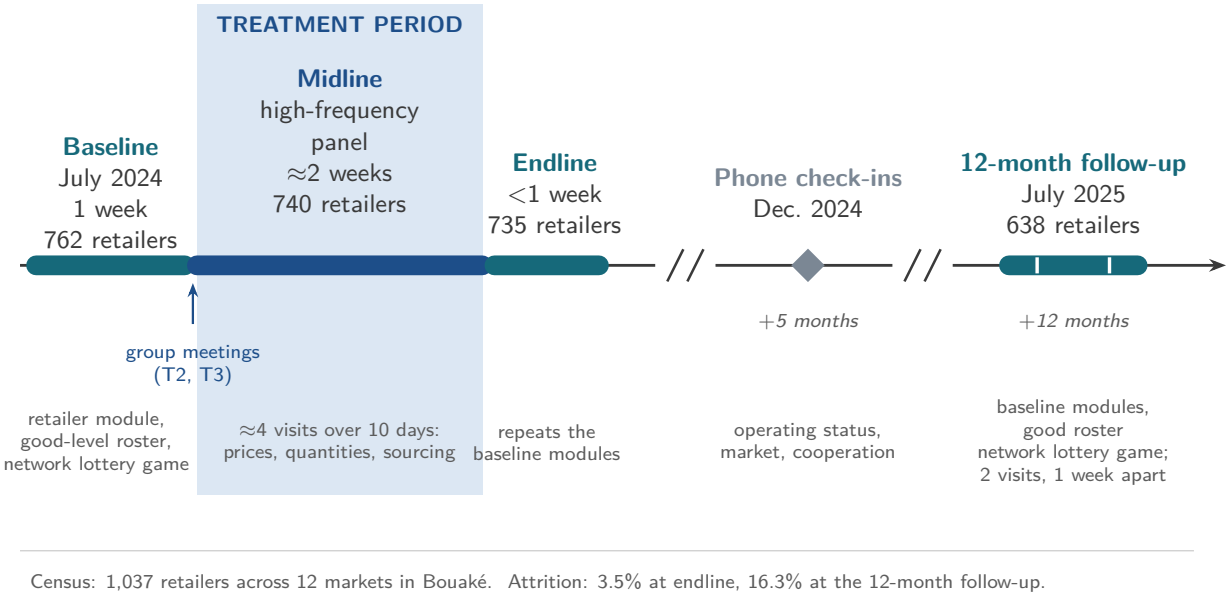
Figure A.4: Correlation between wholesale unit cost and quantity purchased



Source: Retailer panel survey, baseline and midline good-level roster.

Notes: The figure plots kernel densities of the within-pair change from baseline to midline in the *numerator* of the Jaccard index defined in Eq. (4), that is, the count of goods $|G_i \cap G_j|$ sold by both members of a pair. The unit of observation is a within-group retailer pair. *Formed Group* uses the observed CP groups; *Random Group* uses the placebo partition of retailers into groups of three to four drawn within their own marketplace. The vertical dashed line marks zero change. Mass to the right of zero indicates that pair members added goods already sold by their partner.

Figure A.5: Change in the Number of Goods Shared Within Groups, by Group Type



Note: TODO add notes

Figure A.6: Experimental Timeline

B Wholesale Discount Elasticity ψ: Robustness

This appendix reports the estimates underlying the preferred 2SLS estimate $\hat{\psi} = -0.0697$ in Section 6.2. The robustness exercises add a pre-period procurement control and allow the elasticity to vary with baseline contract access and baseline order size. We also report a pass-through diagnostic comparing the level-based wholesale share of marginal cost to the elasticity-based pass-through coefficient.

B.1 Specifications

All specifications share source FE, $g \times m$ FE, source-level IPCW weights, and source-clustered standard errors. They differ in the auxiliary regressors or sample split:

- **Canonical**: no-switch sub-sample, $\tilde{\rho}^{\text{pre}}_{smg} = \rho^{\text{pre}}_{img}$.
- **Pre-period control**: the canonical specification plus the source's baseline maximum contract

access.

- **Contract heterogeneity**: the canonical specification with $\tilde{\rho}^{\text{pre}}_{smg} \times \log Q^{wh}_{smgt}$ as a second endogenous regressor.

- **Order-size heterogeneity**: the canonical specification estimated separately in terciles of baseline log wholesale quantity, with cuts at 4.61 and 6.62.

B.2 Identification of curvature

If the wholesale schedule is nonlinear, the constant-elasticity estimate is a weighted average of local elasticities, with weights determined by where treatment moves the quantity distribution. Two retailer types operate on overlapping but substantially shifted regions of the wholesale-quantity support: contract retailers procure on average 3.15× more quantity per trip than spot retailers, and a Kolmogorov-Smirnov test decisively rejects equality of the two distributions ($D = 0.282$, $p < 10^{-28}$). Yet 99.2% of spot observations fall within the contract support, so identification of ψ is not driven by extrapolation off-support. The relevant question is whether ψ is constant along this shared support. Table B.1 reports the baseline domain by contract status.

Table B.1: Baseline Wholesale-Quantity Domain by Pre-treatment Contract Status

	N	mean (log)	sd (log)	p_{10}	p_{50}	p_{90}	median (level)
Spot retailers	1,827	5.238	1.845	2.90	5.10	7.82	164
Contract retailers	546	6.386	1.773	3.89	6.62	8.53	750

Notes: Mean log shift (contract − spot): +1.147; geometric mean ratio: 3.15×. Range overlap (log): [1.77, 10.39]; spot in overlap: 99.2%; contract in overlap: 100%. Kolmogorov-Smirnov test of distributional equality: $D = 0.282$, $p < 10^{-28}$. Sample: $t = 0$, no-switch, retailer-week level. Quantities in units.

B.3 Current estimates and robustness

The canonical estimate is $\hat{\psi} = -0.0697$ (s.e. 0.0367), with a 90-percent Anderson–Rubin set of $[-0.130, -0.020]$ and a 95-percent set of $[-0.150, 0.000]$. Adding the pre-period procurement control leaves the estimate unchanged at -0.0696. The order-size splits place the largest discount in the bottom tercile, -0.1991, while the middle-tercile estimate is close to zero and the top-tercile estimate is

Table B.2: The wholesale discount elasticity ψ

	$\hat{\psi}$	(s.e.)	AR 90%	AR 95%	FS F	Cells	Sources
Canonical (no-switch sources)	−0.0697*	(0.0367)	[−0.130, −0.020]	[−0.150, +0.000]	41.0	4,799	517
+ pre-period procurement control	−0.0696*	(0.0365)	[−0.130, −0.020]	[−0.140, +0.000]	43.7	4,799	517
By baseline order size							
Bottom third	−0.1991*	(0.1036)	[−0.420, −0.050]	[−0.490, −0.020]	15.7	1,592	259
Middle third	−0.0018	(0.0435)	[−0.070, +0.070]	[−0.090, +0.090]	20.9	1,513	318
Top third	−0.0999	(0.0692)	[−0.220, +0.020]	[−0.250, +0.050]	16.9	1,501	339
By baseline contract (Spec C3)							
Slope, no-contract sources (ψ_0)	−0.0461	(0.0392)	–	–	34.8	4,799	517
× baseline contract	−0.0752	(0.0595)	–	–	33.6	4,799	517
Slope, contract sources (sum)	−0.1213**	(0.0543)	–	–	–	4,799	517

Notes. Source×market×good×week panel. Each row is a separate DiD-IV: $\log \hat{c}^{s}_{smgt}$ on $\log Q^{wh}_{smgt}$, instrumented by $D_s \times \text{Post}_t$, with market×good and source fixed effects, source-IPCW weights and standard errors clustered at the source. The sample restricts to sources whose pre-existing wholesale contract access is unchanged by group formation. The pre-period procurement control is the source's baseline maximum contract access, entered as an exogenous regressor. Baseline-order-size terciles cut the same sample on the source's baseline log wholesale quantity at 4.61 and 6.62 (100 and 750 kg). The baseline-contract panel (Spec C3) treats both $\log Q^{wh}_{smgt}$ and its interaction with the source's baseline contract access as endogenous, instrumented by $D_s \times \text{Post}_t$ and its interaction with the baseline contract; the contract-source slope is the sum of the first two rows, with its standard error from the clustered joint covariance. Anderson–Rubin sets are not reported for that panel (two endogenous regressors), and its first-stage F is per instrument. Anderson–Rubin sets are inverted on a grid over $[-2, 2]$ in steps of 0.01 and are the weak-identification authority here; the first-stage F is the cluster-robust squared t on the excluded instrument and is reported for reference only. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

−0.0999. Allowing the slope to vary with baseline contract access gives −0.0461 for sources without a baseline contract and −0.1213 for contract sources; their difference is not precisely estimated.

B.4 Pass-through diagnostic

A complementary check compares the level-based wholesale share of marginal cost $\alpha^{\text{lvl}} = \overline{w/\hat{c}^{w}}$ to the elasticity-based pass-through coefficient α^{elas} from regressing $\log \hat{c}^{w}$ on $\log w$ within source. Equality of the two is consistent with stable retail-side variable cost. On the no-switch sample the level and elasticity-based estimates align ($\alpha^{\text{lvl}} = 0.81$, $\alpha^{\text{elas}} = 0.75$), supporting the wholesale-stage interpretation of ψ.

Table B.3: Wholesale Share in Marginal Cost: Level vs. Elasticity

	α^{lvl}	α^{elas}	SE(α^{elas})	$p(\alpha^{\text{elas}} = \alpha^{\text{lvl}})$
Spec A (full sample)	0.839	0.338	0.304	0.099
Spec C1 (no-switch)	0.805	0.754	0.913	0.956

Notes: Source FE, $g \times m$ FE, source-level IPCW weights, source-clustered SE.

B.5 Sample construction

The preferred no-switch panel has 4,799 source–market–good–week observations across 517 unique sources (groups or singletons). Within each source–good–week cell, $\log \hat{c}^{w}_{smgt}$, $\log w_{smgt}$, and $\log p_{smgt}$ are IPCW-weighted means of the corresponding retailer-level objects, with the source-level IPCW weight equal to the sum of member weights (preserving the individual-level IPW objective; Wooldridge, 2002, Ch. 17). At baseline the post-treatment group identifier is propagated back, so that a retailer's source at $t = 0$ is her future group. To make baseline and post-treatment magnitudes comparable, Q^{wh}_{smgt} is computed as the per-member mean of individually purchased quantity at baseline and as the sum of member quantities after treatment.

B.6 Sensitivity to η

The Cournot-inverted cost outcome $\hat{c}^{w}_{smgt} = p_{mgt}(1 - s_{imgt}/\eta)$ depends on the assumed demand elasticity. We re-estimate the preferred specification under $\eta \in \{1.5, 2.0, 2.5\}$: the point estimate $\hat{\psi}$ shifts by less than 0.01 in absolute value, and the AR confidence interval continues to exclude zero at 90%. The structural conclusions are robust to this range of the canonical across-product elasticity.

C Activation Block: Estimation Details and Validation

This appendix complements Section 6.3 with details on the candidate panel, the cost imputation ladder, the profiled MLE procedure, and four validation exercises that discipline the activation block. We use the notation of the variable dictionary in Appendix I.

C.1 Candidate panel

The activation panel pairs each retailer i in market m with every good g above 5% market penetration, across all weeks t the retailer is observed: 38,920 rows at the (i, m, g, t) level covering 732 retailers, 12 markets, and 29 goods. Two restrictions then apply. Blocks (m, g, t) in which no market traded that period carry no information about entry and are removed: 21 blocks, 713 rows, 1.80% of the panel, with all 243 cells retained since periods that did trade are kept. Because treatment could in principle drive that exclusion, we test it directly on six margins in Section 6.3; none rejects, though with only

21 excluded blocks the test has limited power and we state that rather than claim the null. Dropping a further 7 cells that lack a valid $t = 0$ component for common support leaves the surveyed sample of 38,496 rows across 236 cells and 697 (m, g, t) blocks.

The pooled-channel indicators are

$$\text{own_exposure}_{img} = \mathbf{1}\{\rho^{\text{own}}_{img} = 1 \text{ or } \text{core}_{ig} = 1\}, \qquad \text{grp_exposure}_{img} = \mathbf{1}\{\rho^{\text{grp}}_{img} = 1 \text{ or } \text{core}^{\text{grp}}_{ig} = 1\},$$

with $\text{core}_{ig} = \mathbf{1}\{i \text{ stocks } g \text{ in } \geq 80\% \text{ of pre-treatment days}\}$. The own-channel indicator is the incumbency indicator $I_{img,0}$ of (8): a retailer–good attribute, measured wholly at baseline and time-invariant in (i, g). It is active on 10.4% of rows; on these rows the search cost is overridden to zero (empirical activation rate 0.967, residual non-activation reflecting the 80% core threshold rather than an interior failure rate). The peer-channel indicator $G_{img,0}$ is active on 10.7% of rows; interacted with the difference-in-differences indicator D_{it}, this is the structural variation that identifies the access channel γ_{inter}.

Both components of $G_{img,0}$ are evaluated at $t = 0$. The core half is time-invariant in (i, g) by construction; the contract half is a per-week object in the raw data and is frozen at its baseline value, so that G is time-invariant in (i, g). Without the freeze, 329 of 14,872 (i, g) pairs (2.2%) switch exposure across periods, which would put a treatment-moving regressor inside the $D_{it} \times G$ interaction. The unfrozen construction is retained behind a switch and reported in the robustness ladder.

C.2 Pre-trip expected costs and post-commitment quote innovations

The activation kernel uses the procurement information available when the basket is fixed, not the cost eventually observed after a purchase. For every candidate retailer–good pair we write

$$\log \hat{c}_{imgt} = \mu^{c}_{imgt} + \omega_{imgt}, \qquad \mathbb{E}[\omega_{imgt} \mid \mathcal{I}^{1}_{imgt}] = 0.$$

For $t \geq 1$, the preferred prediction of μ^{c}_{imgt} uses retailer i's own baseline cost for good g, the lagged market–good cost mean, and good effects. Rows without own history use, successively, the lagged market–good mean, the good-level baseline mean, and the global baseline mean. The same prediction rule is applied to sellers and non-sellers: a seller's realized cost never overrides the pre-trip object in the

activation kernel. At $t = 0$ there is no earlier cost observation, so μ^c_{img0} is a good-by-market projection and the corresponding ω_{img0} is a cross-sectional residual rather than a within-(i, g) innovation.

The empirical distribution $\widehat{G}_\omega$ is re-estimated from $\log \hat{c}_{imgt} - \mu^c_{imgt}$ on observed seller rows and winsorized at its pooled first and ninety-ninth percentiles. Hence the wholesale elasticity and the cost regression used to construct dated predictors are unchanged, but the activation model conditions only on the predictable component and integrates over the transaction-specific innovation.

C.3 Estimation: typed-belief simulated profit

The estimator has an architecture worth stating plainly, because it is what distinguishes it from a nested-fixed-point routine: expected profit is computed once, before the likelihood is ever evaluated. The likelihood then contains no equilibrium solve at all. Equilibrium is imposed only in the counterfactual, where it belongs.

Step 1: beliefs. A retailer choosing whether to stock g does not know which rivals will stock it. She holds a belief P^{CCP}_{jmgt} over each potential rival's participation, which we estimate as a per-period participation logit on predetermined states,

$$P^{CCP}_{jmgt} \;=\; \Lambda\big(\zeta^{CCP}_{mg,0} + x'_{jmg,0}\beta_t\big).$$

The covariates $x_{jmg,0}$ are measured at baseline and the CCP cell intercept $\zeta^{CCP}_{mg,0}$ is identified jointly with β_0 at $t = 0$, then carried into $t = 1, 2$ as a fixed offset while only the slopes β_t are refit. Beliefs are accordingly period-specific and are not the baseline choice probabilities; what is frozen is the cell level and the states, not the mapping between them (Section 5.3). Baseline carries the fullest sample — 14,787 rows against 12,101 and 12,745 in the two midline weeks — so $\zeta^{CCP}_{mg,0}$ is pinned where it is best identified and before attrition selects who contributes to it. Nothing contemporaneous enters: no rival's realised activation in the current period is used to predict any other's, so no free-entry feedback contaminates the beliefs. The reading is that retailers know the baseline composition of their market and anticipate how it maps into participation, and that the mapping itself may shift as the programme rolls out.

Step 2: expected profit by simulation. Given beliefs, expected profit is the Stage-2 profit integrated over who actually competes and at what cost. For each of $D = 100$ draws we draw rival participation $w_j^{(d)} \sim \text{Bernoulli}(P_{jmgt}^{CCP})$ and draw the focal and rival costs as

$$c_i^{(d)} = \exp\left(\mu_{imgt}^c + \omega_i^{(d)}\right), \qquad c_j^{(d)} = \exp\left(\mu_{jmgt}^c + \omega_j^{(d)}\right), \qquad \omega_i^{(d)}, \omega_j^{(d)} \sim \widehat{G}_\omega.$$

Each rival therefore carries her own period-specific predictable cost center; rival costs are not anchored on the focal retailer. No additional incumbency shift is added because systematic differences associated with incumbency are already embedded in the retailer-specific μ_j^c. The regression of μ^c on retailer types is retained only as a diagnostic of that embedded loading. The timing assumption makes the later ω_j draw independent of activation, but does not rule out selection into activation on predictable μ_j^c.

We then solve the Cournot corner draw by draw. The corner matters: not every drawn entrant sells. Ranking drawn rivals from cheapest and adding them one at a time, each addition pushes the price down, and the active set is the longest prefix that survives its own price, $K^* = \max\{K : c_{(K)} < p(K)\}$. Expected profit is the average of realized profit over draws. Draws use common random numbers across all likelihood evaluations, so the objective is smooth in the parameters rather than jittering with the simulation.

Mean-field cross-check. The robustness estimator replaces the random rival set by its expected size and aggregate cost,

$$EN_i = \sum_{j \neq i} P_j, \qquad EC_i = \sum_{j \neq i} P_j \tilde{c}_j,$$

$$\tilde{p}_i = \frac{\hat{\eta}(\tilde{c}_i + EC_i)}{\hat{\eta}(1 + EN_i) - 1}, \qquad \tilde{s}_i = \hat{\eta}\left(1 - \frac{\tilde{c}_i}{\tilde{p}_i}\right),$$

$$\tilde{q}_i = \tilde{s}_i \hat{A}_{mgt} \tilde{p}_i^{-\hat{\eta}}, \qquad \bar{\pi}_{imgt}^{MF} = (\tilde{p}_i - \tilde{c}_i)\tilde{q}_i.$$

Profit is set to zero when the implied price, share, or quantity is non-positive. This approximation is computationally convenient but does not integrate over the nonlinear profit function; it is therefore used only as a cross-check on the simulated-profit estimator.

Step 3: likelihood. With $\bar{\pi}^{\text{ex}}$ in hand the objective is the binary-choice negative log-likelihood, minimised over four free parameters, $\theta = (\gamma_{\text{grp}}, \gamma_T, \gamma_{\text{inter}}, \log \tau)$. Rows with own_exposure = 1 are excluded from the entry margin by construction, since they carry $S \equiv 0$ and $P \equiv 1$ and so contain no information about the setup cost. The 236 cell intercepts α_{mg} are profiled, not estimated: at every evaluation a bisection profiles one intercept for each cell. They are nuisance parameters absorbing the level of entry in each cell, and τ and the channel coefficients are identified strictly from within-cell variation.

Identification. Each free parameter answers to a distinct slice of variation. The scale τ is identified by how sharply activation responds to expected profit, and the assumption that licenses this is timing rather than exclusion: beliefs are frozen at baseline, so the profit variation predates the entry decisions it explains. The access channel γ_{inter} is identified by the difference-in-differences shape within the treated group — the change in activation on group-reachable goods, relative to the same retailers' goods that no group member can reach. What separates it from a generic treatment effect is the off-group null: treatment moves activation on group-exposed goods and not elsewhere, so a story in which the programme simply raised stocking everywhere is rejected by the same variation that identifies the channel. The shape coefficient γ_{grp} is identified by cross-sectional variation in baseline group overlap, which is measured before treatment but defined over realized — hence endogenously chosen — group membership. It is held fixed in every counterfactual for that second reason, not the first: switching it would re-sort group composition, which requires a model of group formation this paper does not estimate. The cell intercepts are pinned by cell activation rates. Finally γ_T is not separately identified: with roughly nine retailers in ten treated, D_{it} is close to collinear with the post indicator, so it absorbs baseline imbalance and secular trend alongside any direct effect. We report it and claim nothing from it. The identified objects of this likelihood are γ_{inter} and τ.

C.4 Validation

Channel capture. The structural model is asked to reproduce the empirical group-vs-control gap in activation rates on the grp_exposure = 1 sub-population. The empirical gap is +12.3 percentage points; the structural prediction is +10.9 pp, capturing 89%.

Per-cell fit. Mean activation by (own_exposure, grp_exposure, T_i) in Table C.1: non-own rows fit within ±1.3 pp; own-exposure rows are mechanically pinned to $P = 1$ by the search-cost override, with a residual ∼ 3 pp gap reflecting the 3–4% of own-exposure rows that empirically chose not to activate.

Table C.1: Mean Activation by Cell Type (Observed vs. Predicted)

own	grp	T_i	N	mean a	mean P^*	gap
0	0	0	3,422	0.082	0.088	+0.006
0	0	1	28,816	0.098	0.099	+0.001
0	1	0	90	0.178	0.174	−0.004
0	1	1	3,177	0.221	0.208	−0.013
1	0	0	350	0.963	1.000	+0.037
1	0	1	2,759	0.967	1.000	+0.033
1	1	0	26	0.962	1.000	+0.038
1	1	1	993	0.968	1.000	+0.032

Stall-size distribution. The per-(retailer, week) sum $\sum_g a_{imgt}$ versus $\sum_g P^*_{imgt}$ is reported in Table C.2. The model's predicted stall size sits in the realistic 3–6 range for 90% of retailer-weeks. The calibration slope is close to one with a small negative intercept.

Table C.2: Per (Retailer × Week) Goods-on-Stall: Observed vs. Predicted

Statistic	Observed $\sum_g a_{imgt}$	Predicted $\sum_g P^*_{imgt}$
Mean	4.02	4.09
Median	3	3.91
p_{90}	8	5.94
p_{99}	12	8.26
Max	16	12.01
$\rho(N_{\text{obs}}, N_{\text{pred}})$ across 1,941 retailer-weeks	0.577	
Calibration $N_{\text{obs}} = \alpha + \beta N_{\text{pred}}$	$\hat{\alpha} = -0.51, \hat{\beta} = 1.11$	

Variety ITT replication. Projecting predicted activation probabilities onto the variety regressions of Section 3 yields the treatment effects in Table C.3. The cumulative two-week-explore margin is

matched within rounding; on the daily margin, the model captures 81% of the observed treatment effect on the structural sample.

Table C.3: Variety ITT: Bundle Size on Pooled Treatment

Specification	Source	Ctrl mean	$\hat{\beta}$	SE
Daily: $n_{ig,t} \sim T \mid t, m$	Observed	3.46	+0.597	0.264
	Predicted	3.65	+0.482	0.126
2-week cumulative: explore$_t \sim T \mid m$	Observed (struct. sample)	4.84	+0.878	0.415
	Predicted (struct. sample)	5.84	+0.905	0.180
	RF paper (full sample, ref.)	6.85	+0.65	—

Notes: Daily specification includes week and market fixed effects and IPCW weights, clustered by retailer. The 2-week cumulative specification includes market fixed effects only. Predicted standard errors are mechanically smaller because P^* is a smoothed, denoised activation rate; predicted SEs do not carry inferential content beyond the observed data.

Arm-by-arm dose-response. Splitting the predicted treatment effect by experimental arm reproduces the observed $C < T_1 < T_2 < T_3$ ranking within ±0.07 goods:

$$\text{T1: } (\hat{\beta}_{\text{obs}}, \hat{\beta}_{\text{pred}}) = (+0.530, +0.528), \quad \text{T2: } (+0.883, +0.867), \quad \text{T3: } (+1.001, +1.073).$$

Entry decomposition. Table C.4 reports the share of midline activations by category {own, group exposed, other}. The 4× jump in the group-exposure share between control (2.4%) and treated (9.8%) is reproduced almost exactly (2.3% → 9.1%): this is the empirical pattern the activation block is designed to capture.

Table C.4: Share of Midline Activations by Category

Category	Treat	N cells	Obs. activations	Share obs.	Share pred.
1_own	C	376	362	55.0%	54.3%
2_group_exp	C	90	16	2.4%	2.3%
3_other	C	3,422	280	42.6%	43.5%
1_own	T	3,752	3,629	50.8%	51.7%
2_group_exp	T	3,177	701	9.8%	9.1%
3_other	T	28,816	2,818	39.4%	39.2%

C.5 Open questions

Three limitations are worth flagging. First, the level of τ is less precisely determined than the access channel. The point estimate is $\hat{\tau} = 6{,}536$ FCFA with a Hessian standard error of 1,410, and the weak-instrument-robust Anderson–Rubin set from the demand-shifter route is [969, 6,785]. What is well determined is the access channel γ_{inter}, which is why the paper leans on it and reports the money metric together with its uncertainty.

Second, uncertainty in the canonical demand elasticity is not propagated. The across-product elasticity η enters the demand shifter in an exponent and therefore propagates into expected profit, $\hat{\tau}$, and every FCFA quantity in the welfare section. The estimates in this appendix condition on the point estimate $\hat{\eta} = 1.1455$.

Third, cross-good complementarity: a 29×29 correlation matrix on retailer-week activations shows the model captures 65% of the variation in pairwise correlations with slope near one, but under-predicts the strongest empirical co-movements. Goods are modelled as separate activation decisions linked only through the retailer's cost and exposure, so genuine bundle effects — a retailer stocking tomatoes because she stocks onions — are outside it. An extension with cross-good shocks is work in progress.

D Test of the Price-taker Assumption

The model treats retailers as price-takers on the wholesale schedule (15). Under this assumption, the monopsony-markdown term in the retailer FOC drops and the FOC collapses to (12). This appendix runs a horse race between the price-taker FOC ($\lambda = 0$) and full strategic internalization of the wholesale markdown ($\lambda = 1$). The data do not reject price-taking.

D.1 Setting and hypotheses

Retailers i purchase good g in week t from wholesale source $s(i, g, t)$, at retailer-level quantity q^{wh}_{imgt} and unit price w_{imgt}. Let $Q^{wh}_{g,s,t} = \sum_{j \in s} q^{wh}_{jmgt}$ denote the source-level total and $\lambda_{imgt} = q^{wh}_{imgt}/Q^{wh}_{g,s,t}$ the within-source wholesale share. The demand-side Cournot–CES inversion yields the implied marginal cost $\widehat{mc}_i = p_{mgt}(1 - s_{imgt}/\hat{\eta})$.

Under the constant-elasticity schedule $w(Q) = w_0 Q^{\psi}$, two competing retailer-cost formulations differ by the markdown term $\log(1 + \psi\,\lambda_i)$:

$$\log \widehat{mc}_{imgt} \;=\; \psi\, \log Q^{wh}_{g,s,t} \;+\; \underbrace{\lambda \cdot \log\!\big(1 + \psi\,\lambda_{imgt}\big)}_{\text{conduct}} \;+\; \mu_i \;+\; \mu_{g,m} \;+\; \mu_s \;+\; \varepsilon_{imgt}. \tag{24}$$

The parameter λ indexes conduct: $\lambda = 0$ is atomistic (price-taking); $\lambda = 1$ is full strategic internalization of the wholesale markdown. The sample is restricted to retailers whose pre-treatment wholesale contract access matches that of their source cell (`no_switch = 1`), which isolates the wholesale-composition channel from the contract-access channel. The preferred Stage 2 estimate $\hat{\psi} = -0.0697$ is plugged in via source-level IV.

D.2 Result

Table D.1 reports the horse-race estimate of λ under OLS (with the data selecting both ψ and λ) and under 2SLS (with the markdown term instrumented by the retailer's own DiD variable d_i and the mean DiD of her source-cell peers $\bar{d}_{-i}$). All specifications use good-by-market, source, and retailer fixed effects, with retailer-clustered standard errors.

Table D.1: Horse Race on $\log \widehat{mc}$: Price-taker ($\lambda = 0$) vs. Strategic Internalization ($\lambda = 1$)

	$\hat{\lambda}$	SE	$p(\lambda = 0)$	$p(\lambda = 1)$
OLS (M1, ψ free)	+0.039	0.289	0.89	< 0.001
2SLS (M2, $\{d_i, \bar{d}_{-i}\}$)	−0.744	3.294	0.82	0.60

Notes: N = 2,304 retailer-weeks (228 retailers, 55 sources, 166 $g \times m$ cells). First-stage F = 15.5; Sargan over-identification test p = 0.079. Good-by-market, source, and retailer fixed effects; retailer-clustered SE.

Neither specification rejects $\lambda = 0$. OLS rejects $\lambda = 1$ at $p < 0.001$. The 2SLS confidence interval is wide, but its point estimate is far from the strategic-internalization benchmark. We conclude that the wholesale markdown term is not present in the data: retailers behave as price-takers on the wholesale schedule, and the Stage 2 FOC (12) is empirically defensible.

E Test of the No-collusion Assumption

The activation model rules out collusive equilibria by construction: Stage 1 is Nash in a sub-modular game, and Stage 2 is Cournot under the canonical across-product elasticity η. Treatment arms T2 and T3 introduce coordination meetings that could in principle facilitate tacit retail collusion among co-located group members. If they do, the model misattributes the resulting markup gains to a lower search cost.

The main text tests this directly with the conduct parameter λ of Section 7, which asks whether sourcing co-members internalize one another's quantity choices and rejects the group-as-one-firm benchmark. The appendix first varies the demand system, then collects robustness to the cost wedge, baseline adjustment, fixed effects, sample, inference, estimator, and cross-good multiproduct effects. It next sets out the wedge test in full, reaching the same question from the cost side without using the conduct parameter. Pass-through under alternative conduct inversions and the coalition-scale exercise of Section F complete the case; the latter also validates power by injecting cartels into the data.

E.1 The λ test under alternative demand parameterizations

The conduct test is not free of the demand system. Equation (20) multiplies λ by $\chi_{imgt} = \ell_1 s^{\text{peer}}_{imgt}/(1 - \ell_0 - \ell_1 s_{imgt})$, where ℓ_0 and ℓ_1 are the intercept and share slope in the affine Lerner index $L(s) = \ell_0 + \ell_1 s$. Which elasticities are allowed to move markups therefore sets how sharply conduct can be identified at all. Under homogeneous-good demand only the across-good elasticity operates, so $\ell_0 = 0$ and $\ell_1 = 1/\eta$. Under nested CES both operate, $\ell_0 = 1/\sigma$ and $\ell_1 = 1/\eta - 1/\sigma$, and the slope collapses as the two elasticities converge.

Table E.1 runs the identical estimator, sample construction and fixed effects under three parameterizations that differ only in which elasticity is free. The first column allows only the across-good elasticity, at the demand-IV estimate $\hat{\eta} = 1.146$ of Section 6.1. The second allows only the across-stall elasticity: the outer nest is shut down at $\eta = 1$, and σ is profiled off the supply moments of Section J, giving $\hat{\sigma} = 4.74$ with a retailer-cluster bootstrap interval of $[4.33, 5.05]$. The third allows both, at the joint supply-moment estimates $(\hat{\sigma}, \hat{\eta}) = (4.00, 3.30)$. None of the three is offered as the preferred demand system — the supply moments reject $\eta = 1$ outright, and reject the demand-IV value of σ against the joint estimate — and that is the point. The question here is not which demand system is right but

whether the conduct verdict turns on the answer.

Table E.1: The conduct parameter under three demand parameterizations

	Across-good only $\eta = 1.146$	Across-stall only $\sigma = 4.74, \eta = 1$	Both, jointly $(\sigma, \eta) = (4.00, 3.30)$
Panel A: Lerner index $L(s) = \ell_0 + \ell_1 s$			
Intercept ℓ_0	0	0.211	0.250
Share slope ℓ_1	0.873	0.789	0.053
Panel B: conduct parameter			
$\hat{\lambda}$	−0.07 (0.15)	−0.06 (0.15)	−3.05 (2.29)
95% confidence interval	[−0.37, 0.22]	[−0.35, 0.24]	[−7.54, 1.43]
p-value, $\lambda = 0$	0.62	0.71	0.18
p-value, $\lambda = 1$	$< 10^{-11}$	$< 10^{-11}$	0.076
Pre-treatment placebo $\hat{\lambda}_0$	+0.33 (0.19)	+0.27 (0.19)	+1.81 (3.24)
Panel C: what the data identify, on a common scale			
$\hat{\lambda}\,\ell_1$	−0.065 (0.133)	−0.044 (0.118)	−0.162 (0.121)
$\lambda\ell_1$ implied by $\lambda = 1$	0.873	0.789	0.053
Markup effect at $\lambda = 1$ (log pts)	+15.8	+18.8	+1.1
Panel D: specification and sample			
Placebo slack $1 - \ell_0 - \ell_1$	0.127	0.000	0.697
Overidentification placebo t	+4.47	degenerate	−2.00
Observations	6,181	6,080	6,181
Retailer × good pairs	2,528	2,488	2,528
Rows with a co-active co-member	2,154	2,153	2,154

Notes: Identical estimator, sample construction, fixed effects — retailer × good and week — and IPCW weights throughout; only the demand parameters entering $L(s) = \ell_0 + \ell_1 s$ change. Standard errors in parentheses, clustered at the retailer level; confidence intervals are $\hat{\lambda} \pm 1.96\,\widehat{\text{se}}$. $\lambda = 0$ is individual Cournot, $\lambda = 1$ is group-as-one-firm. The across-good elasticity is the demand-IV estimate of Section 6.1; the across-stall elasticities are supply-moment estimates from Section J, profiled at $\eta = 1$ in column two and estimated jointly with η in column three.

Panel C rescales the estimate by the Lerner slope. Because λ enters the estimating equation only through $\chi_{imgt} \propto \ell_1$, the product $\lambda\ell_1$ is what the data identify, and it is comparable across columns where λ is not. The markup effect is $-\log(1 - \bar{\chi})$ evaluated at the mean χ over rows with a co-active co-member, and reports how far full internalization would move $\log(\hat{c}^w/w)$ on such a row.

Panel D reports the placebo restriction that the non-group rival share enters with coefficient zero, estimated at the headline fixed effects. The regressor and its placebo sum to $\ell_1(1 - s)/(1 - \ell_0 - \ell_1 s)$ and are therefore collinear by construction, exactly so when $1 - \ell_0 - \ell_1 = 0$; the slack row reports how far each column is from that degeneracy. Column two is exactly degenerate and admits no test.

Column two drops the 101 observations at or above the nested critical share, where the nested cost inversion is not defined; the co-active rows that identify λ are unaffected.

The first two columns agree closely and are the basis for the claim in the main text. It does not matter whether markups respond to shares through substitution across goods or through substitution across stalls: either way ℓ_1 is near 0.8, $\hat{\lambda}$ is statistically indistinguishable from zero, and group-as-one-firm is rejected at $p < 10^{-11}$.

The third column appears to disagree, and Panel C shows that it does not. What the estimator identifies is not λ but the product $\lambda\ell_1$, because λ enters (20) only through χ_{imgt}, which is proportional to ℓ_1. On that common scale the three columns are almost equally precise — standard errors of 0.133, 0.118 and 0.121 — so the data carry the same information about conduct under all three demand systems. Nothing degrades as we move across the table. What changes is the alternative hypothesis. Full internalization means $\lambda\ell_1 = \ell_1$, and at $(\sigma, \eta) = (4.00, 3.30)$ the two elasticities are close enough together that $\ell_1 = 1/\eta - 1/\sigma$ collapses to 0.053, sixteen times smaller than in the first two columns. The cartel benchmark moves sixteen times closer to Cournot while the noise stays put, so the same data can no longer separate them.

The last row of Panel C states that collapse in economic terms. At the mean row with a co-active co-member, moving from individual Cournot to full internalization raises the markup by about 16 log points under across-good demand and 19 under across-stall demand, but by only 1.1 log points at the joint estimates. When stalls are nearly as substitutable as goods, merging two stalls' output barely changes the perceived elasticity either faces, so a group pricing as one firm is almost observationally identical to a group of Cournot competitors. No estimator on this sample distinguishes a one-percent markup change from zero, and none should be expected to.

We therefore read the third column as a statement about the demand system rather than about conduct. It does not say the data are consistent with collusion; it says that under this parameterization the collusive benchmark is not an economically distinct hypothesis, so failing to reject it carries no information. What no parameterization delivers is positive evidence for internalization: $\hat{\lambda}\ell_1$ is negative in all three columns, and its confidence interval excludes the cartel value ℓ_1 in both columns where that value is economically meaningful.

Why these fixed effects. The estimating equation absorbs a retailer × good effect and a week effect, and nothing more. The retailer × good effect is not a robustness choice: it is what sweeps out $\log a_{ig}$,

the unobserved multiplicative cost wedge, and the model's identifying content is exactly the assumption that this wedge is time-invariant. Adding a market × good × week effect on top of it is tempting, since it would absorb cell-level cost and demand shocks, and it does tighten the joint column from a standard error of 2.29 to 1.89. We do not use it, because it breaks the estimator in a way the first column makes visible: under cell effects $\hat{\lambda}$ moves to −0.30 (0.13) there, significantly negative, and the same happens in the joint column. A negative conduct parameter says retailers place negative weight on a co-member's output — competing more aggressively against a sourcing partner than against a stranger — which is not a hypothesis the model contains. The reason is algebraic. Conditional on the own-share term, cell effects reduce the estimator exactly to $\lambda = -\partial \log w / \partial x$, the within-cell covariation of the wholesale price with co-member share. That covariation is the bulk discount, which the experiment moves by construction, so the cell-effect specification isolates the confound rather than the parameter. The measured gradient is +0.24 (0.12) and biases $\hat{\lambda}$ downward, which is the sign the cell-effect estimates show.

The overidentification placebo. Under the model the non-group rival share must enter with coefficient zero, and Panel D reports that test. The conduct regressor and its placebo sum to $\ell_1(1-s)/(1-\ell_0-\ell_1 s)$, so the two are collinear by construction and exactly collinear when $1-\ell_0-\ell_1 = 0$. The across-stall column is exactly degenerate and admits no test. In the homogeneous column, the placebo estimate is +4.74 ($t = 4.47$); in the joint column it is −6.76 ($t = -2.00$). Thus the restriction fails where it can be evaluated, indicating that something in $\log(\widehat{c}^{w}/w)$ moves with in-cell shares beyond the channel the model allows. This is a general caveat on the estimator, not evidence in favor of a particular demand parameterization; the conduct conclusion rests on the distance between the estimate and the economically meaningful full-internalization benchmark.

E.2 Further robustness of the conduct estimate

The rejection of full group internalization is not tied to one cost wedge or implementation choice. Table E.2 varies the cost specification, baseline formation nuisance, fixed effects, sample, clustering level, and exact versus linearized estimator. The multiplicative row in Panel A reproduces the paper's headline estimate. Panels B–F instead use a power-cost reference specification with $\beta = 0.600$ so each panel changes one feature at a time.

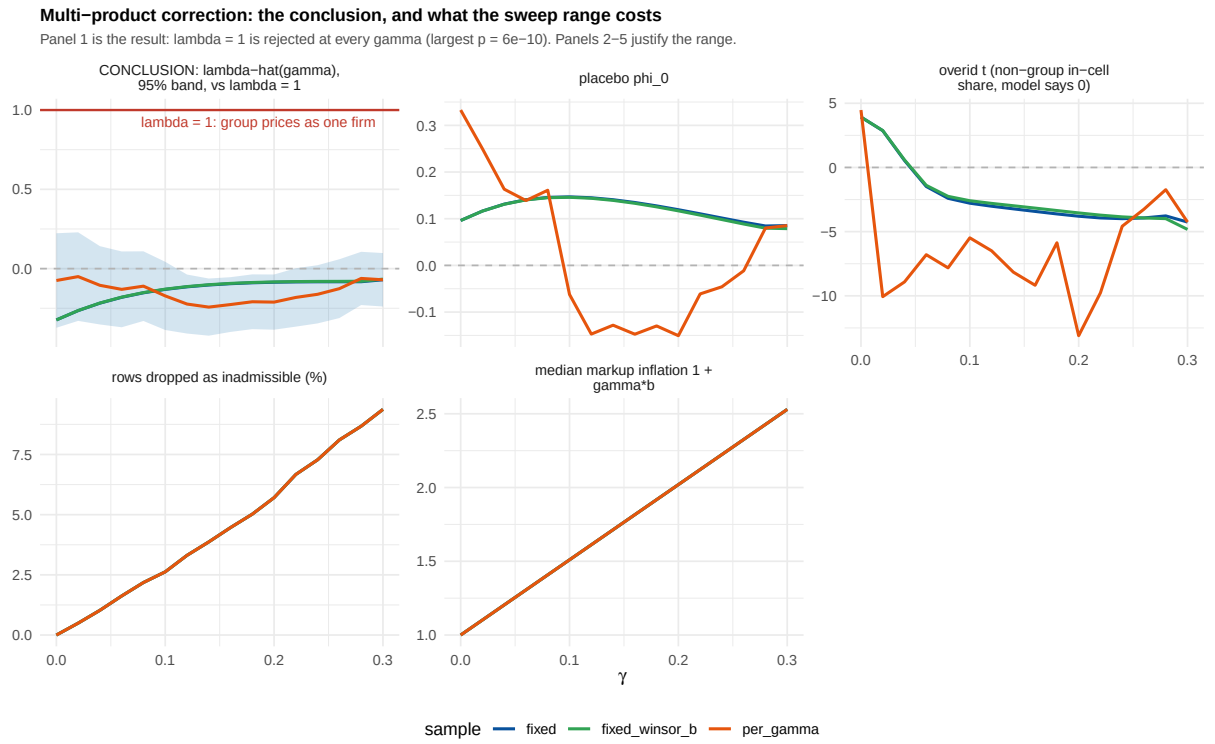
Table E.2: Robustness of the conduct parameter

Specification	$\hat{\lambda}_1$	SE	$p(\lambda = 0)$	$p(\lambda = 1)$	N
A. Cost specification					
Additive $c = w + z_{ig}$	0.491	0.191	0.010	0.008	6,181
Multiplicative $c = a_{ig} w$	-0.075	0.152	0.622	< 0.001	6,181
Power $c = a_{ig} w^{\beta}$, β estimated	0.128	0.093	0.172	< 0.001	6,181
$\beta = 0.542$ imposed (DiD ratio)	0.021	0.112	0.848	< 0.001	6,181
$\beta = 0.600$ imposed	0.009	0.116	0.936	< 0.001	6,181
Cobb–Douglas $c = A_{ig} w^{\alpha} q^{1-\alpha}$	-0.071	0.152	0.641	< 0.001	6,181
$\alpha = 0.600$ imposed	-0.071	0.160	0.659	< 0.001	6,181
B. Baseline nuisance					
ϕ_0 estimated (reference for Panels B–F)	0.009	0.116	0.936	< 0.001	6,181
$\phi_0 = 0$ imposed	-0.123	0.100	0.221	< 0.001	6,181
$\lambda_1 - \phi_0$ (baseline differenced out)	-0.284	0.113	0.012	< 0.001	6,181
C. Fixed effects					
Retailer–good + wave (canonical)	0.009	0.116	0.936	< 0.001	6,181
Retailer + good + wave	0.010	0.107	0.926	< 0.001	6,181
Retailer–good + market×good×wave	-0.205	0.087	0.019	< 0.001	6,033
Retailer + market×good×wave	-0.161	0.068	0.018	< 0.001	6,049
D. Sample					
Full sample	0.009	0.116	0.936	< 0.001	6,181
No-switch (baseline contract status)	0.148	0.137	0.279	< 0.001	4,851
No relational contract	0.080	0.124	0.519	< 0.001	4,745
E. Inference (clustering)					
Retailer (randomization unit, canonical)	0.009	0.116	0.936	< 0.001	6,181
Sourcing group (`source_id`)	0.009	0.140	0.947	< 0.001	6,181
Market×good	0.009	0.126	0.941	< 0.001	6,181
Market×good×wave	0.009	0.124	0.940	< 0.001	6,181
F. Estimator					
Exact NLS (reference for Panels B–F)	0.009	0.116	0.936	< 0.001	6,181
Linearised: $-\log(1 - \lambda x) \approx \lambda x$	0.002	0.119	0.986	< 0.001	6,181

Notes: Every row re-estimates the post-treatment conduct parameter λ_1 on one exposure construction, $x = \kappa D/(1 - A - \kappa s)$, where D is the in-cell share held by the retailer's other sourcing-group members; $\lambda = 0$ is individual Cournot and $\lambda = 1$ is the group maximizing joint profit. The multiplicative specification in Panel A reproduces the paper's headline estimate. Unless the panel says otherwise, Panels B–F use the power wedge $c_{igt} = a_{ig} w_{igt}^{\beta}$ with β imposed at 0.600, retailer–good and wave fixed effects, IPCW weights, the baseline nuisance ϕ_0 estimated freely, and standard errors clustered at the retailer, the randomization unit. Panel E varies only the clustering, so its rows share a point estimate. Across all 23 specifications $\hat{\lambda}_1$ lies in $[-0.284, 0.491]$, the weakest rejection of $\lambda = 1$ is $p = 0.008$, and $\lambda = 0$ is rejected at the five percent level in 4 of 23 rows. The table is therefore sharp against joint profit maximization and weak against individual Cournot: the conduct terms remove well under one percent of the within retailer–good variance, so failing to reject $\lambda = 0$ is not evidence for it. $\hat{\phi}_0$ is the same gradient measured at baseline, when no sourcing group operates; it is estimable because `source_id` is the midline roster carried back to $t = 0$. The additive row is the weakest specification and is reported for completeness: its own baseline placebo is 0.613 (SE 0.255) against the zero the model requires, so it fails the same pre-treatment test its post-period coefficient is being read from. It also rejects both $\lambda = 0$ and $\lambda = 1$, and is therefore not evidence for the preferred conduct interpretation.

Across 23 specifications, $\hat{\lambda} \in [-0.284, 0.491]$; all reject $\lambda = 1$ at 5 percent, with the weakest rejection at $p = 0.008$. Only four reject $\lambda = 0$, so the table rejects joint profit maximization without establishing exact individual Cournot. The additive-cost row is the weakest: it rejects both benchmarks and fails its own baseline placebo, 0.613 (s.e. 0.255), violating the pre-treatment restriction needed for its post-period interpretation.

The multiproduct exercise in Fig. 3 allows a retailer's quantity choice for one good to affect margins on her other goods. Figure E.1 documents what is required to sweep the cross-good spillover γ from zero to 0.30. The model-implied restrictions $c_{imgt}(\gamma) > 0$ and $x_{imgt}(\gamma) < 1$ are slack at $\gamma = 0$ but remove 9.4 percent of rows at $\gamma = 0.30$; over the same range, the median implied markup rises by a factor of 2.53. These are economically substantial corrections, yet the largest p-value against $\lambda = 1$ is only 6.2×10^{-10}.



Notes: The first panel repeats the conduct conclusion; the remaining panels report the baseline formation placebo, the non-group-share overidentification statistic, the share of rows excluded by the model's admissibility restrictions, and the median markup inflation. Blue holds the sample fixed over the full range of γ, green additionally winsorizes the other-goods revenue ratio, and orange uses the model-admissible sample at each γ.

Figure E.1: Multiproduct-conduct diagnostics

The fixed-composition estimate at $\gamma = 0$ is -0.323 (s.e. 0.214), compared with the headline -0.075 (s.e. 0.152), so a small set of high-share, large-bundle observations affects the point estimate. Both samples nevertheless fail to reject $\lambda = 0$ and decisively reject $\lambda = 1$. Because γ and λ are not separately identified, the sweep is a bounded sensitivity analysis, not an estimate of cross-good substitution. Its implication is narrower and robust: omitted multiproduct terms do not explain the rejection of full internalization.

E.3 The wedge

The conduct parameter reads collusion off quantities. The wedge reads it off costs, and needs no conduct parameter to do so. The Stage 2 FOC for active retailer i in cell (m, g, t) delivers two empirical objects for the same row:

$$\begin{aligned} \hat{c}^{w}_{imgt} &\equiv p_{mgt} \cdot (1 - s_{imgt}/\eta) && \text{Cournot-inverted from observed } p,\ s, \\ \tilde{c}_{imgt} &\equiv \exp\!\left(\widehat{\alpha}_{gm} + \hat{\psi}\ \log Q^{wh,\mathrm{imp}}_{imgt}\right) && \text{wholesale-model prediction; uses no retail prices.} \end{aligned}$$

The first inherits whatever conduct is in prices; the second is built from wholesale transactions and is invariant to retail conduct. Under the model the two agree up to wholesale-stage noise. Define the wedge

$$\delta_{imgt} \equiv \log \hat{c}^{w}_{imgt} - \log \tilde{c}_{imgt}.$$

Under collusion, retailers restrict output, p_{mgt} rises above the Cournot benchmark, $\hat{c}^{w}$ overstates the true cost, and $\delta > 0$. The scale of that rise is set by how far a coalition can move the Lerner index: at the canonical $\hat{\eta} = 1.146$ a cell monopolist prices at $1/\eta = 0.87$ against an observed mean Lerner index of 0.075, so the room between competitive and collusive pricing is wide. Section F converts that room into a concrete benchmark by injecting coalitions into the data and re-solving for prices; the smallest signal any of the three injected coalitions produces is $\hat{\beta}_2 = +0.061$.

E.4 Regression

Let $\mathrm{coord}_{imgt} = \mathbf{1}\{\exists j \in M_i \setminus \{i\} : a_{jmgt} = 1\}$ be the coordination indicator (a peer in i's source is co-active in the same cell). Estimate

$$\delta_{imgt} = \alpha_{mg} + \alpha_t + \beta_1 (T_i \times \mathrm{midline}_t) + \beta_2 (T_i \times \mathrm{midline}_t \times \mathrm{coord}_{imgt}) + \gamma\, \mathrm{coord}_{imgt} + \varepsilon_{imgt} \tag{25}$$

on active-seller rows where $\hat{c}^{w}$ is defined, with $(m \times g)$ and t fixed effects, IPCW weights, and standard errors clustered at the retailer level. The null is $H_0 : \beta_2 = 0$ (no collusion); the alternative is $H_1 : \beta_2 > 0$. A robustness specification replaces coord_{imgt} with the pre-treatment $\mathrm{coord}^{\mathrm{pre}}_{img} = \mathbf{1}\{\exists j \in M_i \setminus \{i\} : a_{jmg,0} = 1\}$, fixed at $t = 0$ to break the mechanical endogeneity of the active set to treatment.

E.5 Results

The estimating sample contains 8,027 active-seller rows with mean $\bar{\delta} = 0.005$ and standard deviation 0.263. The realized coordination rate is 32.7%; the pre-treatment coordination rate is 30.3%. Table E.3 reports the coefficients across treatment arms and coordination definitions.

Table E.3: Wedge regression (25)

Specification	N	$\hat{\beta}_1$	$\hat{\beta}_2$	95% CI on β_2	$\hat{\gamma}$
Panel A: realized coordination, $coord_{imgt}$					
Pooled	7,993	−0.009	+0.012	[−0.009, 0.033]	+0.034
		(0.010)	(0.011)		(0.008)
T2	3,403	+0.000	−0.003	[−0.030, 0.025]	+0.042
		(0.011)	(0.014)		(0.010)
T3	4,178	−0.017	+0.023	[−0.011, 0.058]	+0.030
		(0.012)	(0.018)		(0.012)
Panel B: pre-treatment coordination, $coord^{pre}_{img}$					
Pooled	7,993	−0.006	+0.006	[−0.015, 0.027]	+0.035
		(0.010)	(0.011)		(0.008)
T2	3,403	+0.002	−0.009	[−0.037, 0.019]	+0.047
		(0.011)	(0.014)		(0.010)
T3	4,178	−0.013	+0.015	[−0.019, 0.048]	+0.033
		(0.012)	(0.017)		(0.012)

Notes: Coefficients from (25). The dependent variable is the wedge $\delta_{imgt} = \log \hat{c}^{w}_{imgt} - \log \tilde{c}_{imgt}$ on active-seller rows. β_1 is the coefficient on $T_i \times \text{midline}_t$, β_2 the coefficient on $T_i \times \text{midline}_t \times \text{coord}_{imgt}$ — the collusion signal — and γ the level coefficient on the coordination indicator. All specifications include $(m \times g)$ and t fixed effects and IPCW weights. Standard errors in parentheses, clustered at the retailer level; confidence intervals are $\hat{\beta}_2 \pm 1.96\,\widehat{\text{se}}$. Panel B replaces the realized coordination indicator with its pre-treatment counterpart, fixed at $t = 0$. The IV/LATE version of the pooled realized specification gives $\hat{\beta}_2 = -0.008$.

Reading. $\hat{\beta}_2$ is statistically zero in every specification: the largest estimate is +0.023 and every p exceeds 0.18. The realized and pre-treatment versions agree in sign and order of magnitude, so the result is not an artifact of the active set responding to treatment. The widest upper confidence bound is +0.058 (T3, realized coordination), well below the smallest signal any injected coalition produces in Section F. The level coefficient $\gamma \approx +0.03$ to +0.05 ($p < 0.01$) reflects cell-time selection that the $(m \times g)$ fixed effect does not fully absorb — group members co-locate in cells with slightly different cost structure — but it does not move with treatment, so it does not contaminate $\hat{\beta}_2$.

The data reject tacit retail collusion among coordinating group members. The model's no-collusion Stage 2 is empirically defensible, and the search-cost coefficient γ_T is not absorbing a hidden collusive markup.

E.6 Pass-through

If treated retailers banked their procurement savings as margin, the wholesale-cost cut would not reach consumers. The treatment-induced $\mathrm{d}\log p/\mathrm{d}\log w$ is 0.50 (s.e. 0.26, ratio of difference-in-differences) and 0.69 (s.e. 0.33, 2SLS), both statistically indistinguishable from the competitive benchmark $\alpha_c \approx 0.6$. The benchmark is α_c, not one, because wholesale cost is only a fraction $\alpha_c = w/c \approx 0.6$ of marginal cost: by the chain rule $\mathrm{d}\log p/\mathrm{d}\log w = (\mathrm{d}\log p/\mathrm{d}\log c)(\mathrm{d}\log c/\mathrm{d}\log w) = 1\cdot\alpha_c$, where the first factor is one because under CES–Cournot a constant multiplicative markup passes through fully in logs.[25] The estimates bracket α_c, so pass-through is complete and savings reach consumers.

The direct estimates above use wholesale cost as the denominator. Table E.4 instead expresses the same experiment as pass-through to conduct-inverted marginal cost, $\mathrm{d}\log p/\mathrm{d}\log \widehat{c}$, under the two conduct benchmarks. The observed wholesale-price response and first stage are identical across columns; only the share entering the Cournot cost inversion changes.

Under individual conduct, the cost elasticity is $\hat{\psi} = -0.0697$ and the implied pass-through is 1.298; complete pass-through cannot be rejected ($p = 0.612$). Under full group internalization, the cost elasticity becomes -0.2218 and pass-through falls to 0.408, rejecting complete pass-through ($p = 0.008$). This is an internal-consistency diagnostic rather than separate identification of conduct: replacing own share with group share mechanically relabels inferred marginal cost. Its value is that full internalization makes the structural cost and pass-through results difficult to reconcile, whereas individual conduct does not.

This is corroborating rather than decisive, and for a reason worth stating: the same CES property means a constant cartel markup also passes through fully in logs. Pass-through therefore detects treatment-induced margin widening, not a static cartel — which is precisely the gap the λ test of Section 7 closes, since λ is identified from the cross-sectional co-member share rather than from a time difference.

[25] Under homogeneous-good CES–Cournot the cell price $p^* = \eta \sum_k c_k/(\eta N - 1)$ is homogeneous of degree one in the cost vector, so a proportional cost change moves price one-for-one in logs.

Table E.4: The wholesale schedule under individual and full group internalisation

	Conduct assumption in the Cournot inversion	
	Individual ($\lambda = 0$)	Full group ($\lambda = 1$)
	own share s_{imgt}	group share S_{imgt}
A. Cost elasticity: DiD-IV of $\log \hat{c}$ on $\log Q^{wh}$		
$\hat{\psi}$	-0.0697^{*} (0.0367)	-0.2218^{***} (0.0491)
Difference (col. 2 − col. 1)	-0.1521 (0.0246)	
B. Implied pass-through $\mathrm{d}\log p/\mathrm{d}\log \hat{c}$		
$\hat{\rho}$	1.298 (0.645)	0.408 (0.180)
Wald 95% CI	[+0.03, +2.57]	[+0.05, +0.76]
Anderson–Rubin 95% set	$[-0.04, \infty)$	[+0.00, +0.78]
p-value, $H_0 : \rho = 1$	0.612	0.008
C. Common to both columns		
Wholesale price elasticity $\hat{\beta}$	-0.0712 (0.0829)	
First stage, $\mathrm{d}\log Q^{wh}/\mathrm{d}(D \times \mathrm{Post})$	0.6034 (0.0942)	
First-stage F	41.0	
Cells / sources	4,799 / 517	

Notes. Source×market×good×week panel, restricted to sources whose pre-existing wholesale contract access is unchanged by group formation (Spec C1). Every column is the same DiD-IV: the outcome on $\log Q^{wh}_{smgt}$, instrumented by $D_s \times \mathrm{Post}_t$, with market×good and source fixed effects, source-IPCW weights and standard errors clustered at the source. Q^{wh} is the per-member mean at baseline and the within-source sum thereafter. The two columns differ ONLY in the share entering the Cournot inversion $\hat{c}_{imgt} = \bar{p}_{mgt}(1 - \cdot/\sigma)$: the retailer's own in-cell share, or the summed in-cell share of all active members of her sourcing group, switched on post-treatment. Column 1 reproduces the paper's $\hat{\psi}$. **Panel C is not repeated by column because it cannot differ**: $\hat{\beta}$'s outcome is the observed wholesale price, which contains no share, so re-defining the inversion leaves that regression numerically unchanged. **The difference in Panel A is a re-labelling, not new information**: since the instrumental-variables estimator is linear in the outcome and the two columns share the same cells, the gap equals the DiD-IV of $\log[(1 - S/\sigma)/(1 - s/\sigma)]$ on the same instrument, which is what the difference row reports. $\hat{\rho}$ is the ratio of the retail-price and cost responses to the same instrumented volume, estimated as one IV of $\log p$ on $\log \hat{c}$. Its relevant null is complete pass-through, so p-values are for $H_0 : \rho = 1$ and are computed from the reduced form of $\log p - \log \hat{c}$ on the instrument, in which the first stage cancels and no weak-instrument correction applies. $\hat{\rho}$ is a ratio with a weakly determined denominator in column 1, so its Anderson–Rubin set is unbounded above; the AR sets are the weak-identification authority and the first-stage F is the cluster-robust squared t on the excluded instrument, reported for reference. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

F Alternative Coalition Structures

The tests of Section E ask whether the sourcing group colludes. This appendix asks a prior question — how large a coalition the treatment arms could seed at all — and then uses the answer to measure the power of the wedge test. Both exercises run over the same three coalition scales, so the concentration a scale implies and the signal it would leave in the wedge can be read against each other.

The relevant benchmark is not a single market-wide cartel. Retailers are not specialized and markets are fragmented: a typical market×good×week cell has about seven active sellers and a median within-good share of only 5.7%. Under asymmetric Cournot competition each seller's Lerner index is firm-specific, s_{imgt}/η, so small shares mechanically imply small markups; at the canonical $\hat{\eta} = 1.146$ a perfect cartel internalizes all cross-effects and prices at the joint-monopoly Lerner index $1/\eta = 0.87$, against an observed mean of 0.075. A partial coalition lives in between.

F.1 Three scales of treatment-induced coalition

The treatments could seed coalitions at three scales, each collapsing a different set of sellers into one player. Within-arm coalitions let the retailers assigned to T_2 and to T_3 collude separately, which is the largest bloc the randomization itself could create. Across-group coalitions pool every group-affiliated retailer in the cell, which ignores group boundaries and is therefore the most concentrated form the cooperative structure could support. Within-group coalitions — the scale the coordination meetings actually operate at — merge the members of the same CP group who sell the same good in the same cell. Table F.1 reports what each does to the cell.

Table F.1: Three scales of treatment-induced coalition

Scale	Sellers who collude within a cell	Players/cell	1/HHI
Observed	—	7	5.7
Within-arm	retailers assigned to T_2 and to T_3 collude separately	3	3.0
Across groups	all group-affiliated retailers in the cell pool together	4	3.2
Within group	members (≥ 2) of the same CP group on a shared good	5	4.8

Notes: Simulation over the 559 midline market×good×week cells; medians across cells reported. A coalition forms only where it would hold at least two active sellers in the cell; all other sellers remain singletons. "Players/cell" counts a colluding coalition as a single player; 1/HHI is the effective number of competitors, computed on post-merger equilibrium shares from the isoelastic Cournot solver at $\hat{\eta} = 1.146$. Even the most concentrated scale leaves a median of three players, far from monopoly.

Every scale leaves several independent competitors in the median cell. The most concentrated of the three — within-arm — still leaves three players and an effective number of competitors of 3.0, and the scale the coordination meetings actually operate at leaves five.

F.2 Simulated-cartel validation of the wedge test

A null is informative only if the test could detect collusion. We therefore inject each coalition of Table F.1 into the midline data as a quota cartel — one player that keeps all its members producing, at the within-coalition share-weighted mean of their marginal costs — re-solve the cell for the Cournot equilibrium price, recompute $\hat{c}^w$ from the simulated price and shares, and re-run (25) on the resulting wedge. Baseline rows are never modified, so the difference-in-differences shape that identifies β_2 survives the injection. The imputed cost $\tilde{c}$ is left untouched, since it is a wholesale object and no retail cartel moves it. As a check on the harness, running the solver with no coalition returns the observed price to within 6×10^{-14}.

Table F.2: Simulated-cartel validation of the wedge test

Scenario	Coalition cells	Colluder rows	$\Delta \log p$ (colluders)	$\hat{\beta}_2$ (wedge)	Lerner, obs. → sim.
Real data	—	—	—	+0.012 (0.011)	0.075
Within-arm	408	3,949	+0.172	+0.061 (0.014)	0.082 → 0.200
Across groups	350	2,393	+0.288	+0.221 (0.021)	0.075 → 0.170
Within group	301	1,912	+0.052	+0.081 (0.011)	0.069 → 0.096

Notes: Each simulated row injects the corresponding coalition of Table F.1 into the midline cells as a quota cartel and re-solves the isoelastic Cournot equilibrium at $\hat{\eta} = 1.146$; baseline rows are unchanged. $\Delta \log p$ is the mean log price rise among colluding sellers. $\hat{\beta}_2$ is the coefficient on $T_i \times \text{midline}_t \times \text{coord}_{imgt}$ from (25), re-estimated on the simulated wedge with the same $(m \times g)$ and t fixed effects, IPCW weights and retailer clustering; standard errors in parentheses. The real-data row is the pooled realized-coordination specification of Table E.3. The final column reports the mean market-wide Lerner index on the affected cells, observed and simulated, against a joint-monopoly benchmark of $1/\eta = 0.87$. No cell required a seller to exit in any scenario.

Reading. The injection is genuine: it raises colluder prices by 5 to 29 log points and roughly doubles the market-wide Lerner index in the two market-wide scenarios. The wedge test sees it. Every injected coalition turns $\hat{\beta}_2$ positive and significant, from +0.061 for the within-arm cartel to +0.221 for the across-group cartel, against +0.012 (s.e. 0.011) in the real data. The gap is not marginal: the smallest injected signal is more than four standard errors above the real-data estimate, and it lies outside the real-data confidence interval, whose upper bound is +0.033.

The within-group scenario is the one that matters most for the model, because it is the coalition the coordination meetings could plausibly organize. It is also the mildest — colluder prices rise by only 5 log points and the effective number of competitors falls from 12.7 to 9.1 — and the wedge test still detects it at $\hat{\beta}_2 = +0.081$ (0.011), seven times the real-data estimate. The test has power against exactly the coalition the treatment could have created, and it does not fire.

G Across-Stall Elasticity σ: Identification from the Retailer Panel

The nested-CES alternative uses an across-stall elasticity estimated from the retailer panel rather than a calibrated value. This appendix presents that estimation, which uses cost-side instruments for the endogenous retailer-level price. The canonical homogeneous-good model instead uses the across-product elasticity η estimated in Section 6.1.

Specification. Within each market m and good g, retailers r compete and households allocate expenditure across them. Under the CES demand system, retailer r's within-cell revenue share satisfies the log-linear inversion

$$\log s_{rgm} \;=\; \alpha_{gm} \;+\; (1-\sigma)\,\log p_{rgm} \;+\; \varepsilon_{rgm}, \qquad s_{rgm} \;=\; \frac{p_{rgm}\,q_{rgm}}{\sum_{r'\in gm}\,p_{r'gm}\,q_{r'gm}}, \tag{26}$$

where α_{gm} is a (g,m) fixed effect that absorbs the CES demand shifter A_{gm} and any common cost level, and ε_{rgm} is a stall-level demand shock. OLS on (26) is biased because p_{rgm} responds to ε_{rgm}.

Instruments. We instrument $\log p_{rgm}$ with three cost shifters: (i) the retailer's own wholesale unit price for (g,m) in the same week; (ii) the leave-one-out mean wholesale price for (g,m) across the retailer's market-cohort, which captures common upstream cost shocks; and (iii) the retailer's reported transport cost per trip. The exclusion restriction is that, conditional on (g,m) fixed effects, the cost shifters affect the retailer-level revenue share only through the equilibrium price. Each shifter is a procurement-side primitive — wholesale prices are set at Bromakoté for buyers downstream of the cell, and transport prices are set by taxi-fare schedules unrelated to stall-level demand.

Estimates. Table G.1 reports OLS and 2SLS estimates of σ on two independent survey waves: W1 collected in 2023 (1,008 retailers, $N = 3{,}432$ stall-good observations) and W2 collected in 2024 (710 retailers, $N = 2{,}676$). OLS yields biased estimates well below one ($\hat{\sigma}^{\mathrm{OLS}} = 0.78$ in W1 and 0.57 in W2), consistent with positive correlation between unobserved demand and price. The cost-shifter IV recovers

$$\hat{\sigma}^{\mathrm{IV}}_{\mathrm{W1}} \;=\; 2.03 \quad (0.44), \qquad \hat{\sigma}^{\mathrm{IV}}_{\mathrm{W2}} \;=\; 2.08 \quad (0.61),$$

with first-stage F statistics of 11.6 (W1) and 7.3 (W2). The two waves yield statistically indistinguishable estimates, both close to 2, and a Wald test does not reject $\sigma = 2$ at the 10% level in either wave.

Table G.1: Across-Stall Elasticity σ from the Retailer Panel

	W1 (2023)	W2 (2024)
OLS $\hat{\sigma}$	0.78	0.57
	(0.04)	(0.06)
IV $\hat{\sigma}$	**2.03**	**2.08**
	(0.44)	(0.61)
First-stage F	11.6	7.3
p-value Wald test $H_0 : \sigma = 2$	0.946	0.896
$g \times m$ FE	Yes	Yes
Observations	3,432	2,676
Retailers	1,008	710

Source: Retailer panel survey, W1 (2023) and W2 (2024) good-level rosters, baseline waves.

Notes: The table reports OLS and 2SLS estimates of the demand elasticity σ from (26), with the dependent variable $\log s_{rgm}$ regressed on $\log p_{rgm}$ and $(g \times m)$ fixed effects. The instruments for $\log p_{rgm}$ are (i) the retailer's own wholesale unit price for (g, m) in the same week, (ii) the leave-one-out mean wholesale price for (g, m) across the retailer's market-cohort, and (iii) the retailer's reported transport cost per trip. Standard errors (in parentheses) are clustered at the retailer level. First-stage F is the Kleibergen-Paap robust statistic. The Wald test reports the p-value of $H_0 : \sigma = 2$ against the two-sided alternative. The mapping from the coefficient on $\log p$ to σ is $\sigma = 1 - \hat{\beta}$.

Reading. The retailer-panel IV pins σ to a value close to 2, independently in two waves separated by one year. This estimate concerns the across-stall elasticity in the nested-CES alternative. The canonical homogeneous-stall model instead uses the across-product elasticity η; Appendix B reports the wholesale-elasticity estimate $\hat{\psi}$ for $\eta \in \{1.5, 2.5\}$ as a further robustness check.

H Stochastic-choice Bounds on Net Retailer Surplus

Section 8 distinguishes gross variable profit, deterministic net retailer surplus, and the inclusive value generated by the stochastic activation model. Conditional on the model, these objects are in FCFA: expected profit is measured in FCFA, its coefficient identifies the activation-shock scale τ, and the setup-cost index is consequently recovered in the same units. The remaining issue is uncertainty in those estimates, not an unidentified money normalization.

H.1 Why deterministic sign inequalities do not apply

For a non-incumbent, the activation rule is

$$a_{imgt} = 1 \iff \bar{\pi}_{imgt} - S_{imgt} - \tau\varepsilon_{imgt} \geq 0, \qquad \varepsilon_{imgt} \sim \text{Logistic}(0,1). \tag{27}$$

It follows that neither $a_{imgt} = 1 \Rightarrow \bar{\pi}_{imgt} \geq S_{imgt}$ nor its converse is valid: the private cost realization can reverse the sign of the deterministic payoff difference. A moment inequality built from those implications would silently discard the shock that generates the activation likelihood and cannot be combined with the model estimated in Section 6.3.

The valid revealed-preference restriction concerns the conditional choice probability. For an interior probability,

$$P^*_{imgt} = \Lambda\left(\frac{\bar{\pi}_{imgt} - S_{imgt}}{\tau}\right), \qquad S_{imgt} = \bar{\pi}_{imgt} - \tau \operatorname{logit} P^*_{imgt}. \tag{28}$$

Thus uncertainty about the setup cost must be propagated jointly from expected profit, the fitted choice probability, and τ. There is no scale-free bound that can validly remove τ while retaining the stochastic activation model.

H.2 Projection bound

Let θ collect the demand, cost, activation-scale, and setup-cost parameters, and let $\Theta_{1-\alpha}$ denote a joint confidence or sensitivity set for them. For every $\theta \in \Theta_{1-\alpha}$, the implementation must repeat the operations used by the headline counterfactual: reconstruct profits, re-profile the cell effects α_{mg}, solve the activation fixed point under each channel configuration, and calculate the welfare object. For $H \in \{RP^{\text{net}}, RP^{\text{incl}}, W^{\text{incl}}\}$, define

$$\underline{\Delta H} = \inf_{\theta\in\Theta_{1-\alpha}} \left\{H(E_{11};\theta) - H(E_{00};\theta)\right\}, \qquad \overline{\Delta H} = \sup_{\theta\in\Theta_{1-\alpha}} \left\{H(E_{11};\theta) - H(E_{00};\theta)\right\}. \tag{29}$$

This projection is conservative when $\Theta_{1-\alpha}$ is a rectangular product of marginal confidence sets; a joint likelihood or bootstrap region is preferable when available. It also makes the welfare interpretation transparent. If ε_{imgt} is genuine setup-cost heterogeneity, the relevant retailer object is the inclusive

value in (**??**). If it is decision error, the relevant object is expected deterministic net surplus $RP^{\text{net}} = \sum P^{*}_{imgt}(\bar{\pi}_{imgt} - S_{imgt})$. The old one-sided numerical bounds are not reported because they relied on deterministic sign implications contradicted by (27); numerical projection bounds require rerunning the model over $\Theta_{1-\alpha}$.

I Notation Dictionary

Symbol	Definition
Indexing	
i	Retailer.
m	Market (12 marchés in Bouaké).
g	Good.
t	Week.
s	Source: group id if retailer is in a group, retailer id otherwise.
M_s	Set of retailers in source s.
$\mathcal{A}_{mgt}$	Active set in cell (m, g, t): $\{i : a_{imgt} = 1\}$.
Stage 2 (Cournot–CES)	
η	Across-product elasticity. It is the single demand elasticity in the canonical homogeneous-stall model, where $\hat{\eta} = 1.1455$ (s.e. 0.4934), and the outer elasticity under nested CES, where the joint supply-moment estimate is $\hat{\eta} = 3.30$.
σ	Across-stall elasticity used only under nested CES; its joint supply-moment estimate is $\hat{\sigma} = 4.00$.
A_{mgt}	CES demand shifter.
p_{mgt}	Cell-level retail price (homogeneous-good Cournot).
s_{imgt}	Retailer's within-cell quantity share in the homogeneous model; the nested model instead uses revenue share r_{imgt}.
q_{imgt}	Retail quantity sold by retailer i.
q^{wh}_{imgt}	Individual wholesale quantity purchased by retailer i.
Q^{wh}_{smgt}	Source-level wholesale quantity, $\sum_{j \in M_s} q^{wh}_{jmgt}$.
$w(Q^{wh}_{smgt})$	Posted wholesale unit-price schedule; $w'(\cdot) < 0$.
w_0	Level of the schedule in $w(Q) = w_0 Q^{\psi}$.
ψ	Wholesale discount elasticity, $\partial \log w / \partial \log Q^{wh}$. Preferred $\hat{\psi} = -0.0697$ (s.e. 0.0367); Anderson–Rubin 90% set $[-0.130, -0.020]$.
κ	Wholesale-to-retail conversion factor, $q_{imgt} = \kappa \cdot q^{wh}_{imgt}$.
c_{imgt}	Model-implied retail marginal cost, $w(Q^{wh}_{smgt})/A$.
$\hat{c}^{w}_{imgt}$	Cournot-inverted cost, $p_{mgt}(1 - s_{imgt}/\eta)$ in the homogeneous model and (32) under nested CES. Reduced-form outcome.
$\tilde{c}_{imgt}$	Imputed retailer-level marginal cost; equals $\hat{c}^{w}_{imgt}$ on seller rows.
Identification of ψ	
D_s	Treatment assignment indicator (pooled T1–T3) at the source level.

J Alternative Demand Structure: Nested CES

The canonical model in Section 6.1 treats stalls selling the same physical good as perfect substitutes and therefore uses only the across-product elasticity η. This appendix allows stalls to be differentiated CES varieties within a market–good cell. The nested model has an outer, across-product elasticity η and an inner, across-stall elasticity σ, with the economically admissible ordering $1 < \eta \leq \sigma$. Unlike the main demand IV, the exercise below jointly estimates both elasticities from supply-side margin moments.

Nested-Cournot restriction. Let

$$r_{imgt} = \frac{p_{imgt} q_{imgt}}{\sum_{j \in (m,g,t)} p_{jmgt} q_{jmgt}}$$

denote retailer i's revenue share within market–good–period cell (m, g, t). Under nested CES and Cournot competition, its Lerner index is

$$L_{imgt} = \frac{r_{imgt}}{\eta} + \frac{1 - r_{imgt}}{\sigma} = a + b r_{imgt}, \qquad a \equiv \frac{1}{\sigma}, \quad b \equiv \frac{1}{\eta} - \frac{1}{\sigma} \geq 0. \tag{30}$$

The intercept a identifies the across-stall elasticity and the share slope b identifies the difference between the two inverse elasticities. The inequality $b \geq 0$ imposes the nesting restriction $\eta \leq \sigma$.

Joint minimum-distance estimator. We put observed wholesale cost and retail revenue on the same unit basis using the conversion factor $\kappa = 0.740/0.711 = 1.041$ and form the empirical margin $\widetilde{L}_{imgt} = 1 - \kappa w_{imgt}/p_{imgt}$, winsorizing the cost–price ratio at its first and ninety-ninth percentiles. The estimator matches three moments. First, the revenue-weighted aggregate empirical margin is matched to $a + b\overline{H}$, where $H_{mgt} = \sum_i r_{imgt}^2$ is the cell Herfindahl index and $\overline{H}$ is its revenue-weighted mean. Second, the coefficient from a revenue-weighted regression of the cell margin on H_{mgt} with good fixed effects is matched to b. Third, the coefficient on $\log N_{mgt}$ in a retailer-level regression of $\log \widetilde{L}_{imgt}$ on the number of active sellers, with good, market, and period fixed effects, is matched to the corresponding coefficient generated by $\log(a + b r_{imgt})$.

Writing these empirical moments as $\widehat{M}_k$ and their model counterparts as $M_k(a, b)$, we solve

$$(\hat{a}, \hat{b}) = \arg \min_{a>0,\, b\geq 0} \sum_{k=1}^{3} \left(\frac{M_k(a, b) - \widehat{M}_k}{\widehat{\mathrm{se}}(\widehat{M}_k)} \right)^2, \tag{31}$$

where the moment standard errors come from 400 retailer-cluster bootstrap draws. The three empirical moments are 0.2597 for the aggregate margin, −0.1065 for the margin–Herfindahl slope, and −0.0474 for the elasticity of the margin with respect to the number of sellers; their bootstrap standard errors are 0.0105, 0.1007, and 0.0221, respectively.

Results. The constrained minimum-distance estimate is $\hat{a} = 0.2500$ and $\hat{b} = 0.0534$, implying

$$\hat{\sigma} = 4.00, \qquad \hat{\eta} = 3.30.$$

The optimum is interior, so the nesting constraint does not bind. Retailer-cluster bootstrap 95-percent intervals are $[3.64, 4.34]$ for σ and $[2.90, 3.97]$ for η. The overidentification statistic is $J = 2.82$ on one degree of freedom ($p = 0.093$). The separate retailer-side demand IV gives $\hat{\sigma} = 2.03$ (s.e. 0.443). Adding that estimate as a fourth moment moves the joint estimates to $(\hat{\sigma}, \hat{\eta}) = (3.74, 3.35)$ but raises the objective to 19.88 on two degrees of freedom; the demand moment alone adds 17.06 ($p < 0.001$). Thus, the demand-IV and supply-moment information are not mutually consistent, which is why the joint nested estimates are reported as an alternative specification rather than substituted into the canonical model.

The across-stall elasticity alone. The conduct exercise of Section E.1 also needs the parameterization in which stall differentiation is the only source of markup response. Shutting the outer nest down at $\eta = 1$ makes the top tier Cobb–Douglas, so expenditure shares across goods are fixed and $L(s) = 1/\sigma + (1 - 1/\sigma)s$. Profiling the same three supply moments in $a = 1/\sigma$ alone gives $\hat{a} = 0.2111$, so $\hat{\sigma} = 4.74$, with a retailer-cluster bootstrap 95-percent interval of $[4.33,\ 5.05]$. The restriction itself is decisively rejected by these moments: the objective rises to $J = 302.30$ on two degrees of freedom, and the distance test against the unrestricted joint estimate is 299.47 on one degree of freedom. We nevertheless report the conduct estimate at this calibration because $\eta = 1$ is the parameterization under which the across-stall elasticity carries the entire Lerner slope, and the point of Table E.1 is to vary

which elasticity does that work.

Cost inversion. At the joint estimates, the nested Lerner index in (30) remains below one for every feasible revenue share. The associated marginal-cost inversion is

$$\hat{c}^{\text{nest}}_{imgt} = p_{imgt}\left(1 - L_{imgt}\right). \tag{32}$$

Because $\hat{\eta} < \hat{\sigma}$, stalls selling the same good are more substitutable than different products, as required by the nesting structure.

K An Independent Route to τ: The Demand-Shifter IV

The estimator of Section 6.3 identifies the friction scale τ from a timing assumption: beliefs are frozen at baseline, so the expected profit driving entry is predetermined. That assumption is not testable from inside the likelihood. This appendix reports a second strategy that reaches the same parameter from a different direction, using the experiment's demand variation and no timing restriction at all. It is weakly instrumented and we do not use its point estimate; the reason to run it is that its robust confidence set is an independent check on the canonical scale, and the two approaches share only the demand kernel. Agreement between them is therefore informative in a way that a within-likelihood diagnostic cannot be.

Design. Write the setup cost on the logit index scale rather than in logs,

$$S_{imgt} \;=\; \tau\Big(s_{mg} - \gamma_{grp}\, G_{img,0} - \gamma_T\, D_{it} - \gamma_{inter}\,(D_{it} \times G_{img,0})\Big), \tag{33}$$

where s_{mg} are cell intercepts. This is (8) carried onto a linear index: the channel terms are the same three objects on the same predetermined shape and the same difference-in-differences indicator, with the cell effect reappearing as τs_{mg}. The one substantive difference is scale — γ_{inter} here is a setup-cost cut in FCFA, where in (8) it is a proportional one — so the two are not directly comparable as magnitudes. Substituting (33) into the entry margin gives a linear probability model of activation on expected profit and the channel terms, on the non-incumbent margin, with retailer, cell and week effects, clustered by retailer, rescaled onto the index by $1/0.0841$, the inverse mean of $p(1-p)$.

Expected profit is endogenous, so we instrument it with the demand shifter $\log \hat{A}_{mgt}$ of (14). The exclusion restriction is the one already imposed in the model rather than an auxiliary assumption invented for the estimator: demand reaches activation only through profit, so $\log \hat{A}_{mgt}$ shifts the entry margin without entering the setup cost.

Why the point estimate is not used. The instrument is weak, and transparently so. The demand component accounts for only about 4 percent of the within-cell variance of expected profit, and the first stage clustered honestly at the market–good level has $F = 4.6$. We therefore never lean on the two-stage point estimate for the level of τ. What survives weak identification is the Anderson–Rubin set on the raw activation margin,

$$\mathcal{T}_{AR} \;=\; \Big\{\tau = 1/b \;:\; H_0\big[a_{imgt} - b\,\bar{\pi}_{imgt} \perp \log \hat{A}_{mgt} \,\big|\, \mu_{mg}, \lambda_t\big] \text{ not rejected at } 5\%\Big\}, \tag{34}$$

which collects every τ whose implied profit sensitivity the instrument cannot reject. The 95 percent set is [969, 6,785] FCFA, and the canonical $\hat{\tau} = 6{,}536$ lies inside it. The demand-shifter point estimate is $\hat{\tau}^{IV} = 2{,}897$ FCFA; because the instrument is weak, the confidence set rather than this point estimate is the relevant comparison with the canonical scale.

The design also delivers a diagnostic that is exact rather than estimated: the difference-in-differences component of the setup cost is linear in τ,

$$S_{did}(\tau) \;=\; b_{\pi}^{did} \;-\; \tau\, b_{Y}^{did}, \tag{35}$$

where b_{π}^{did} and b_{Y}^{did} are τ-free difference-in-differences components of expected profit and the fitted index, so any candidate τ translates into an implied setup-cost effect without re-estimation.