Qinrui XIAHOU

1120-16 K.K. Leung Building, Pok Fu Lam Road, Hong Kong SAR
(+852) 5316-3598 · qrxiahou@connect.hku.hk

12 September 2026

Professors Emily Breza, Arun Chandrasekhar, and Davide Viviano
Workshop on Development Economics and Networks for Early-Career Researchers
National Bureau of Economic Research
Cambridge, Massachusetts

Dear Professors Breza, Chandrasekhar, and Viviano,

I am writing to apply to the Workshop on Development Economics and Networks for Early-Career Researchers in December 2026. I am a PhD candidate and incoming postdoctoral fellow at the University of Hong Kong, with my fellowship beginning in mid-October. My research spans environmental, development, and organizational economics. I use randomized controlled trials to study economic trade-offs and incentive design, conducting field experiments in supplier networks, stock exchanges, and digital platforms operated by firms and regulators. The workshop's focus on empirical network methods would provide valuable tools for studying how interventions spread through these networks.

I would like to present *Simple Contracts for Busy Firms: Evidence from a Supply Chain Experiment*, joint with Hao Chen and Guojun He. In a 2×2 field experiment among the 155 active core suppliers of a Chinese publicly listed apparel brand, we independently randomize decision rights over suppliers' reporting choices and the scope of third-party verification. Contrary to the predictions of standard delegation and multitasking frameworks, simpler arrangements elicit stronger responses. A mandated indicator set and carbon-targeted verification increase reporting completeness and the verifiability of supporting documentation, whereas delegated indicator choice and comprehensive verification leave both outcomes close to the control-group level. The paper has been invited for submission to a special issue of the *Journal of Development Economics*. Because the project remains under active development, feedback from the workshop would be particularly valuable.

The paper studies how a lead firm reaches suppliers through the contractual terms it sets. Its central question is how the design of those terms shapes the share of a supplier network that meaningfully engages with a new requirement. My other field experiment, *Errors, Efforts, and Earnings: AI-Driven Field Experiment in Suboptimal Labor Supply Decisions*, approaches a related question from the opposite direction. Treating 9,000 taxi drivers in one city changes the market environment they share, so a five-wave staggered rollout measures the general equilibrium spillovers that a simultaneous launch would hide. My postdoctoral work continues and expands these experiments, and what I learn in Cambridge about estimating treatment spillovers would directly inform my ongoing and planned field experiments.

Any of the following three faculty members can provide a recommendation. Professor Guojun He is my primary supervisor and a coauthor on the enclosed paper. Professor Jin Li served on my thesis advisory committee, and Professor Jing Cai served as an examiner of the dissertation.

Prof. Guojun He, Hong Kong Jockey Club Professor of Economics, HKU Business School, The University of Hong Kong. gjhe@hku.hk

Prof. Jin Li, Zhang Yonghong Professor in Economics and Strategy, and Director of the Centre for AI, Management and Organization, HKU Business School. jli1@hku.hk

Prof. Jing Cai, John R. and Marjorie C. Moore Professor in Agricultural and Resource Economics, University of Maryland, and Research Associate at the NBER. cai516@umd.edu

My curriculum vitae, research statement, and working paper are enclosed. Thank you for considering my application.

Sincerely,

Qinrui Xiahou
PhD in Economics, The University of Hong Kong

Qinrui XIAHOU

(+852) 5316-3598 1120-16 K.K. Leung Building, Pok Fu Lam Road, Hong Kong SAR qrxiahou@connect.hku.hk

EDUCATION

The University of Hong Kong, HKU Business School Hong Kong SAR
PhD in Economics 2022 – 2026

- ***Focus:*** Environmental Economics & Development Economics
- ***Thesis Advisory Committee***: Prof. Guojun He, Prof. Jin Li, and Prof. Yanhui Wu
- ***Thesis Examination Committee***: Prof. Jing Cai (UMD), Prof. Uta Schönberg, Prof. Hui Li, and Prof. Bingjing Li
- ***Scholarships:*** Hong Kong PhD Fellowship (HKPFS, 2022-26), HKU Presidential PhD Scholarship (HKU-PS, 2022-26), FBE PhD Research Progress Award (2024-25), FBE PhD Entrance Scholarship (2022)
- ***Grants:*** National Natural Science Foundation of China (NSFC) Young Student Basic Research Program for PhD Candidates (Sole PI, RMB 300,000, 2025-27), MIT J-PAL Project AI Evidence Full RCT Grant (Co-PI, USD 165,000, 2026-27), Research Postgraduate Student Innovation Award (PI, HKD 50,000, 2026), Additional Funding from RGC for Supporting Research Needs of HKPFS Awardees (PI, HKD 19,454, 2024-25)
- ***Leadership:*** Mentor to 2024 YDC Youth Ambassadors, Home and Youth Affairs Bureau of Hong Kong SAR

Yale University, School of the Environment New Haven, CT
Master of Environmental Science 2019 – 2021

- ***Focus:*** Climate Change Economics and Policy (Academic Advisor: Professor Robert O. Mendelsohn)
- ***Leadership:*** Teaching Fellow at Yale Department of Economics, Student Research Fellow at Yale Hixon Center for Urban Ecology, Student Associate of the Yale Climate Change and Health Initiative

Renmin University of China, School of Environment & Natural Resources Beijing, China
Bachelor of Economics of Resources and Environment 2015 – 2019

- ***GPA:*** 3.92/4.0 (Ranked 1/24)
- ***Honors:*** Outstanding Graduate, GCL Scholarship (one of the ten recipients in the whole university)
- ***Leadership:*** Chinese Youth Delegate to the 22nd UN Climate Change Conference (one of ten), Guest Speaker at the International Youth Summit on Energy and Climate (IYSECC 10.0), Head of Press Corps

WORKING PAPERS

He, G., & **Xiahou, Q**. (2026). *The Demand, Supply, and Market Responses of Corporate ESG Actions: Evidence from a Nationwide Experiment in China*. Available at SSRN 5291594. (**Submitted** to the ***Review of Financial Studies***)

- ***Conference Presentations:*** SFA Annual Meeting (Palm Springs, 2026, Scheduled), Future Scholars in Finance Forum (Tsinghua University, 2026), Asia Meeting-China of the Econometric Society (HKU, 2026), NBER Chinese Economy Working Group Meeting (CUHK, 2025), Paris December Finance Meeting (Paris, 2025), CAFM (Seoul, 2025, **Best Paper Award**), Monash Environmental Economics Workshop (Monash University, 2025), Chinese Economy Workshop (HKU, 2025), HEC-HKUST Sustainable Finance Workshop (HKUST, 2025), EEA Congress (Bordeaux School of Economics, 2025), GRASFI (Paris Dauphine University, 2025), CES China Annual Conference (Sun Yat-sen University, 2025), CFRC (Tsinghua University, 2025), Asian Economic Development Conference (PKU, 2025), CCER Summer Institute (PKU, 2025), ICCDS (SJTU, 2025), CAERE (SUFE, 2025), IFABS (Oxford, 2025), and Inaugural ASU-HKU Interdisciplinary Conference (HKU, 2025)

Kang, L., Li, S., **Xiahou, Q.**, & Zhang, P. (2024). *Gone with the Aid: How Paired Assistance Program Affects High-Stakes College Choices*. Available at SSRN 4977681. (**Revise & Resubmit** at the ***Journal of Urban Economics***)

- ***Conference Presentations:*** Jinan-SMU Conference on Urban and Regional Economics (Jinan University, 2025), Sino-Japanese SPACE Workshop (Hitotsubashi University, 2025), Forum for Young Scholars on China's Education Finance Research (PKU, **Best Presentation Award**, 2025), CCER Summer Institute (Peking University, 2025), and Summer Meeting in Urban Economics (PHBS, 2025)

Chen, H., He, G., & **Xiahou, Q**. (2026). *Delegation, Monitoring, and Engagement: A Field Experiment in Supply Chain Governance*. (**Invited to Submit** to the Special Issue of the ***Journal of Development Economics***)

- ***Conference Presentations:*** NBER Organizational Economics Working Group Meeting (Cambridge, 2026, Scheduled), Conference in memory of Professor Elisabeth "Betty" Sadoulet (UC Berkeley & JDE, 2026, Scheduled), IMO–ESF Workshop (IESE Business School, 2026, Scheduled), CCER Summer Institute (PKU, 2026), *Organization Science* Paper Development Workshop (PKU GSM, 2025)

WORK IN PROGRESS

- Errors, Efforts, and Earnings: AI-Driven Field Experiment in Suboptimal Labor Supply Decisions (with Guojun He and Cheng Zheng)
- Climate Change, Migration and the Design of Intercity Transportation Networks (with Guojun He and Shuo Li)

- The "Shadow Principal": Party Organization and the Dual-Agency Problem in Chinese Corporate Governance

AD-HOC JOURNAL REVIEWER

- *Management Science* (MS), *Journal of Environmental Economics and Management* (JEEM), *Journal of Economic Behavior and Organization* (JEBO), *China Economic Review* (CER), *Scientific Reports*

PRE-DOCTORAL PUBLICATIONS

- Fraas, A., Lutter, R., Murphy, J., **Xiahou, Q.**, Potter, J., & Gosling, S. D. (2025). What Are the Effects of Early Childhood Exposure to Environmental Lead on Personality? *Journal of Benefit-Cost Analysis*, 1–15.
- Dolphin, G., & **Xiahou, Q.** (2022). World carbon pricing database: sources and methods. *Scientific Data*, *9*(1), 573.
- **Xiahou, Q.**, Springer, C. H., & Mendelsohn, R. (2022). The effect of foreign investment on Asian coal power plants. *Energy Economics*, *105*, 105752.
- Liu, J., Yin, M., **Xia-Hou, Q.**, Wang, K., & Zou, J. (2021). Comparison of sectoral low-carbon transition pathways in China under the nationally determined contribution and 2° C targets. *Renewable and Sustainable Energy Reviews*, *149*, 111336.
- Jun-Ling, L. I. U., Ke, W. A. N. G., **Xiahou, Q. R.**, Fang-Ming, L. I. U., Ji, Z. O. U., & Ying, K. O. N. G. (2019). China's long-term low carbon transition pathway under the urbanization process. *Advances in Climate Change Research*, *10*(4), 240-249.
- Liu, J., **Xiahou, Q.**, Wang, K., Zou, J., & Kong, Y. (2019). Medium and long-term low-carbon development pathway of China's industrial sector. *China Soft Science (in Chinese)*, (11), 31-41.
- Wang, K., & **Xiahou, Q**. (2017). Progress and Prospects of Global Climate Negotiations after the Paris Agreement. *Journal of Environmental Economics (in Chinese)*, 2(4), 141-152.

RESEARCH AND WORK EXPERIENCE

Resources for the Future Washington, DC
Research Analyst (35 hours/week) June 2021 – August 2022

- Conducted the empirical analysis and co-authored a paper on the long-term effect of childhood lead emissions
- Developed a comprehensive jurisdiction-sector-year level carbon pricing database using GitHub, R, and Python
- Coded a structural model and implemented simulations of the US land use conversion and its carbon implications

Yale School of the Environment New Haven, CT
Research Assistant, Professor Kenneth Gillingham (20 hours/week) September 2019 – May 2021

- Scraped data on rooftop solar suitability and estimated an instrumental variable model for the Solarize Campaign
- Modeled the energy-saving effects of the Clean Energy for America Act for the US Senate Committee on Finance

Yale Department of Economics New Haven, CT
Teaching Fellow for Economics of Natural Resources (20 hours/week) August 2020 – December 2020

- Taught 48 TA sections on microeconomic theories and environmental economics
- Led weekly discussions on environmental issues such as the global oil market and climate change modeling

Yale Emergency Support Team New Haven, CT
Data Analyst (20 hours/week) April 2020 – May 2020

- Designed and developed the interactive COVID-19 curve model for the City of New Haven with three teammates
- Recoded the curve model from Python to R and Excel for the New Haven Mayor's Office

The Environmental Research Center, Duke Kunshan University Jiangsu, China
Research Assistant (Remote), Professor Junjie Zhang (20 hours/week) August 2017 – June 2019

- Prepared a policy memo on rationales, actions, and obstacles of the US withdrawal from the Paris Agreement
- Led the writing of the technical report that evaluated the impact of the US reversing climate change policies

Programme of Energy and Climate Economics, Renmin University of China Beijing, China
Research Assistant, Professor Ke Wang & Ji Zou (20 hours/week) December 2015 – June 2019

- Co-authored three papers on China's long-term low-carbon transition based on energy modeling using LEAP
- Drafted the section Global Climate Governance Regime in the textbook *Economics of Climate Change Mitigation*

CERTIFICATES

- CFA Certificate in Sustainable Investing (Issued 2024, ID: 101743872)
- MITx MicroMasters Program in Statistics and Data Science (2021–2023)

PHD COURSEWORK

The University of Hong Kong, HKU Business School Hong Kong SAR
PhD in Economics (Field of Study: Economics) 2022 – 2026

- ***Microeconomics and Macroeconomics**:* ECON6011 Microeconomic Theory; ECON6012 Macroeconomic Theory; ECON6036 Game Theory and Applications
- ***Econometrics**:* ECON6005 Econometric Theory I; ECON6801 Applied Econometrics
- ***Field Courses**:* ECON6098 Advanced Topics in Labour Economics; STRA6016 Organizational Economics and Business Strategy I; STRA6017 Organizational Economics and Business Strategy II; ECON6010 Monetary Policy: Theory and Practice
- ***Research Seminars**:* ECON6054 Graduate Seminar; GRSC6100 Doing Research @ HKU
- ***Research Training**:* GRSC6101 Responsible Conduct of Research; GRSC6102 Stream-based Responsible Conduct of Research; GRSC6120 Introduction to Thesis Writing; GRSC6140 Advanced Thesis Writing; GRSC6108 Online Transferable Skills Programme
- ***Grading**:* HKU doctoral coursework is assessed on a pass/fail basis; all courses listed above were completed with a grade of Pass

Statement of Research Interests and Motivation

Qinrui Xiahou (HKU Business School)

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

I am a PhD candidate and incoming postdoctoral fellow at the University of Hong Kong. My research interests are environmental economics, development economics, and organizational economics, and I use field experiments to address important economic trade-offs and incentive design problems in the real world. I ask why firms and workers leave valuable opportunities unrealized, and which contracts, disclosure rules, and information close the gap. I conduct field experiments inside institutions that carry requirements and information among parties outside a single hierarchy, including public exchanges, supply chain contracts, and digital platforms.

Paper for the workshop. *Simple Contracts for Busy Firms: Evidence from a Supply Chain Experiment* randomizes the terms of a live supplier program. Our partner is a Chinese publicly listed apparel brand that introduced ESG reporting for its 155 active core suppliers, most of them small manufacturers with limited administrative staff. A 2×2 design varies who selects the reported indicators, the buyer or the supplier, and whether third-party verification covers all submitted indicators or carbon alone. The theories of authority and multitasking favor supplier choice and broad verification. For the typical supplier in this network, both rankings run the other way. A mandated indicator set raises reporting completeness by 10.4 percentage points, 19.4 percent of the control mean, while delegating the same choice moves neither completeness nor documentation. Targeted carbon monitoring does as much as the mandate, and comprehensive monitoring does little. We explain the reversal with a model in which comprehension of a complex contract draws on the same capacity budget as execution.

Several tests support that mechanism. Returns to supplier choice and to targeted monitoring rise with supplier size, output, and data infrastructure, and engagement is stronger where a dedicated employee manages the filing. An endline survey identifies missing data, insufficient staff, and processing time as the binding constraints. Governance thus has to match supplier capacity before unlocking efficiency gains.

Platform experiments and their spillovers. Two further experiments put the same question to different agents. A nationwide experiment brought ESG concerns to more than 4,800 Chinese listed firms through the public investor platforms of the Shenzhen and Shanghai exchanges. Treated firms disclose more, issue more ESG reports, and improve measured performance, financing part of this improvement by cutting advertising. Productive and opaque firms respond most strongly, which fits costly signaling under information asymmetry. A second experiment asks whether inaccurate beliefs or costly search keeps taxi drivers queueing at airports for lower hourly earnings than city cruising pays. A 2×2 design with roughly 9,000 drivers separates personalized earnings feedback from AI forecasting and routing tools, and its five-wave staggered rollout measures the general equilibrium spillovers that simultaneous deployment would hide. J-PAL funded the project this year, and large-scale field testing begins in the coming months.

Regional networks and migration. Two quasi-experimental projects carry the agenda from firms and workers to households. The first estimates how climate change reshapes migration, and how the design of intercity transportation networks shapes that response. The second moves from realized migration to migration intentions, showing that China's paired assistance program shifts high-stakes college choices and the destinations students set for themselves.

Motivation for attending. My work studies institutions that move requirements and information among independent parties, a central topic of this workshop. I would like to learn how researchers estimate spillovers along network links, design experiments under interference, and use network structure to guide targeting. My supply chain experiment studies how far a buyer's requirement reaches within its supplier network, and I would value guidance on supplier interactions such a program may create. My taxi driver experiment uses a staggered rollout to estimate general equilibrium spillovers, and I would welcome feedback on that design. My postdoctoral work will continue this agenda through field experiments with large firms on related questions. Better methods for handling spillovers would improve how I design, randomize, and analyze these studies, where treatment often affects units beyond the assigned group. This workshop brings together researchers whose work would be especially useful for my research. I would be grateful for the opportunity to present, receive feedback, and meet potential collaborators.

Simple Contracts for Busy Firms: Evidence from a Supply Chain Experiment*

Hao Chen† Guojun He‡ Qinrui Xiahou§

Abstract

Contracts are often read, understood, and acted on by firms managing many competing demands. In a 2×2 field experiment conducted in the supplier network of a listed company, we independently vary decision authority (mandated versus delegated KPI selection) and monitoring scope (targeted versus comprehensive compliance audits) to understand how contract design shapes supplier engagement and documentation. We show that simple governance wins: mandates and targeted audits improve engagement and documentation, while delegation and comprehensive audits do not. These results run counter to the predictions of classical theories of delegation and monitoring. They follow from a model in which a complex contract must be understood before it can be acted on, and the capacity spent on comprehension is drawn from the same budget as execution. Heterogeneity analyses, survey evidence, and behavioral measures corroborate this capacity mechanism.

JEL Codes: C93, D23, D82, L14, M14, Q56
Keywords: supply chain governance, field experiment, delegation, multitasking, monitoring, organizational capacity, ESG reporting, private regulation

*We thank participants at the Guanghua-Organization Science Paper Development Workshop and CCER Summer Institute for their helpful feedback, as well as the partner firm's management team for their cooperation. Financial support from the National Natural Science Foundation of China (72425014) and the HKU Jockey Club Enterprise Sustainability Global Research Institute is gratefully acknowledged. The experiment is pre-registered at the AEA RCT Registry (ID: AEARCTR-0016528) and received an ethical exemption from the Institutional Review Board of the University of Hong Kong. The authors contributed equally and are ordered alphabetically.

†University of Hong Kong. Email: hattiechenhao@connect.hku.hk.
‡University of Hong Kong. Email: gjhe@hku.hk. Corresponding author.
§University of Hong Kong. Email: qrxiahou@connect.hku.hk.

1 Introduction

Modern production is organized through supply chains that cross many firm boundaries. A single consumer product often passes through several legally independent firms, spanning multiple countries and tiers of suppliers, before reaching the final assembler (Antràs and Helpman, 2004; Antràs and Chor, 2013; Gereffi et al., 2005). Lead firms rely on suppliers for outcomes that are hard to contract on, including product quality, delivery timing, labor conditions, and environmental performance (Baker et al., 2002; Gibbons, 2005). This reliance defines a basic problem of organizational design: how much decision authority the lead firm allocates to suppliers, and how broadly it monitors what suppliers actually do. These two choices, over authority and over monitoring, are among the oldest questions in the economics of organization, and every buyer settles them in practice each time it writes the terms of a supplier program.

We study these choices through a randomized field experiment inside the supply chain of a Chinese publicly listed company, using the firm's own reporting platform as the field. Our partner is a leading apparel brand that holds top-tier environmental, social, and governance (ESG) ratings from major international agencies. It recently introduced ESG reporting requirements for 155 of its suppliers, giving us a natural multidimensional task in which to vary the governance structure. We cross two dimensions in a 2×2 design. On the vertical dimension, suppliers either receive a mandate specifying eight fixed indicators to report, or hold the authority to select any eight from a menu of 78. Both groups enter the data on the platform and may upload supporting documentation. On the horizontal dimension, suppliers face either comprehensive verification of all submitted data or targeted verification of carbon metrics alone. A separate control group faces the baseline requirement, with ESG filing otherwise voluntary. We measure supplier response along two margins. Completeness

is the share of applicable fields a supplier fills in, and documentation is the breadth and verifiability of the supporting material it uploads.

The design maps onto the two canonical trade-offs of organizational economics. On the vertical dimension, the theory of authority applies. A supplier knows better than the buyer which data it can readily gather, so delegating the choice of indicators lets it use that knowledge. Classical theory therefore expects delegation to draw fuller and more informative reports than a fixed mandate (Aghion and Tirole, 1997; Dessein, 2002; Mookherjee, 2006; Alonso and Matouschek, 2008). On the horizontal dimension, the theory of multitasking applies. When only some reported items are verified, a supplier has reason to neglect the rest, so broad verification should sustain wider effort than a narrow audit (Holmström and Milgrom, 1991). On both counts, classical theory favors the more flexible and the more comprehensive arrangement.

Contrary to the predictions of classical theory, the simpler arrangements draw the stronger response from the typical supplier in this network. Mandating a fixed set of indicators raises the share of applicable fields a supplier fills in by 10.4 percentage points, 19.4 percent of the control mean, and broadens the documentation it provides, while delegating the choice has small and imprecise effects. Targeted carbon monitoring produces gains of similar size, and comprehensive monitoring does not. The strongest response comes from the structure that mandates the indicators and targets verification. Our estimates also suggest that the two burdens compound, since relieving both together delivers less than the sum of relieving each alone on every outcome we report. The gains concentrate in environmental reporting, where measurement is clearest and coordination costs are lowest. Blinded assessments by the lead firm's own supply chain team rate the simpler arms higher, and the engagement gains do not come at the expense of revenue, employment, operating performance, or continued participation. Autonomy and breadth, the two features that classical theory favors, draw the weakest

engagement among the small suppliers that make up most of this network.

We attribute the departure to a resource classical theory holds fixed, the organizational capacity a supplier can devote to a new administrative task. Small suppliers run thin administrative structures, and choosing among indicators or preparing for a broad audit consumes staff time and coordination before any reporting begins (Bloom et al., 2013; Atkin et al., 2017; McKenzie and Woodruff, 2017). We develop a model in which a complex contract must be understood before it can be acted on, and the capacity spent on comprehension is drawn from the same budget as execution. Delegation and comprehensive monitoring both impose a fixed cost the supplier must pay before starting, and both raise the marginal productivity of the capacity that remains. These two forces cross once as capacity varies, so the simpler structures dominate below the crossing and the classical ranking returns above it. The model further implies that the two comprehension costs compound, and that engagement concentrates on the dimensions a supplier can most readily convert into verifiable content.

Four further tests bear on the channel. The returns to delegation and to targeted monitoring rise with supplier capacity, measured by pre-treatment registered capital, and the engagement gains appear where a staff member completes the filing and fade when the task falls to a busy executive. An endline survey traces unreported items to unavailable data and insufficient staff, and firms without dedicated ESG personnel or centralized data systems leave more items unreported. Behavior during structured reminder calls moves the same way. We also weigh the readings most likely to compete with capacity, among them the signaling of buyer priorities, the clarity of the request, and the selective disclosure of convenient fields. Each fits part of the evidence, but none reproduces the way the response scales with a supplier's own resources or predicts that the advantage of the simpler structure should narrow and change sign among the best-resourced suppliers.

Our paper makes three contributions. The first is to ask whether the canonical prescriptions of organizational design survive outside the firm. A growing share of world production is coordinated there, written into contracts between legally independent firms that once operated under authority inside a single hierarchy (Antràs and Helpman, 2004; Lafontaine and Slade, 2007; Antràs and Chor, 2013). Lead firms increasingly do what managers do inside a firm: they specify indicators, verify performance, and rate their suppliers. Whether theory built for the interior of the firm governs these arrangements is the question Gibbons (2005) posed, yet commercial contract terms are rarely available for manipulation. Direct tests of delegation and monitoring scope come almost entirely from the laboratory, from inside firms, and from public bureaucracies (Fehr et al., 2013; Hong et al., 2018; Rasul and Rogger, 2018; Manthei and Sliwka, 2019; Bandiera et al., 2021; Kala, 2024). In each of these settings, the principal has hired the agent, trained it, and paid for its administrative infrastructure. A buyer inherits its suppliers and can do none of these things. Field experiments in supplier networks work within that constraint, typically varying what a buyer asks of a supplier and how hard it enforces the request (Ghosh and Singh, 2026; Amengual and Distelhorst, 2025; Boudreau, 2024). We instead vary who decides what gets reported and how widely performance is verified, the two margins on which organizational theory takes a stand, and neither of its classical rankings survives.

The second is to extend the classical framework so that its prescriptions emerge as a special case. Agency theory rations the agent's effort and prices it through incentives (Holmström and Milgrom, 1991), but it assumes the agent can already work out what the contract asks of it, and no clause of the contract prices that ability. We add one parameter for it, the capacity a supplier must spend before any reporting begins. The classical rankings return in the limit where this cost is small relative to capacity. Because both levers draw on the same budget, the parameter moves both

dimensions of the design, their burdens compound, and authority and monitoring become one design problem. The extension also organizes findings that have looked unconnected. Inside firms, authority raises productivity where the agent holds the slack to act on what it is given (Acemoglu et al., 2007; Bloom et al., 2012; Aghion et al., 2021; Kala, 2024), the region our model assigns to abundant capacity. Work on the cost of deliberation (Garicano, 2000; Tirole, 2009; Bolton and Faure-Grimaud, 2010) and on rationed organizational attention (Dessein et al., 2016; Bandiera et al., 2020) supplies the microfoundation, and our experiment supplies the field evidence that this cost determines whether a contract works.

The third is to the practice of private regulation, where firms now set and enforce standards that public authority does not reach (Kitzmüeller and Shimshack, 2012; Locke, 2013; Boudreau et al., 2023). Research on supplier compliance treats the supplier's response as a problem of incentives or of information, to be solved with audit intensity, certification, contingent payment, and disclosure (Guo et al., 2016; Plambeck and Taylor, 2016; Kalkanci and Plambeck, 2020), and field evidence on what buyers can enforce points the same way (Short et al., 2016; Distelhorst et al., 2017; Boudreau, 2024). We identify a margin that operates before incentives do. A requirement that outruns what a supplier can absorb fails among firms that are willing and well informed, so breadth demanded at the outset selects the suppliers with staff to spare and passes over those the buyer most needs to observe. The reach of a disclosure program is therefore an upgrading question as much as an enforcement one, consistent with evidence that firm capability limits what small producers can take up (Bloom et al., 2013; Atkin et al., 2017; McKenzie and Woodruff, 2017; Bruhn et al., 2018; Verhoogen, 2023) and that administrative burden governs take-up of public programs (Bhargava and Manoli, 2015; Deshpande and Li, 2019; Finkelstein and Notowidigdo, 2019). A buyer that wants coverage has to build the supplier's

ability to comply as part of asking for it.

The paper proceeds as follows. Section 2 describes the institutional setting, and Section 3 describes the experimental design. Section 4 reports the main results and their departure from the classical predictions. Section 5 develops the capacity model. Section 6 tests the capacity channel and weighs alternative mechanisms. Section 7 examines performance and business outcomes. Section 8 discusses implications. Section 9 concludes.

2 Institutional Context

2.1 The Lead Firm

The lead firm in this study is an industry-leading consumer goods and apparel brand based in China, with a large outsourced supplier network serving domestic and international markets. It holds top-tier ratings from major international ESG evaluation agencies, reflecting recognized progress in its own operations. However, as with most global consumer brands, much of its environmental and social footprint is embedded in its supply chain, where hundreds of independent manufacturers produce goods and source materials.

Prior to this study, the firm had limited formal mechanisms for supplier ESG management. While it maintained basic records of traditional performance metrics such as on-time delivery and product quality, it possessed no systematic mechanism for monitoring or verifying upstream ESG practices. This gap became increasingly difficult to sustain as investor, regulatory, and consumer pressure for supply chain transparency intensified. The firm's strategic objective was accordingly to move beyond surface-level compliance toward genuine accountability. This required establishing ESG standards

and designing governance structures capable of motivating substantive participation among independent suppliers operating under significant resource constraints.

Fuller supplier disclosure is valuable to the firm in itself. Its third-party ESG ratings depend in part on the extent of supply-chain disclosure, so broader and more complete supplier reporting can strengthen the firm's standing with rating agencies. Evidence from other settings points in the same direction. Disclosure programs have improved the outcomes they cover, from restaurant hygiene to industrial emissions (Jin and Leslie, 2003; Doshi et al., 2013; Chen et al., 2018; Tomar, 2023; Christensen et al., 2021), and reporting that requires real effort can serve as a credible signal of commitment (Lys et al., 2015). The buyer made its priorities clear to all suppliers. The platform's welcome page announced awards for top performers, and the briefing described in Section 2.3 stated that weak ESG performance could affect future orders. The experimental arms therefore differ only in how the reporting request is structured, in who selects the indicators and in how widely the entries are verified.[1]

2.2 The Supplier Network

The firm's supplier network comprises more than 300 manufacturing partners across China in four main categories. Cut-make-trim manufacturers assemble garments from pre-cut fabric. Original design manufacturers handle product design, material sourcing, and manufacturing. Fabric and material suppliers produce raw textiles and functional synthetic materials. Ancillary material suppliers provide zippers, labels, and other accessories.

For this experiment, the firm introduced ESG management across its 155 active

[1] The reporting data are consistent with this interpretation. Suppliers allowed to select their own indicators tended to choose those the buyer would otherwise have mandated, and suppliers audited on carbon alone completed social and governance fields at least as fully as suppliers audited on all indicators (Appendix Figure F1). Thus, neither greater choice nor a narrower audit shifted reporting away from the buyer's stated priorities.

core suppliers. As Figure 1 illustrates, these suppliers are predominantly micro and small firms, typically operating on thin margins and handling labor-intensive garment assembly or design. Median employment is 31 workers and median firm age is 15 years. Most operate with limited formal management hierarchies, and few possess dedicated ESG departments or specialized compliance personnel. This profile is representative of manufacturing supply chains in emerging economies, where local suppliers play a central role in production but have much thinner administrative resources than large corporate entities (Gereffi et al., 2005).

2.3 The Digital Platform

The experiment is embedded within the lead firm's proprietary digital Supply Chain Management platform, which serves as the standard operational interface for routine business reporting between the firm and its suppliers. Integrating the experiment into an existing corporate system ensured that suppliers perceived the new ESG reporting requirements as standard policy updates. This naturalistic setting limits experimenter demand effects that can arise when subjects know they are being observed (Harrison and List, 2004).

Upon logging into the platform, suppliers are presented with a Welcome Page that outlines the reporting requirements for the current cycle, as shown in Figure 2. While the general format is identical for all users, the specific text varies dynamically with the supplier's assigned treatment condition. This page is the primary mechanism for introducing the experimental variations, informing each supplier whether it must report a fixed set of indicators or may choose its own, and whether its data will be subject to comprehensive or targeted carbon monitoring.

Once suppliers click "Start Filing" on the Welcome Page, they are directed to the

main reporting dashboard shown in Figure 3. The interface organizes the reporting requirements into distinct tabs for environmental, social, governance, financial, and certification metrics. For every active indicator, suppliers can enter data, add qualitative remarks, and upload supporting documentation. The interface also displays historical submissions alongside current-period fields.

The platform tracks a library of 108 supply chain indicators listed in Appendix A and summarized in Table F1. Thirty indicators covering basic financial information and standard factory certifications are mandatory for all suppliers and constitute the common baseline requirement. The remaining ESG indicators are mandatory or optional according to the supplier's assigned treatment group.

After baseline data collection, the lead firm's supply chain management team conducted a structured telephone briefing with a designated manager at each supplier. The briefing outlined reporting requirements and performance evaluation standards without disclosing the randomized nature of the governance structure. It also made the buyer's own stake explicit. Because the firm's third-party ESG ratings depend on the coverage and completeness of supply chain disclosure, it asked suppliers to report broadly and to substantiate what they reported, and it warned that weak ESG performance would count against a supplier in future order allocation. What the buyer wanted was therefore common knowledge in the network, and a supplier that delivered it was also protecting its position in the relationship. Each supplier understood its own obligations while remaining unaware of the broader experimental design, which preserved both the integrity of the experiment and the realism of the governance environment.

2.4 Third-Party Monitoring and Verification

Third-party verification is widely used in global supply chain governance and can affect firms' compliance and reporting behavior (Duflo et al., 2013; Short et al., 2016; Chen and Lee, 2017). During the experimental period, the lead firm implemented a formal verification process to strengthen data integrity and accountability. Independent professional auditors reviewed the quality of submitted data for suppliers assigned to the treatment conditions.

The scope of verification varied by experimental assignment. Under comprehensive monitoring, auditors verified ESG indicators across the environmental, social, and governance domains through document reviews, facility inspections, and employee interviews. Under targeted monitoring, auditors verified carbon-emissions data, with the remaining ESG indicators left self-reported. This variation identifies the effect of broader versus narrower oversight while holding fixed the presence of third-party verification. It allows us to assess whether broader oversight improves overall reporting or imposes coordination costs that reduce supplier engagement.

3 Experimental Design

3.1 The 2×2 Factorial Design

We implement a 2×2 factorial randomized controlled trial that independently varies two governance dimensions (Figure 4). The first dimension is the vertical allocation of authority. Under *Mandate* the lead firm specifies a fixed set of eight ESG KPIs for the supplier to report. Under *Delegation* the supplier selects eight indicators from a broader menu of 25 environmental, 48 social, and 5 governance indicators. The

second dimension is the horizontal scope of monitoring. *Comprehensive Monitoring* subjects the full ESG reporting scope to third-party verification, and *Targeted Carbon Monitoring* centers verification on carbon-related metrics.

Crossing these two dimensions yields four treatment cells and a separate control group. All five arms use the same platform, receive the same briefing, and compete for the same awards. Within this common environment, control suppliers face the baseline arrangement: ESG filing remains voluntary, with no assigned KPI-selection rule or verification structure. The control group therefore provides the voluntary-reporting benchmark. Comparing each treatment cell with control identifies the effect of introducing a governance structure, while comparisons across treatment cells identify how authority allocation and monitoring scope affect reporting once one is in place.

Operationally, the five arms are implemented through distinct platform interface designs (Figure 5). The control group is presented with a standard baseline interface. The four treatment groups receive interfaces matched to their treatment cells. The control group is presented with a standard baseline interface. The four treatment groups receive interfaces matched to their treatment cells. The treatment variation is introduced on the Welcome Page and carries through to the filing interface, where the platform adjusts required fields according to the assigned KPI-selection rule. Screenshots and English translations of the group-specific welcome screens are provided in Appendix B.

3.2 Randomization and Sample Balance

We randomly assigned the 155 participating suppliers across the five experimental conditions using a stratified procedure based on supplier type and firm size, proxied by historical production value. Stratification improves precision and ensures that

treatment effects are not confounded by systematic pre-existing differences in supplier characteristics. The control group received 46 suppliers and the four treatment cells approximately 27 each, for a total of 109 treated suppliers.

Table 1 reports pre-treatment balance for each of these cells against the control group. They are indistinguishable from control on production volume, registry-filed employment, firm age, registered capital, certification counts, supplier type, supplier ranking, and size class. One of the sixty mean comparisons is significant at the ten percent level, and it survives neither a median nor an inverse hyperbolic sine transform of the same variable. Regressing each cell indicator on the full set of characteristics against control returns joint p-values between 0.43 and 0.94, and a multinomial logit of arm assignment on the same characteristics returns $p = 0.98$. Assignment is therefore orthogonal to observables cell by cell, supporting the validity of the randomization procedure.

3.3 Timeline and Implementation

The experiment began with retrospective reporting for fiscal year 2024, which ran from April 2024 through March 2025. Treatment-specific interfaces were in place from each supplier's first observed filing. The output, revenue, employment, and sales reported in this round pertain to fiscal year 2024, before the program began. They provide the pre-program benchmark for the business outcomes analyzed in Section 7.2.

The experiment then continued with four quarterly reporting rounds over fiscal year 2025, from April 2025 through March 2026. From the first filing onward, each supplier's Welcome Page and subsequent communications specified whether verification would be comprehensive or targeted to carbon metrics. In September 2025, the lead firm's supply chain management team conducted structured telephone briefings with designated

supplier managers. The calls reiterated each supplier's reporting requirements and verification scope without revealing the broader experimental design; Section 6.7 analyzes the recordings. Independent third-party verification took place in November 2025.

An endline survey administered in June 2026 collected information on reporting barriers, internal staffing, data systems, and perceived task burden. Registration at the AEA RCT Registry was filed before the first fiscal-year 2025 reporting round opened and before the research team obtained access to platform data.

3.4 Outcome Measures and Estimation Sample

We measure supplier engagement on two margins. The first is *completeness*, the number and the share of applicable platform fields a supplier fills in.[2] This margin captures mainly voluntary reporting. Treated suppliers complete all eight assigned indicators, whether the lead firm or the supplier chose them, so completeness differences among treated suppliers come from fields beyond these eight. Control suppliers complete about 77 of the applicable fields on average with no assignment at all, and treated suppliers complete about 89. Completeness is therefore high across all arms and driven largely by reporting effort rather than compliance with any specific requirement.

The second is *documentation*, the breadth and the depth of the evidence a supplier uploads. Of the 108 metrics on the platform, seventy-five accept a supporting attachment, and we count how many of these a supplier backs with at least one file. We also count the ESG categories covered (out of twelve), and the categories backed by hard evidence. Three counts are needed because the quantity of documentation a

[2] The applicable set contains nineteen fields the platform fills by computing them from other entries. Recomputing both completeness measures on the remaining eighty-nine leaves the mandate effect at 9.6 percentage points and targeted monitoring at 10.4, close to the 10.4 and 10.2 in the main specification.

supplier submits need not reflect its substance. A supplier can add files to a topic it has already documented, which raises the file count without widening coverage, and it can cover a topic with descriptive material that carries no verifiable record. The first two counts therefore measure how much and how broadly a supplier documented, and the third measures how much of that a verifier could check.

We disaggregate both margins by environmental, social, and governance domain to assess whether responses vary with coordination complexity. Endline supplier scores, business records, respondent-role information, survey responses, and reminder-call recordings provide supplementary measures of performance, capacity, and process engagement. Appendix C details the attachment-based measures, and Appendix E describes the survey and reminder-call variables.

Figure F2 describes the two engagement margins across filing rounds. The platform estimation sample contains 735 supplier-round observations from 152 suppliers with valid treatment assignments.[3] Supplier participation also varies across reporting rounds with seasonal production cycles and factory operations.[4] Section 7.2 shows that treatment assignment does not affect supplier exit, indicating that this variation does not generate differential attrition.

3.5 Empirical Specification

Given that the experimental interfaces were deployed from the first reporting round, all observed platform behavior occurs under the assigned governance conditions. Our

[3] Three suppliers never entered any platform data and therefore provide no outcome variation. Each is in a different treatment cell, so their exclusion is unlikely to introduce selection bias.

[4] Five rounds for 152 suppliers give 760 potential observations, of which 735 are observed. Of these suppliers, 132 appear in all five rounds, fifteen in four, and five in three. The unobserved rounds concentrate in the retrospective fiscal-year 2024 filing and the final quarterly round, and they are spread across arms, with five in the control group and between three and nine in each treatment cell.

main estimating equations are OLS regressions with stratum and round fixed effects:

$$Y_{ist} = \alpha + \beta_1 \text{Mandate}_i + \beta_2 \text{Delegation}_i + \delta_s + \lambda_t + \varepsilon_{ist}, \tag{1a}$$

$$Y_{ist} = \alpha + \beta_1 \text{Comprehensive}_i + \beta_2 \text{Targeted}_i + \delta_s + \lambda_t + \varepsilon_{ist}. \tag{1b}$$

Here, Y_{ist} is the engagement outcome for supplier i in stratum s during reporting round t. The control group, which faces the baseline requirement alone, is the omitted category, so each β measures the effect of introducing a governance structure relative to voluntary ESG filing. Stratum fixed effects δ_s absorb cross-sectional variation used in the randomization, and round fixed effects λ_t control for common time patterns. Treatment was assigned to suppliers within strata, so we cluster standard errors on the supplier. Section 4.4 reports the stratum-clustered estimates and a wild-cluster bootstrap over the strata. To estimate joint effects, we replace the pooled indicators with four treatment-cell indicators.

The capacity-threshold analysis of Section 6.3 uses a fully interacted version of (1). Let Above_i indicate that supplier i lies above a capacity threshold and $\text{Below}_i = 1 - \text{Above}_i$. For the vertical dimension we estimate

$$\begin{aligned} Y_{ist} = \alpha &+ \beta_{1L} \left(\text{Mandate}_i \times \text{Below}_i\right) + \beta_{1H} \left(\text{Mandate}_i \times \text{Above}_i\right) \\ &+ \beta_{2L} \left(\text{Delegation}_i \times \text{Below}_i\right) + \beta_{2H} \left(\text{Delegation}_i \times \text{Above}_i\right) \\ &+ \gamma \, \text{Above}_i + \delta_s + \lambda_t + \varepsilon_{ist}, \end{aligned} \tag{2}$$

and the horizontal dimension replaces the two arms with Comprehensive and Targeted. Because both capacity groups carry their own arm indicators and γ absorbs the level difference between them, the control group is the omitted category *within* each capacity group. Each of the four coefficients is therefore the effect of one arm measured against

control suppliers of the same capacity, and $\beta_{1L} - \beta_{2L}$ and $\beta_{1H} - \beta_{2H}$ are the simple-minus-complex gaps below and above the threshold. Dilution-related heterogeneity specifications interact the treatment indicators with the role of the person completing the report. Endline ESG scores and supplier exit are estimated in cross-sectional regressions with stratum fixed effects. Business outcomes are estimated by OLS controlling for the fiscal-year 2024 level of the outcome. Reminder-call regressions include caller and stratum fixed effects.

4 Main Results

This section reports the main effects of the two governance dimensions on the completeness and documentation measures defined in Section 3. Three results emerge. A fixed mandate raises engagement and a delegated choice does not. Targeted carbon monitoring raises it and comprehensive monitoring does not. Their combination produces the largest response of the four governance bundles. We show that these patterns are robust across specifications and then compare them with the predictions of classical organizational theory.

4.1 Mandate versus Delegation

We begin with the vertical allocation of authority. Each arm is measured against the control group.

Figure 6 presents the treatment effects, and Table F2 reports the corresponding regressions. On the completeness margin, mandate raises the number of completed indicators by 15.939 and the completion rate by 10.4 percentage points. The latter represents a 19.4 percent increase relative to the regression-adjusted control mean of 53.7 percent. Delegation raises the completion rate by 4.5 percentage points, and the

estimate is not statistically distinguishable from zero. The raw number of completed indicators shows the same ordering. The two arms differ at the ten percent level on both completeness outcomes and at the five percent level on attachment categories and hard evidence.

The documentation outcomes reinforce this pattern. Mandated suppliers document 1.075 additional indicators, cover 0.958 additional attachment categories, and cover 0.766 additional categories with hard, verifiable evidence. The category effect is approximately one-half of the control mean of 1.892 categories. The category and hard-evidence estimates are statistically significant at the 5 percent level. The indicator estimate carries a standard error of 1.044 and is not statistically distinguishable from zero. The delegation coefficients for indicators, categories, and hard evidence are small and imprecise.

Across both margins, only the mandate raises engagement. Requiring the indicator set yields more complete reporting and more verifiable documentation, while delegating the choice to suppliers leaves engagement near the control level.

4.2 Targeted versus Comprehensive Monitoring

We turn to the horizontal scope of monitoring, again measuring each treatment arm against the control group.

Figure 7 and Table F3 report the estimates. Targeted carbon monitoring raises the number of completed indicators by 15.554 and the completion rate by 10.2 percentage points, a 19.0 percent increase relative to the control mean of 53.7 percent. Both completeness estimates are significant at the 1 percent level. It also raises attachment coverage by 0.806 categories, roughly two-fifths of the control mean, and hard-evidence coverage by 0.615 categories, each significant at the 5 percent level. Documented

indicators rise by 1.075 with a standard error of 1.057, directionally consistent but statistically insignificant. Comprehensive monitoring produces a 4.7 percentage point increase in completion and a 0.227 category increase in attachment coverage, and both estimates are imprecise, as are its other documentation coefficients. The two monitoring arms differ at the ten percent level on the completeness outcomes and are not statistically distinguishable on documentation.

The monitoring dimension mirrors the authority dimension. Targeted monitoring moves engagement on both margins, while broadening the audit across all three ESG domains leaves it close to the control level.

4.3 The Combined Treatment

The two dimensions cross in the factorial design, so we can also read the four treatment cells directly.

Figure 8 plots the cell-level estimates, and Table F4 reports their magnitudes. The mandate and targeted-monitoring cell raises the completion rate by 11.3 percentage points, documents 1.521 additional indicators, expands documentation coverage by 1.240 categories, and adds 0.961 categories backed by hard evidence. The two completeness estimates are significant at the 5 percent level and the documentation-coverage and hard-evidence estimates at the 1 percent level, while the indicator estimate is imprecise.[5]

The remaining three cells fall short of this combination. Mandate with comprehensive monitoring raises the completion rate by 9.5 percentage points and attachment coverage by 0.695 categories, though the estimates are less precise. Delegation with

[5] This cell contains the network's largest suppliers by baseline production, so we checked that its estimates do not rest on them. Dropping the largest one, two, or five suppliers, winsorizing the outcome at the first and ninety-ninth percentiles, and adjusting for baseline covariates interacted with treatment following Lin (2013) all leave the estimates essentially unchanged.

targeted monitoring raises the completion rate by 9.1 percentage points and attachment coverage by 0.405 categories. Delegation with comprehensive monitoring produces little movement on either margin, the weakest result among the four cells.

The four cells line up in the same order on completion, indicators documented, documentation coverage, and hard evidence, with the fully simplified cell first, the two single-lever cells close behind, and the fully complex cell last. The fully simplified and the fully complex cell differ at the one percent level on four of the five outcomes (Table F4). The design's power lies with the main effects, so we treat the cell ordering as descriptive.

4.4 Robustness

Treatment was assigned to suppliers within strata, so the main regressions cluster standard errors on the supplier. Two features of the design could still shape inference. The panel weights suppliers by how often they file, and the thirteen strata leave a stratum-level convention resting on twelve degrees of freedom. Tables F5 and F6 therefore re-estimate the main specifications with standard errors clustered on the stratum, with that same fit bootstrapped over the strata, and with the outcome collapsed to one observation per supplier. Separately, we re-estimate the twenty core arm effects, both arms on all five outcomes in each dimension, under the same bootstrap.

Coefficients barely move across the alternatives. The mandate estimate for completion stays near 10.4 percentage points and the targeted-monitoring estimate near 10.2, while delegation and comprehensive monitoring remain smaller under every specification. Clustering on the thirteen strata widens the intervals in both tables: the two completeness effects fall from the one to the five percent level, and nothing

loses significance. Indicators documented is more sensitive to the clustering choice, reaching significance at the stratum level but not at the supplier level for targeted monitoring. Wild-bootstrap p-values track their analytic counterparts closely, with targeted monitoring on completeness strengthening from five to one percent, and bootstrapping over suppliers changes no estimate by more than 0.025. The main results therefore survive the clustering convention, the weighting implicit in repeated supplier-round observations, and few-cluster inference alike. The one statistic sensitive to few-cluster inference is the crossing test of Section 6.3, a difference-in-differences estimated off the small above-threshold cell, and we report bootstrap p-values there.

4.5 Relation to the Classical Benchmarks

The three findings run counter to the two organizational theories that speak most directly to the design. The theory of delegation holds that granting authority to a better-informed agent improves outcomes, since the agent can tailor the task to private information about local conditions (Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008). The multitasking theory of performance measurement holds that broad monitoring sustains effort across all valued dimensions, since narrow measurement invites the agent to divert effort toward whatever is measured (Holmström and Milgrom, 1991). On both counts the more flexible and the more comprehensive structures would be expected to draw the fuller response. In this setting the opposite holds. The simpler and narrower structures move engagement, and the more flexible and more comprehensive ones do not.

The two theories share an assumption that this setting does not meet. Both take for granted that the agent has the capacity to act on the authority it is granted and to prepare for the monitoring it faces. That assumption is natural inside a firm, where the

theories were developed and where they find consistent empirical support (Acemoglu et al., 2007; Bloom et al., 2012; Bandiera et al., 2021; Kala, 2024). It is less natural among small independent suppliers, whose administrative resources are thin and for whom the work of choosing among tasks or preparing for a broad audit can absorb the very capacity the program is meant to mobilize (Bloom et al., 2013; Atkin et al., 2017; McKenzie and Woodruff, 2017). Evidence from other settings runs the same way. Administrative simplification raises take-up when decision costs are salient (Bhargava and Manoli, 2015; Bettinger et al., 2012), and application costs screen out the agents with the least administrative slack (Finkelstein and Notowidigdo, 2019; Deshpande and Li, 2019). The next section makes the assumption explicit, recovers the classical predictions when comprehension is cheap or capacity abundant, and shows how the simpler structures come to dominate as capacity binds.

5 A Model of Organizational Capacity

The main results run opposite to the predictions of delegation theory and multitask monitoring theory, and Section 4 traced the difference to a resource those theories hold fixed: the capacity a supplier can devote to the task. This section builds that resource into an otherwise standard principal-agent framework that embeds the classical theories of authority allocation (Aghion and Tirole, 1997; Dessein, 2002) and of monitoring scope (Holmström and Milgrom, 1991). A single friction, a binding constraint on the supplier's organizational capacity, governs both governance margins at once. The model recovers the classical rankings as limiting cases and identifies the capacity region in which those rankings reverse.

This single friction also speaks to a related question: how the two instruments interact. The joint design of decision rights and performance measures is itself a

classical theme (Holmström and Milgrom, 1991, 1994; Athey and Roberts, 2001; Dessein et al., 2010), and that literature finds the two instruments complementary, so they cluster in practice and piecemeal adoption returns little (Milgrom and Roberts, 1990; Ichniowski et al., 1997; Brynjolfsson and Milgrom, 2013). That literature places the binding scarcity with the principal, who can price it through the contract. Here, the scarcity sits with the agent, and no clause of the contract prices it. A complex contract must be understood before it can be acted on, and both governance levers draw on the same capacity to do this understanding before either delivers anything, so a more complex structure starts later and climbs faster. Because the levers share one budget, the relief each offers substitutes for the other, and relieving one margin recovers most of what relieving both would deliver. The same capacity constraint also determines which structure is more effective, and this ranking is not fixed but depends on where the supplier's capacity sits.

Section 5.1 lays out the environment and the payoffs of the two parties. Section 5.2 derives the classical benchmark and the conditions under which complexity raises the value of the report. Section 5.3 introduces comprehension costs and derives the reversal. Section 5.4 develops four extensions, and Section 5.5 collects the testable hypotheses. Appendix D contains all proofs and the formal microfoundations.

5.1 Setup

Agents and capacity. A lead firm (the buyer) transacts with a continuum of suppliers. Each supplier is endowed with *organizational capacity* k, the managerial and staff time, records infrastructure, and coordination ability available for the reporting program during the period. Capacity is heterogeneous across the network, distributed according to F with support $[0, \bar{k}]$ and continuous, strictly positive density

on $(0, \bar{k})$. Section 5.4.4 decomposes k into the organization's endowment and the competing load of the person who handles the filing.

Capacity is a broader endowment than managerial attention. Attention is the flow that binds day to day (Ocasio, 1997; Dessein et al., 2016), while capacity also includes the standing apparatus of dedicated personnel, centralized records, and structured processes that determines how much administrative work the organization can absorb (Bloom and Van Reenen, 2007; Bloom et al., 2013). It is distinct, too, from the individual cognitive bandwidth that scarcity consumes (Mani et al., 2013; Gabaix, 2014; Kaur et al., 2025). Capacity resides in the organization, and a firm can raise it by hiring staff or centralizing records. A fixed capacity constraining execution is a standard premise in operations, where a rising load on a fixed service capacity degrades both the speed and the quality of the work (KC and Terwiesch, 2009).

Governance structures. The buyer designs a governance structure $c = (V, H)$. The vertical dimension $V \in \{M, D\}$ allocates authority over indicator selection. Under mandate (M) the buyer specifies the reporting indicators, and under delegation (D) the supplier selects them. The horizontal dimension $H \in \{C, T\}$ sets the scope of third-party verification. Comprehensive (C) verification covers the full program, and targeted (T) verification covers a single focal task.

Comprehension and execution. A contract can be acted on only after it has been understood. Processing the governance structure consumes capacity

$$r(c) \;=\; r_0 \;+\; \delta\,\mathbf{1}\{V = D\} \;+\; \mu\,\mathbf{1}\{H = C\}, \qquad \delta,\ \mu \geq 0, \tag{3}$$

where δ is the cost of evaluating the indicator menu and selecting from it, and μ is the cost of mapping a broad verification scope onto internal records and functions. Capacity

remaining after this comprehension cost is what remains available for execution, and a supplier that devotes effort e to the program produces engagement

$$\phi(c)\, g(e), \qquad \phi(c) \;=\; 1 \;+\; \pi_D\, \mathbf{1}\{V = D\} \;+\; \pi_C\, \mathbf{1}\{H = C\}, \qquad \pi_D,\ \pi_C \geq 0. \quad (4)$$

The factor $\phi(c)$ is execution productivity. A supplier that has worked through a richer governance structure knows which of its records support which indicator and which material the verifier will examine, so each unit of effort converts into more reported and better documented content. The two productivity gains enter additively, so complexity generates no interaction in execution. Section 5.4.1 allows for an interaction in comprehension.

Engagement is a single scalar and covers both margins measured in the experiment. Filling a platform field and attaching a record that documents it draw on the same capacity, and the model treats them as one output. The empirical counterparts are the share of platform indicators a supplier fills in and the breadth and verifiability of the documentation it attaches.

Assumption 1 (Execution technology)**.** $g(0) = 0$, $g' > 0$, $g'' < 0$, and $g(e) \to \infty$ as $e \to \infty$.

Only the restriction of g to $[0, \bar{k}]$ enters the results, since residual capacity never exceeds $\bar{k}$.

Payoffs. Each unit of engagement is worth $w(c)$ to the buyer. Delegation affects this value because the supplier holds private information about which indicators its records can support (Aghion and Tirole, 1997; Dessein, 2002), and verification scope affects it because material outside a targeted audit retains only part of its value (Holmström and Milgrom, 1991). Section 5.2 states the conditions under which $w(c)$ rises with

complexity on each dimension.

The supplier values the buyer's assessment because the assessment feeds future business (Baker et al., 2002; Macchiavello and Morjaria, 2015). Let $\alpha \in [0,1]$ denote the supplier's stake in the relationship, so that the supplier's private return per unit of engagement is $\alpha\, w(c)$. Capacity assigned neither to processing nor to execution earns the core-business return $\rho > 0$ per unit. A supplier with capacity $k \geq r(c)$ facing structure c therefore solves

$$\max_{0 \leq e \leq k - r(c)} \; \rho \left[k - r(c) - e \right] \; + \; \alpha\, w(c)\, \phi(c)\, g(e), \tag{5}$$

and supplies $e = 0$ when $k < r(c)$. Writing $E(k; c)$ for equilibrium engagement, the buyer chooses the structure to maximize program value,

$$\Pi(c) \;=\; \int_0^{\bar{k}} w(c)\, E(k; c)\, dF(k). \tag{6}$$

Timing. The buyer announces c. The supplier processes the contract at cost $r(c)$, and under delegation it then selects the indicator set, a subgame set out in Appendix D. It allocates residual capacity between execution and the core business, after which engagement is realized, verified according to H, and assessed.

The supplier's solution. The objective in (5) is strictly concave, with derivative $\alpha\, w(c)\, \phi(c)\, g'(e) - \rho$. We restrict attention to suppliers for whom the program competes for scarce capacity at every capacity level in the support.

Assumption 2 (Binding capacity)**.** For every structure c and every k in the support of F with $k \geq r(c)$, $\alpha\, w(c)\, \phi(c)\, g'\big(k - r(c)\big) > \rho$.

Under Assumption 2 the supplier devotes all residual capacity to the program, and

equilibrium engagement is

$$E(k;c) \;=\; \begin{cases} 0, & k < r(c), \\ \phi(c)\, g\big(k - r(c)\big), & k \geq r(c). \end{cases} \tag{7}$$

Lemma 1 (The engagement curve)**.** *$E(\cdot\,;c)$ is continuous, equal to zero up to $r(c)$, and strictly increasing and strictly concave thereafter. An increase in $r(c)$ shifts the curve rightward by that amount, and an increase in $\phi(c)$ raises and steepens it, leaving the entry point unchanged.*

Figure 9 draws the curve and its two comparative statics. The entry point $r(c)$ is what the supplier pays before any content is produced, and the slope $\phi(c)\, g'(k - r(c))$ is the rate at which residual capacity converts into engagement. Governance moves the entry point and the rate of climb in opposite directions, and every result below follows from that opposition. Appendix D also treats the case in which the supplier may decline to process the contract. Entry points then shift further to the right, by more for structures with larger $r(c)$, and the results below are unchanged.

5.2 The Classical Benchmark

The classical theories evaluate governance structures through the value $w(c)$ and treat the cost of absorbing a structure as negligible relative to the organization's capacity. The benchmark below sets these costs to zero. What complexity is worth to the buyer differs across the two dimensions of governance, so we take them in turn.

On the vertical dimension the gain is not automatic. Under delegation the supplier chooses between an informative indicator set, which matches indicators to the buyer's information needs, and a convenient set, which minimizes its own selection effort. Appendix D analyzes this subgame.

Lemma 2 (Congruent selection)**.** *There exists a threshold* $\underline{\alpha} \in (0,1)$ *such that the supplier selects the informative indicator set, and hence* $w(D,H) > w(M,H)$*, if and only if* $\alpha \geq \underline{\alpha}$*.*

Delegation therefore raises the value of the report only when the supplier's interests are sufficiently aligned with the buyer's. The stake α carries that alignment, and the setting described in Section 2.3 places the network in the aligned case. We therefore assume alignment above that threshold, and that broad verification raises the value of the report, so that complexity is valuable on each dimension.

Assumption 3 (Value of complexity)**.** $\alpha \geq \underline{\alpha}$, and $w(V,C) > w(V,T)$ for each V. The value $w(c)$ is then strictly increasing in complexity on each dimension.

Proposition 1 (Classical benchmark)**.** *Suppose comprehension is costless across structures,* $\delta = \mu = 0$*. Then for every capacity level* $k > r_0$ *and every structure* c*,*

$$E(k;\, D,C) \;\geq\; E(k;\, c) \qquad \textit{and} \qquad w(D,C)\, E(k;\, D,C) \;\geq\; w(c)\, E(k;\, c),$$

with strict inequalities whenever $\pi_D + \pi_C > 0$*. The buyer chooses delegation and comprehensive verification,* $c^* = (D,C)$*, for every capacity distribution* F*.*

Proposition 1 reproduces the two classical prescriptions. The engagement ranking follows from execution productivity. Whatever the supplier's stake, a structure it has worked through converts capacity into content at a higher rate. The value ranking also uses the informational channel, which rests on congruent selection on the vertical dimension and on audit scope on the horizontal one. Costless comprehension is stronger than the result requires. With $\delta, \mu > 0$ the same ranking holds once supplier capacity is large relative to those costs, as Proposition 6(ii) below shows.

Corollary 1 (Congruence failure)**.** *If $\alpha < \underline{\alpha}$, then $w(D, H) = w(M, H)$ and delegation delivers no informativeness gain. Sufficient congruence is a precondition for the vertical component of the trade-off analyzed below.*

5.3 Capacity Costs and the Reversal

The experiment concerns suppliers for whom complexity is costly to absorb.

Assumption 4 (Comprehension costs)**.** $\delta > 0$, $\mu > 0$, and $\pi_D, \pi_C > 0$.

Evaluating an indicator menu against internal data availability, and planning documentation for a broad verification scope, each draw on the same capacity that produces engagement. Equation (3) orders the structures by total overhead, $r(M, T) < r(M, C)$, $r(D, T) < r(D, C)$, while (4) orders them in the opposite direction by execution productivity. A more complex structure therefore starts later and climbs faster, and two such curves can be ranked only by capacity.

Proposition 2 (Single crossing)**.** *Let s and x be two structures with $r(s) < r(x)$ and $\phi(s) \leq \phi(x)$, and define $\Delta_{xs}(k) = E(k; x) - E(k; s)$.*

(i) Δ_{xs} is strictly increasing on $[\, r(x), \bar{k}\,]$, with $\Delta_{xs}(r(x)) < 0$. The curves therefore meet at most once, and where they meet Δ_{xs} changes sign from negative to positive.

(ii) A crossing exists in the support if and only if $\phi(x) > \phi(s)$ and $\Delta_{xs}(\bar{k}) > 0$. It is then unique, denoted k^* and defined by $\phi(x)\, g(k^* - r(x)) = \phi(s)\, g(k^* - r(s))$, and both structures are still climbing there. The threshold rises with the complex structure's comprehension cost and falls with its execution productivity: $\partial k^*/\partial r(x) > 0$ and $\partial k^*/\partial \phi(x) < 0$.

(iii) The buyer's pointwise ranking reverses weakly earlier: there exists $k^*_\Pi \leq k^*$ *such that* $w(x)E(k;x) > w(s)E(k;s)$ *for all* $k > k^*$ *and* $w(s)E(k;s) > w(x)E(k;x)$ *for all* $k \in (r(s), k^*_\Pi)$.

Concavity alone delivers part (i), since the complex structure operates on a steeper stretch of the climb and the gap between the curves widens. Without a productivity gain the complex structure would start later and climb no faster, so it would be dominated at every capacity level, which is the case in Corollary 1.

Figure 10 displays the crossing. Left of k^* the simple structure's head start dominates. The supplier either cannot yet act on the complex contract or retains too little capacity to convert its higher productivity into content. Right of k^* the complex structure's faster climb dominates. Arm averages weight these curves by the capacity distribution.

Proposition 3 (Reversal in constrained populations)**.** *Let* $\underline{k}^*$ *denote the smallest pairwise engagement crossing among the four structures and* $\underline{k}^*_\Pi$ *the smallest pairwise value crossing.*

(i) If $F(\underline{k}^*) = 1$*, expected engagement is highest under the capacity-minimizing bundle and lowest under the fully complex bundle:*

$$\mathbb{E}_F\, E(k;\, M,T) \;>\; \mathbb{E}_F\, E(k;\, c') \;>\; \mathbb{E}_F\, E(k;\, D,C), \quad c' \in \{(M,C),\, (D,T)\}.$$

(ii) If $F(\underline{k}^*_\Pi) = 1$*, the buyer's program value* $\Pi(c)$ *ranks the cells in the same pattern, and the buyer optimally chooses the capacity-minimizing bundle* (M,T).

Figure 11 collects the four curves against the capacity distribution. Proposition 3 delivers the three baseline predictions. Mandate raises engagement relative to delegation because removing the choice overhead frees capacity for execution (Hypothesis 1).

Targeted verification raises engagement relative to comprehensive verification because a single salient objective concentrates capacity (Hypothesis 2). The bundle that minimizes total overhead maximizes engagement, and the fully complex bundle ranks last (Hypothesis 3).

The proposition leaves the two single-burden cells unranked against each other. Below the smallest crossing they are ordered by their overheads, so (M, C) delivers more engagement than (D, T) when $\mu < \delta$ and less when $\mu > \delta$. The four-cell regression identifies this contrast, which is why the middle cells carry information about the relative size of the two frictions.

Proposition 1 and Proposition 3 together locate the classical prescriptions. They apply when comprehension is cheap or when capacity is abundant, and they reverse in networks whose mass sits left of the thresholds.

5.4 Extensions

Four extensions sharpen the baseline. They concern the interaction between the two overheads, the allocation of effects across tasks, the location of the reversal, and the role of the filer's own capacity.

5.4.1 Compounding Burdens

The baseline stacks the two overheads additively. Two forces make their joint burden exceed the sum of the parts for constrained suppliers. The first operates through the shared budget. Both overheads are drawn from one concave execution technology, so a second withdrawal of a given size lands on a steeper segment of the curve than the first. The second operates through comprehension itself. Under delegation with comprehensive verification, the chosen indicators determine what the audit will inspect

and the audit scope determines which indicators are worth choosing, so the supplier faces a joint interpretation problem that neither clause poses alone. Let $\gamma \geq 0$ denote this interaction, so that $r(D, C) = r_0 + \delta + \mu + \gamma$ while the other entries in (3) are unchanged.

Define the engagement interaction at capacity k as

$$I(k) \;=\; E(k;\, M, T) \;+\; E(k;\, D, C) \;-\; E(k;\, M, C) \;-\; E(k;\, D, T). \tag{8}$$

Proposition 4 (Compounding). *(i) For every $k > r(D, C)$, $I(k) < 0$.*

(ii) $I(k)$ is decreasing in γ, so a positive comprehension interaction deepens the compounding.

(iii) Measured from the fully complex cell, the burdens compound exactly when the reliefs substitute: $I(k) < 0$ if and only if the engagement gains from the two single simplifications, $E(k; M, C) - E(k; D, C)$ and $E(k; D, T) - E(k; D, C)$, sum to more than the joint gain $E(k; M, T) - E(k; D, C)$.

(iv) Consequently $\mathbb{E}_F\, I(k) < 0$ for every F supported on $(r(D, C), \bar{k}\,]$.

Part (i) holds at $\gamma = 0$, so compounding does not require the two clauses to interact in comprehension. Equal withdrawals from one budget do increasing damage as the budget shrinks. The additive form of ϕ in (4) keeps execution free of interactions. If productivity were supermodular across the two dimensions, that channel would push in the opposite direction. Part (iii) restates the same inequality from the opposite corner of the design, which is how the test is implemented, with the four cells re-expressed against the fully complex bundle.

Proposition 3 places all capacity mass below the smallest crossing, and Proposition 4(iv) places it above $r(D, C)$. The two conditions hold together whenever

$r(D, C) < \underline{k}^*$, so that every supplier can process even the fully complex contract while the network still sits left of every crossing. Hypotheses 3 and 4 are tested on the same sample under that joint restriction.

5.4.2 The Task-Complexity Gradient

The reporting program spans N tasks that differ in tractability $\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0$, which measures how readily a unit of capacity converts into reported, verifiable content. Engaging with task j requires assembling material held across functions, so effort s_j yields content $\psi(\lambda_j s_j)$ with ψ increasing, strictly concave, and a finite marginal product at zero, $\psi'(0) = \bar{\psi} < \infty$. A task is therefore worth undertaking only when the shadow value of capacity is low enough. Appendix D solves the supplier's allocation problem and establishes the following.

Proposition 5 (Activation order and incidence)**.** *Let $\chi(k; c)$ denote the shadow value of the supplier's residual capacity, decreasing in $k - r(c)$, and let the value of content be equal across tasks. There exist thresholds $\eta_1 > \eta_2 > \cdots > \eta_N$, increasing in task tractability, such that the supplier engages with task j if and only if $\chi(k; c) \leq \eta_j$. Tasks therefore activate in order of tractability as capacity grows. A governance change that lowers $r(c)$ raises $k - r(c)$, lowers χ, raises engagement on every active task, and activates the most tractable task still inactive.*

A governance change that frees capacity therefore acts on tasks in two ways. It raises engagement on tasks already undertaken, by more where tractability is higher, and it brings the most tractable inactive task above its threshold. Treatment effects should concentrate on tractable tasks and fade as task complexity rises. Where the value of content differs across tasks, the thresholds scale with that value as well, so targeted verification raises the threshold of the focal task above the ordering by

tractability alone.

5.4.3 The Capacity Threshold

The crossing in Proposition 2 makes the ranking of structures a statement about where the capacity distribution sits.

Proposition 6 (The capacity threshold)**.** *Let s and x be as in Proposition 2.*

(i) The engagement gap $\Delta_{xs}(k)$ is strictly increasing on $[\, r(x),\, \bar{k}\,]$: the gains from delegation and from comprehensive verification rise with supplier capacity at every capacity level.

(ii) Let $k^*_{\max}$ be the largest pairwise value crossing against (D, C) and suppose $k^*_{\max} < \bar{k}$. For every F' supported on $(k^*_{\max}, \bar{k}\,]$, the classical ranking of Proposition 1 holds in population: $\Pi(D, C) > \Pi(c)$ for all $c \neq (D, C)$.

Propositions 3 and 6 together delimit the two regimes. The classical rankings govern to the right of the threshold, whether because comprehension costs vanish or because the network's mass has moved past k^*, and the reversal governs to the left. Which regime applies is an empirical property of the network.

5.4.4 Dilution

The baseline treats the supplier's capacity as a primitive. Inside the firm, the capacity that reaches the task depends on who handles the filing. Let K denote the organization's capacity endowment and $\ell \geq 0$ the competing load of the person assigned to the filing, so that the capacity available to the task is $k = K - \ell$. Assignment therefore positions the same organization at different points on the capacity axis. A dedicated staff member carries low ℓ, while an owner or general manager filing personally carries high ℓ.

Proposition 7 (Dilution)**.** *Fix $H = h$ and let $\Delta_V(k) = E(k;\, M, h) - E(k;\, D, h)$ denote the vertical treatment contrast at capacity $k = K - \ell$. Then Δ_V is continuous, $\Delta_V(k) = 0$ for $k \leq r(M, h)$, and Δ_V is strictly increasing on $[\, r(M, h),\ r(D, h)\,]$. As the filer's competing load ℓ rises toward $K - r(M, h)$, every treatment contrast is attenuated toward zero. The horizontal contrast satisfies the same properties.*

Governance affects engagement only when the task is assigned capacity. Contrasts estimated on filings handled by dedicated staff recover the model's comparisons, and contrasts estimated on filings handled by top executives are attenuated toward zero and estimated imprecisely.

5.5 Testable Hypotheses

We take the model to the experiment, where $N = 3$ orders the environmental, social, and governance reporting dimensions by decreasing tractability. Environmental reporting rests on quantitative measures drawn from routine records and verified against codified protocols. Social reporting requires qualitative material held across the human-resource and administrative functions, and governance reporting requires policy documentation that must often be created before it can be evidenced. The propositions above, all proved in Appendix D, yield seven hypotheses.

Hypothesis 1 (Vertical: The Cost of Choice)**.** In a constrained network, mandate raises engagement relative to delegation within each monitoring scope (Proposition 3).

Hypothesis 2 (Horizontal: The Value of Focus)**.** In a constrained network, targeted verification raises engagement relative to comprehensive verification within each authority allocation (Proposition 3).

Hypothesis 3 (The Capacity-Minimizing Bundle)**.** Expected engagement is highest under the capacity-minimizing bundle (M, T) and lowest under the fully complex bundle (D, C), with the single-burden cells in between (Proposition 3).

Hypothesis 4 (Compounding Burdens)**.** The two overheads compound among constrained suppliers, equivalently the two reliefs substitute: the interaction I in (8) is negative, so the gains from the two single simplifications, measured from the fully complex cell, sum to more than the joint gain (Proposition 4).

Hypothesis 5 (The Task-Complexity Gradient)**.** Treatment effects concentrate on the more tractable reporting dimensions and fade with task complexity (Proposition 5).

Hypothesis 6 (The Capacity Threshold)**.** The gains from delegation and from comprehensive verification increase with supplier capacity, and the baseline ranking reverses among high-capacity suppliers (Proposition 6).

Hypothesis 7 (The Dilution Effect)**.** Treatment contrasts are large when a dedicated staff member handles the filing and attenuate when a top executive files personally (Proposition 7).

Two further observations connect the framework to the remaining evidence. First, the frictions themselves are observable. Survey measures of the constraints suppliers report bear on whether the choice cost is positive, $\delta > 0$, and on whether a broad verification scope raises the processing burden, $\mu > 0$. Suppliers more often attribute unreported items to capacity constraints than to a lack of willingness to report them. Second, capacity freed from processing may be used outside the report as well as inside it. Consistent with that reading, the lead firm's own endline ESG scores are higher in the mandate and targeted-audit arms. Both patterns are presented as empirical evidence in their own right.

6 Testing the Capacity Channel

Section 5 set out seven predictions. The first three are the main results already reported. The remaining four concern how the two burdens combine and how suppliers with different capacity levels respond to governance. This section tests them in turn, starting with whether the two burdens compound and then turning to heterogeneity across reporting tasks, across suppliers, and within suppliers. It then draws on endline-survey and reminder-call evidence to examine capacity directly and distinguish it from competing mechanisms.

6.1 Compounding Burdens

Hypothesis 4 concerns the two burdens together. Both draw on the same capacity budget, and because the execution technology is concave, a second unit of relief lands on a steeper segment of the curve than the first. Relieving one source of complexity delivers part of what relieving the other would have delivered, so the sum of the two separate reliefs exceeds the relief they deliver jointly. Proposition 4 states this in the form the design can test.

The four treatment cells identify the comparison directly. We re-express them against the fully complex cell, which combines delegated KPI selection with comprehensive monitoring. Removing the choice burden alone is the move to mandate with comprehensive monitoring. Removing the audit-scope burden alone is the move to delegation with targeted carbon monitoring. Removing both is the move to mandate with targeted carbon monitoring. The second difference is the joint gain minus the sum of the two single gains. It is zero when the two burdens are additive and negative when they compound.

Figure F3 and Table F7 report the three gains and the second difference for the five engagement outcomes. Relieving the choice burden alone adds 14.933 completed items and 9.831 percentage points of the completion share. Relieving the audit-scope burden alone adds 14.269 items and 9.429 percentage points. The two sum to 29.202 items and 19.260 percentage points. Relieving both together adds 17.625 items and 11.643 percentage points, about 60 percent of the sum. The documentation outcomes point the same way with smaller shortfalls. The joint relief is 1.750 indicators against a sum of 1.783, 1.490 attachment categories against a sum of 1.600, and 1.203 hard-evidence categories against a sum of 1.363.

The second difference is negative on all five outcomes, which is the sign Proposition 4 predicts. The proposition signs the interaction for a population that can process every structure. The near-zero coefficient on the fully complex cell is also consistent with a mass of suppliers below its entry point, in which case the second difference is bounded by the concavity channel alone. Both readings give the same sign. The individual second differences are not precisely estimated, so we read the common sign across the five outcomes as suggestive of compounding and leave its size open. The five outcomes come from the same filings, so their common sign is close to a single draw.

Two mechanisms generate this sign. Equal withdrawals from one capacity budget do increasing damage as the budget shrinks, which produces a negative interaction through the concavity of the execution technology alone. Delegated choice under comprehensive verification adds a joint interpretation problem, since the indicators the supplier selects determine what the audit will inspect and the audit scope determines which indicators are worth selecting. Both are consistent with what we find, and both imply that a buyer who simplifies one governance margin should expect less from simplifying the second than the first simplification delivered.

6.2 The Task-Complexity Gradient

Hypothesis 5 concerns heterogeneity across the reporting tasks, predicting that the response concentrates in the domains easiest to execute, where a supplier can most readily turn a request into retrievable and verifiable evidence. Tractability depends on both data accessibility and the clarity of the verification protocol. Environmental reporting often relies on quantitative records such as utility bills, metering data, and procurement documents, while social and governance reporting more often requires qualitative materials or coordination across HR, legal, and compliance functions.

Figure 12 and Table F8 separate the treatment effects by ESG domain. Mandate raises environmental completion by 14.5 percentage points and increases the probability of an environmental attachment by 28.1 percentage points. Targeted carbon monitoring raises environmental completion by 15.9 percentage points and raises environmental attachment submission by 24.3 percentage points. These four estimates are statistically significant at the 5 percent level or better.

The documentation margin follows the predicted order. Attachment estimates for social and governance information are close to zero under both simpler structures, so the gain in verifiable evidence is confined to the environmental domain. The completeness margin orders less cleanly. Mandate raises social completion by 10.5 percentage points, the smallest of its three estimates, while governance completion rises by 14.6 points under mandate and 15.6 under targeted monitoring. The governance block carries five indicators, so its completion share moves in twenty-point steps and its intervals are wide. The gradient is thus identified on the documentation margin and uninformative on the governance block of the completeness margin.

This is the pattern Hypothesis 5 predicts, with the simpler treatments most effective where a supplier can translate a clear request into verifiable evidence. The result fits

the multitask insight that effort responds to differences in measurability (Holmström and Milgrom, 1991), and adds a capacity channel. What matters alongside whether a task can be evaluated is how much internal coordination the supplier needs before there is evidence to evaluate.

6.3 The Capacity Threshold

Hypothesis 6 predicts that the gains from the complex structure on each dimension rise with supplier capacity and that the baseline ranking reverses among high-capacity suppliers. We measure capacity by registered capital, a pre-treatment item from the buyer's supplier roster. It scales with the formal administrative apparatus a supplier maintains, the standing staff and records that a new reporting requirement has to draw on. It is also filed on a common legal standard, whereas employment and production reach the buyer on definitions that differ across firms.[6] Proposition 2 places the crossing above the complex structure's entry point, and this network is one of micro and small firms, so most of the capital distribution should sit in the flat region where the complex arm does nothing. A median split would average that region into the contrast, so we set the threshold at the top quartile. It holds 38 of the 150 suppliers with reported capital, of which 9 are control, 18 mandate, and 11 delegation on the vertical dimension, and 9 control, 12 comprehensive, and 17 targeted on the horizontal one.[7] Estimates come from equation (2), so each coefficient is the effect of one arm

[6] Registered capital in China has been subscribed capital since the 2013 Company Law reform, so paid-in capital is the natural alternative. Only 128 of the 152 suppliers disclose it, too few to support a well-powered test of the threshold, and where both are observed the two rank suppliers almost identically, with a rank correlation of 0.92. The self-reported alternatives are not comparable across firms. Plant employment and annual output correlate 0.08 and 0.06 in logs with registered capital, and the median supplier reports three and a half times as many employees to the buyer as to the public registry.

[7] Assignment does not predict which suppliers land above the cut. The shares are 20, 34 and 21 percent for control, mandate and delegation, with $\chi^2(2) = 3.23$ and $p = 0.20$, and 20, 23 and 32 percent for control, comprehensive and targeted, with $\chi^2(2) = 2.03$ and $p = 0.36$. Within the group,

on suppliers in one capacity group, measured against control suppliers of the same capacity.

Figure F4 reports the authority dimension, and the two capacity groups behave differently. Below the threshold delegation moves nothing. It changes completed indicators by -0.248 and attachment coverage by -0.312 categories, neither distinguishable from zero, while mandate raises completed indicators by 14.024 and coverage by 1.132 categories. Above the threshold delegation works on every outcome, raising completed indicators by 33.454 and the completion share by 21.4 percentage points, and raising documented indicators by 4.643, coverage by 1.961 categories, and hard-evidence categories by 1.577.

These effects imply a reversal in which structure dominates. Below the threshold mandate leads delegation by 14.272 completed indicators ($p = 0.009$), 1.444 attachment categories ($p = 0.003$), and 1.342 hard-evidence categories ($p < 0.001$). Above the threshold delegation is the stronger arm on all five outcomes, by 7.536 completed indicators and 0.850 categories. The change in the gap across the two capacity groups carries the predicted sign on every outcome and is significant at the ten percent level on four of the five, the exception being documented indicators. Clustering on the strata makes all five significant, and a wild-cluster bootstrap leaves three of them so.

Figure F5 reports the monitoring dimension, where the capacity gradient is at least as strong but the ranking does not turn. Above the threshold targeted monitoring raises completed indicators by 42.573 and the completion share by 27.8 percentage points, and raises documented indicators by 5.552, coverage by 2.615 categories, and hard-evidence categories by 1.738, all significant at the one percent level. Comprehensive monitoring above the threshold reaches 11.578 completed indicators but is not distinguishable from

no authority or monitoring arm differs from control on production, firm age, or registered capital, and employment differs only for targeted monitoring, and only at ten percent.

zero on any of the five outcomes. Below the threshold the two monitoring arms are indistinguishable from each other and from the control group. The pooled advantage of targeted monitoring reported in Section 4.2 is therefore concentrated among suppliers above the capital threshold, where it reaches 30.996 completed indicators ($p < 0.01$) and 2.624 attachment categories ($p < 0.01$). Targeted monitoring remains the stronger arm at every capacity level we observe.

Figures F6 and F7 re-estimate both dimensions with the threshold set at the top 50, 30, 25, and 20 percent of registered capital. The authority pattern holds at every cut. Delegation is flat below the threshold and raises completed indicators above it by 17.1, 31.8, 33.5, and 25.9, significant at five percent at the first three cuts and at ten percent at the last. The gap between the arms narrows and changes sign each time. On the monitoring dimension targeted monitoring leads throughout.

These estimates place the vertical crossing of Proposition 6 inside the capacity range this network spans, so Hypothesis 6 holds on the authority dimension. Below the threshold, the capacity constraint binds and the simpler structure wins; above it, the advantages of the more complex structure emerge, consistent with the model's prediction. The horizontal dimension shows no such crossing within our sample. Comprehensive monitoring imposes a larger overhead than delegation does, since a broad audit charges by volume, exposing every field a supplier reports to verification on every dimension, so its crossing point likely lies above the capacity levels present in our network. Suppliers in our sample may simply lack the capacity needed for comprehensive monitoring's productivity gain to dominate its overhead. The vertical pattern is consistent with evidence that the returns to management practices and to decentralization depend on baseline organizational capability (Bloom et al., 2013, 2020; Aghion et al., 2021) and that upgrading in global value chains requires complementary firm capabilities (Verhoogen, 2023).

6.4 Respondent Role and the Dilution of Treatment Effects

Hypothesis 7 moves the capacity test inside the supplier. Senior managers have broad responsibilities outside ESG reporting, so the capacity available to the platform task may be especially limited when an owner or top executive completes the filing directly. Proposition 7 predicts that treatment contrasts shrink when the reporting task receives too little dedicated capacity.

We measure this using the identity of the supplier's standing contact on the buyer's platform, a role recorded on the contact roster when the account was opened, before any treatment interface went live. This role is classified for 702 of the 735 supplier-rounds, and about one supplier in seven files through an owner, general manager, or director. Because the contact was named prior to assignment, this split reflects a firm characteristic fixed before treatment rather than a response to it, and assignment indeed does not predict who files: the share of firms whose filer is an owner, general manager, or director is similar across the five cells, and regressing that indicator on the arm indicators returns coefficients of at most 0.05, none of them significant.

We estimate equation (2) with the executive indicator in place of the capacity indicator, so each of the four coefficients is an arm effect measured against control suppliers filing through the same kind of respondent. Figures F8 and F9, with Table F9, report the estimates. Engagement is concentrated where the task has someone to absorb it. When a staff member handles the filing, mandate raises completed indicators by 17.355 and the completion rate by 11.4 percentage points, and adds 1.236 attachment categories and 0.930 categories of hard evidence. Targeted monitoring raises completed indicators by 16.298 and the completion rate by 10.7 percentage points. These are the main results of Section 4, recovered among the suppliers who give the task dedicated time. Delegation and comprehensive monitoring

move nothing even there, and within the staff group mandate exceeds delegation at the one percent level on attachment categories and hard evidence. When an owner, general manager, or director files, no arm moves any outcome and none differs from the control group. The four completeness estimates lie between -11.179 and -1.116 completed indicators, and the executive-group arms are jointly indistinguishable from zero ($p = 0.458$) against staff-filer arms that are jointly significant ($p = 0.019$).

This is what Proposition 7 describes. Engagement requires capacity above the structure's entry point, and the capacity that binds is what the filer can give the task, whatever the firm holds. A dedicated respondent clears the entry point of the simpler structure and the mandate delivers. A principal whose attention is spread across every function of the firm clears none of them, and the governance structure stops mattering. Read as four arm effects, the estimates say exactly that, with engagement appearing where the simplest structure meets a dedicated respondent and nowhere else.

These within-firm patterns align with evidence on the scarcity of executive attention. Time-use studies show that top managers' time is the binding resource at the apex of the firm and that its allocation across competing demands predicts firm performance (Bandiera et al., 2020). Theories of organizational focus reach the same conclusion from the opposite direction. Because attention is rationed, organizations gain when routine information tasks are lodged with dedicated personnel, because a leader's attention must span many functions (Dessein et al., 2016). Our estimates carry this logic across the firm boundary. An owner or general manager who files personally allocates a thin slice of a heavily demanded resource, while a staff assignment gives the task dedicated execution capacity. Where the task is lodged with dedicated staff the governance structure matters and the simpler one wins. Where it competes with everything else on a principal's desk, the structure stops mattering at all.

6.5 Capacity Is Scarce and Predicts Reporting Gaps

To examine the mechanisms underlying the treatment effects, we administered an endline survey through the lead firm, which yielded 135 responses from suppliers in the experimental sample. Response patterns are balanced across arms (Figure F10). Asked why they left fields unreported despite wanting to provide the information, the most frequently cited reason is unavailable data, followed by the judgment that an item is irrelevant to the business and then by unclear requirements. Unavailable data, missing staff, and time cost account for 74 of the 142 citations, and unclear requirements, the survey's counterpart to the comprehension cost, for a further 29 (Figure F11). The organizational infrastructure measures point the same way. Sixty-five percent report no dedicated ESG personnel and 62 percent no centralized data system, and suppliers lacking either leave more items unreported and cite more binding constraints (Figures F12 and F13). These descriptive relationships connect the model's capacity stock to observable staffing and information-system limitations, and Appendix E reproduces the instrument and describes the measures.

6.6 Choice, Focus, and Task Tractability

Given that capacity is scarce, we ask whether the treatments operate through it. The model attributes the engagement effects to a mediating channel. Simpler structures consume less of the supplier's limited capacity in overhead, leaving more for execution. If that channel is operative, the capacity constraints we identified should register as a different perceived burden across the arms, which in turn shapes engagement. The survey measures this intermediate step directly, so we regress the survey-based burden and constraint measures on the same treatment contrasts used for the platform outcomes. Appendix E documents each measure.

The pattern runs in the direction the channel requires. The arms that produce weaker platform engagement are exactly those in which suppliers describe tighter capacity. Delegation and comprehensive monitoring raise the capacity barriers, time shortages, and demand for fewer fields that mandate and targeted monitoring leave near control levels (Figures F14 and F15). These contrasts show that the arms differ in reported burden. How much of the engagement gap that burden carries is beyond what they identify. Appendix E discusses each figure in detail.

6.7 Engagement During Reminder Calls

The September 2025 briefings described in Section 3.3 were structured reminder calls placed ahead of the upcoming filing. Each call reviewed the supplier's own indicators and verification scope, addressing only that supplier's arrangement to preserve the experiment's naturalistic framing. The team recorded and transcribed 150 conversations, of which 146 match a supplier in the experimental sample and enter the estimation sample. From these calls we construct four engagement measures: the supplier's speech share, its number of substantive questions, an indicator for written follow-up, and an equal-weighted index of the three. Appendix E gives the exact definitions. We estimate treatment contrasts in the call-level cross-section with randomization-stratum and caller fixed effects, clustering standard errors by stratum. Suppliers were assigned to the two callers at random, and this assignment is balanced across arms, but the two callers differ in conversational style, so the caller fixed effects absorb this difference and sharpen the estimates.

Figures F16 and F17, together with Table F10, report the results. Suppliers assigned to targeted carbon monitoring speak for a 7.1 percentage point larger share of the conversation. Their estimates are also positive for questions, written follow-up, and

the engagement index. Mandate produces positive point estimates on the same four measures. The call evidence complements the platform outcomes. Simpler governance changes how a supplier engages with the request, upstream of what it eventually enters in the reporting system.

6.8 Alternative Mechanisms

The findings so far support the capacity channel, under which governance shapes engagement by conserving or consuming the capacity suppliers can devote to reporting. The same patterns could arise through other channels, so we examine the most plausible alternatives below. Each can rationalize part of what we find, and we do not claim that any is absent. None, however, reproduces the heterogeneity and the survey evidence at the same time, which leaves a binding internal capacity constraint as the most plausible reading of the full set of results.

One alternative is that the schemes signal the lead firm's priorities. Mandate identifies the indicators the buyer values, and a carbon audit identifies the outcome it will scrutinize. Suppliers may therefore respond more under mandate than delegation and under targeted than comprehensive monitoring because the buyer's intent is clearer. This account can explain the main effects and their concentration in environmental reporting, which the carbon audit makes especially salient. It does not, however, explain the full pattern across supplier capacity. The returns to Delegation and Targeted Monitoring rise with pre-treatment registered capital on all five engagement outcomes (Section 6.3). A capacity-dependent response to a common signal could generate heterogeneity, but it provides no reason for the Mandate advantage to narrow and then reverse, with Delegation overtaking Mandate above the threshold. Suppliers in a given treatment arm receive the same signal. The reversal instead follows from

binding capacity constraints: better-resourced suppliers can turn greater autonomy and a narrower monitoring task into fuller reporting and broader documentation.

A second explanation is that Mandate and Targeted Monitoring reduce uncertainty by clarifying what suppliers must file and what the buyer will check. This is closely related to the same primitive in the model. The comprehension cost in equation (3) captures the burden that an unclear request imposes on a supplier. This interpretation has independent support in settings where agents fail to act on aspects of a task that they have not learned to notice (Hanna et al., 2014). Clarity alone, however, should generate similar gains regardless of who completes the filing or how much administrative slack the firm has. Instead, the gains appear when dedicated staff handle reporting and fade when executives file personally (Section 6.4). They also rise with registered capital (Section 6.3). Clarity is one benefit of the simpler structures, but it does not by itself account for the results.

A third explanation is selective disclosure under Delegation. Suppliers may select indicators for which records are readily available. This is the reporting analogue of the strategic concealment emphasized in models of supplier auditing (Plambeck and Taylor, 2016). Completion is computed over the platform's full applicable field set. Suppliers can therefore reduce their reporting burden by concentrating on convenient fields without increasing measured coverage. This channel may explain part of Delegation's weaker performance. It cannot explain the monitoring result, however, because suppliers do not select the verification scope. Targeted Monitoring generates more engagement than Comprehensive Monitoring among suppliers facing identical indicators (Section 4.2). The common mechanism behind both margins is the reporting capacity that the governance structure absorbs before reporting begins.

7 Firm Performance and Business Outcomes

7.1 ESG Performance Rated by the Lead Firm

Does the capacity released by simpler governance improve buyer-valued performance? Platform entries cannot answer this question, since suppliers choose which indicators to complete and report different indicators in different units. We therefore use endline ESG ratings from the lead firm's professional supply chain team, who were blind to treatment and could evaluate performance consistently across supplier types. Although the firm does not disclose the rating methodology, the scores draw on submitted information and on soft information from the team's ongoing interactions with suppliers. They reflect the team's assessment of suppliers' production and operations, engagement with the program, and reporting practices. The ratings thus provide a buyer-side measure of performance that is distinct from suppliers' platform entries and does not mechanically overlap with the platform governance measures.

Figure F18 and Table F11 report two endline outcomes. Mandate raises the transparency score by 0.105 and the performance score by 0.083. Targeted carbon monitoring raises the same outcomes by 0.104 and 0.102. Delegation and comprehensive monitoring generate smaller estimates. The individual coefficients are imprecisely estimated in the cross-section of 155 suppliers.

The ordering is suggestive. Mandate and targeted monitoring, the two structures that generate the strongest engagement, also earn the highest marks from the blinded assessment, and the transparency and performance components move together within each arm. This is consistent with the additional reporting induced by simpler governance containing substance that the buyer's own experts recognize, beyond pro-forma completion of required fields. The alignment matters for program design.

The same structures that draw constrained suppliers into the system also produce the information the procurement organization rates highly.

7.2 Financial and Operational Outcomes

The model does not sign the effect of ESG reporting on production. A central multitasking concern is that new administrative requirements draw resources away from production, employment, or commercial activity, so we examine whether the engagement gains coincide with such displacement. Five outcomes are internal to the supplier's own business, and the sixth, exit from the buyer network, records continued participation. Revenue and employment enter in logs and the remaining outcomes as they are measured. Each outcome is one post-program value per supplier, so we cluster by randomization stratum.

Figures F19 and F20, together with Table F12, report annual revenue, total employment, the voluntary worker turnover rate, the overall on-time delivery rate, the overall return rate, and exit from the buyer network. No arm moves any of them. Mandate raises log revenue by 0.418 with a standard error of 0.382 and log employment by 0.072 with a standard error of 0.061, and it changes the probability of exit by -0.012 with a standard error of 0.031. The remaining twenty-one estimates are of the same order, and none of the twenty-four is distinguishable from zero. The estimates establish the absence of displacement.

These estimates do not document the crowding out that the multitask framework predicts when agents absorb new demands on a fixed resource base, and that empirical work often finds when effort is redirected toward newly measured tasks (Holmström and Milgrom, 1991; Hong et al., 2018). The relational nature of the buyer-supplier tie offers a plausible explanation. The lead firm and its suppliers transact repeatedly

under incomplete contracts, so cooperation is sustained by the value of the relationship (Baker et al., 2002; Boudreau et al., 2023). Visible engagement with the buyer's reporting program then carries a return of its own, since it signals reliability, builds trust, and raises the continuation value of a match that carries substantial economic value (Macchiavello and Morjaria, 2015, 2021). The model rations the supplier's capacity and so implies some reallocation away from competing uses, and a buyer that steers orders toward the suppliers it sees engaging would offset that reallocation. The estimates cannot pin down the route from reporting to commercial performance. They do show that the program did not come at the expense of the core business.

8 Discussion

8.1 Implications for Organizational Theory

Our results generate three implications for organizational economics. First, the scope conditions of delegation theory bind in this inter-firm setting. Classical models predict that granting agents authority over task selection exploits private information and improves decisions (Aghion and Tirole, 1997; Dessein, 2002). Our heterogeneity results show that these benefits accrue to suppliers with organizational slack, while for the median supplier the cost of evaluating unfamiliar indicators dominates the gain from local tailoring. The model carries this tradeoff in the comprehension cost δ of (3) and the productivity gain π_D of (4), and Proposition 6 resolves it monotonically in capacity. The charge is fixed and paid before any effort, so its bite falls as capacity grows, while the gain applies to a widening base of execution. Delegation becomes productive once the supplier can absorb the charge and use the gain.

This result helps reconcile the experiment with evidence that delegation can improve

productivity inside well-resourced firms (Bloom et al., 2012; Bandiera et al., 2021). The relevant distinction is whether the informed agent has the administrative structure needed to use local information while completing the assigned task. Organizational recommendations drawn from within-firm settings may therefore travel poorly to small independent suppliers unless the agent's implementation capacity is made explicit.

Second, the results extend multitasking theory to the extensive margin of engagement. Standard models emphasize effort substitution once an agent is already participating (Holmström and Milgrom, 1991). In our setting, the breadth of monitoring also determines whether the supplier begins the task and how broadly it documents what it reports, and, among participating suppliers, where the documentation response lands. Targeted monitoring provides a clear entry point and generates a broader engagement response, and its advantage over comprehensive verification widens with supplier capacity. The two overheads scale differently. The cost of evaluating an unfamiliar indicator set is paid once, so capacity relaxes it. The cost of preparing for a broad audit rises with the number of fields the supplier files, and therefore grows as the supplier reports more. This administrative-burden channel complements research showing that the design and targeting of verification affect regulated behavior (Duflo et al., 2013, 2018) and that a buyer's monitoring program is judged by its structure as much as by its presence (Short et al., 2016; Plambeck and Taylor, 2016).

Third, the experiment shows how governance instruments combine through a common resource constraint. Mandate and targeted monitoring operate on different institutional margins, but both preserve the capacity available for execution. Their combination is the only cell that improves both completeness and documentation by a statistically significant margin. Because both levers draw on the same budget, their comprehension costs compound, and the engagement ranking of the four bundles tracks that accumulated cost. This is what Proposition 4 predicts, and Section 6.1

reports the test, where the second differences are negative on every outcome. Decision rights and monitoring systems should therefore be analyzed jointly. A governance component that performs well in isolation may deliver less when paired with a second component that recreates a large coordination burden.

8.2 Implications for Supply Chain Governance Practice

For practitioners, the central implication is that governance complexity is itself a cost. Lead firms often seek richer information by offering suppliers flexibility and expanding verification across many dimensions. Those choices can be valuable when suppliers have dedicated staff, integrated data systems, and reporting experience, but early in a program they absorb the same resources meant for compliance.

Several design principles follow. First, governance should be matched to supplier capability. Directive and focused arrangements provide a practical starting point for small suppliers with limited staff and fragmented records. Delegation and comprehensive monitoring can expand as suppliers develop internal capacity and demonstrate familiarity with the reporting process. This makes governance a lifecycle choice, revisited as capability grows (Locke, 2013).

Second, the sequencing of requirements matters. Metrics with clear standards and accessible records are a natural entry point, letting suppliers build routines for data retrieval, documentation, and verification before moving to qualitative or cross-functional requirements. The concentration of responses in environmental reporting suggests that a phased program can deepen initial participation while preserving comprehensive ESG coverage as a longer-run objective.

Third, authority and monitoring should be designed together. A mandate that clarifies the KPI set loses part of its value when paired with a verification scope that

still requires broad internal coordination, and targeted monitoring is less effective when suppliers must first evaluate a large menu of tasks. The mandate and targeted-monitoring combination performs best because it clarifies both what the supplier should report and where to focus its preparation.

The findings also inform private regulation in global value chains, where lead firms increasingly pass public disclosure and due-diligence requirements down to supplier networks of small enterprises. Uniform requirements turn regulatory breadth into administrative burden at the supplier level (Boudreau, 2024; Chen et al., 2018; Kalkanci and Plambeck, 2020), while frameworks that permit prioritization, staged implementation, and capacity-sensitive monitoring can raise early participation and still hold broad accountability as the objective.

8.3 Limitations and Extensions

Three questions follow from what this experiment settles, and each marks a direction for further work.

The first is the range over which the ranking applies. We estimate it in one apparel network in China, and Section 6.3 shows where inside that network the ranking changes sign. Its location depends on the comprehension cost of a governance structure and on the capacity a supplier brings to it, and both differ across industries with the maturity of reporting standards. The more valuable extension treats that cost as a choice. A buyer that supplies templates, trains its suppliers, or drafts plainer terms moves the crossing. Whether that investment pays better than simplifying the contract is worth knowing, and a lead firm has to choose between the two.

The second is the reach of the outcomes. Completeness and documentation record what a supplier reports and can substantiate, one step short of emissions and labor

conditions. A longer panel would show whether the reporting a simpler contract elicits carries through to certification, sourcing decisions, and measured environmental performance, and whether the governance ranking is the same there.

The third is the measurement of capacity. We recover it from baseline firm characteristics, the role of the filer, and endline reports on staffing and data systems, each capturing part of the construct the model defines. Time-use data and records of cross-functional coordination would measure it directly and would let future work ask whether governance builds the capacity it draws on. If a mandate leaves a supplier better able to handle a delegated choice later, then the phased governance of Section 8.2 is an investment, and a buyer can raise the capacity it faces.

9 Conclusion

A growing share of world production is coordinated by contract between independent firms, where a buyer faces the two questions a manager faces inside a firm. How much discretion should the agent hold, and how broadly should its performance be verified? We randomized both terms of a live supplier program, and the answers organizational economics developed for the interior of the firm did not carry across its boundary. Mandating the reported indicators raised the share of applicable fields a supplier fills in by 10.4 percentage points, 19.4 percent of the control mean, and widened documented evidence from 1.9 to 2.9 categories. Delegating the same choice moved neither margin. Narrowing verification to carbon did as much as the mandate, while broadening it did no more than asking for nothing at all. The buyer's blinded assessors rated the simpler arms highest, and the core business showed no displacement.

A binding constraint on organizational capacity accounts for these patterns. Delegation and comprehensive monitoring consume capacity before execution, and mandate

and targeted monitoring preserve it. The survey, respondent-role, and reminder-call evidence all point the same way. The classical benefits of autonomy and broad measurement return as capacity grows, so the governance that works depends on the suppliers a buyer has. As lead firms increasingly serve as private regulators beyond their legal boundary (Gereffi et al., 2005; Boudreau et al., 2023), governance that clarifies priorities, focuses verification, and expands with supplier capability is more likely to turn external pressure into sustained participation and substantive improvement.

References

Acemoglu, Daron, Philippe Aghion, Claire Lelarge, John Van Reenen, and Fabrizio Zilibotti, "Technology, Information, and the Decentralization of the Firm," *Quarterly Journal of Economics*, 2007, *122* (4), 1759–1799.

Aghion, Philippe and Jean Tirole, "Formal and Real Authority in Organizations," *Journal of Political Economy*, 1997, *105* (1), 1–29.

_ , Nicholas Bloom, Brian Lucking, Raffaella Sadun, and John Van Reenen, "Turbulence, Firm Decentralization, and Growth in Bad Times," *American Economic Journal: Applied Economics*, 2021, *13* (1), 133–169.

Alonso, Ricardo and Niko Matouschek, "Optimal Delegation," *Review of Economic Studies*, 2008, *75* (1), 259–293.

Amengual, Matthew and Greg Distelhorst, "Labor Standards in the Gap Inc. Supply Chain," *Strategic Management Journal*, 2025, *46* (11), 2663–2689.

Antràs, Pol and Davin Chor, "Organizing the Global Value Chain," *Econometrica*, 2013, *81* (6), 2127–2204.

_ and Elhanan Helpman, "Global Sourcing," *Journal of Political Economy*, 2004, *112* (3), 552–580.

Athey, Susan and John Roberts, "Organizational Design: Decision Rights and Incentive Contracts," *American Economic Review*, 2001, *91* (2), 200–205.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen, "Organizational Barriers to Technology Adoption: Evidence

from Soccer-Ball Producers in Pakistan," *Quarterly Journal of Economics*, 2017, *132* (3), 1101–1164.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 2002, *117* (1), 39–84.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat, "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," *Quarterly Journal of Economics*, 2021, *136* (4), 2195–2242.

_ , Stephen Hansen, Andrea Prat, and Raffaella Sadun, "CEO Behavior and Firm Performance," *Journal of Political Economy*, 2020, *128* (4), 1325–1369.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment," *Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bhargava, Saurabh and Dayanand Manoli, "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

_ , Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

_ , **Raffaella Sadun, and John Van Reenen**, "The Organization of Firms Across Countries," *Quarterly Journal of Economics*, 2012, *127* (4), 1663–1705.

_ , **Renata Lemos, Raffaella Sadun, and John Van Reenen**, "Healthy Business? Managerial Education and Management in Health Care," *Review of Economics and Statistics*, 2020, *102* (3), 506–517.

Bolton, Patrick and Antoine Faure-Grimaud, "Satisficing Contracts," *Review of Economic Studies*, 2010, *77* (3), 937–971.

Boudreau, Laura, "Multinational Enforcement of Labor Law: Experimental Evidence on Strengthening Occupational Safety and Health Committees," *Econometrica*, 2024, *92* (4), 1269–1308.

_ , **Julia Cajal-Grossi, and Rocco Macchiavello**, "Global Value Chains in Developing Countries: A Relational Perspective from Coffee and Garments," *Journal of Economic Perspectives*, 2023, *37* (3), 59–86.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico," *Journal of Political Economy*, 2018, *126* (2), 635–687.

Brynjolfsson, Erik and Paul Milgrom, "Complementarity in Organizations," in Robert Gibbons and John Roberts, eds., *The Handbook of Organizational Economics*, Princeton, NJ: Princeton University Press, 2013, pp. 11–55.

Chen, Li and Hau L. Lee, "Sourcing Under Supplier Responsibility Risk: The Effects of Certification, Audit, and Contingency Payment," *Management Science*, 2017, *63* (9), 2795–2812.

Chen, Yi-Chun, Mingyi Hung, and Yongxiang Wang, "The Effect of Mandatory CSR Disclosure on Firm Profitability and Social Externalities: Evidence from China," *Journal of Accounting and Economics*, 2018, *65* (1), 169–190.

Christensen, Hans B., Luzi Hail, and Christian Leuz, "Mandatory CSR and Sustainability Reporting: Economic Analysis and Literature Review," *Review of Accounting Studies*, 2021, *26* (3), 1176–1248.

Deshpande, Manasi and Yue Li, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 2019, *11* (4), 213–248.

Dessein, Wouter, "Authority and Communication in Organizations," *Review of Economic Studies*, 2002, *69* (4), 811–838.

_ , Andrea Galeotti, and Tano Santos, "Rational Inattention and Organizational Focus," *American Economic Review*, 2016, *106* (6), 1522–1536.

_ , Luis Garicano, and Robert Gertner, "Organizing for Synergies," *American Economic Journal: Microeconomics*, 2010, *2* (4), 77–114.

Distelhorst, Greg, Jens Hainmueller, and Richard M. Locke, "Does Lean Improve Labor Standards? Management and Social Performance in the Nike Supply Chain," *Management Science*, 2017, *63* (3), 707–728.

Doshi, Anil R., Glen W. S. Dowell, and Michael W. Toffel, "How Firms Respond to Mandatory Information Disclosure," *Strategic Management Journal*, 2013, *34* (10), 1209–1231.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Ex-

perimental Evidence from India," *Quarterly Journal of Economics*, 2013, *128* (4), 1499–1545.

_ , _ , _ , and _ , "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India," *Econometrica*, 2018, *86* (6), 2123–2160.

Fehr, Ernst, Holger Herz, and Tom Wilkening, "The Lure of Authority: Motivation and Incentive Effects of Power," *American Economic Review*, 2013, *103* (4), 1325–1359.

Finkelstein, Amy and Matthew J. Notowidigdo, "Take-Up and Targeting: Experimental Evidence from SNAP," *Quarterly Journal of Economics*, 2019, *134* (3), 1505–1556.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, *108* (5), 874–904.

Gereffi, Gary, John Humphrey, and Timothy Sturgeon, "The Governance of Global Value Chains," *Review of International Political Economy*, 2005, *12* (1), 78–104.

Ghosh, Sukti and Jasjit Singh, "Eliciting Supplier Cooperation for Value Chain Decarbonization: A Field Experiment with Smallholder Farmers in India," *Management Science*, 2026, *72.* forthcoming.

Gibbons, Robert, "Incentives Between Firms (and Within)," *Management Science*, 2005, *51* (1), 2–17.

Guo, Ruixue, Hau L. Lee, and Robert Swinney, "Responsible Sourcing in Supply Chains," *Management Science*, 2016, *62* (9), 2722–2744.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning Through Noticing: Theory and Evidence from a Field Experiment," *Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, 2004, *42* (4), 1009–1055.

Holmström, Bengt and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 1991, *7* (Special Issue), 24–52.

_ **and** _ , "The Firm as an Incentive System," *American Economic Review*, 1994, *84* (4), 972–991.

Hong, Fuhai, Tanjim Hossain, John A. List, and Tomomi Tanaka, "Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories," *International Economic Review*, 2018, *59* (2), 511–536.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 1997, *87* (3), 291–313.

Jin, Ginger Zhe and Phillip Leslie, "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*, 2003, *118* (2), 409–451.

Kala, Namrata, "The Impacts of Managerial Autonomy on Firm Outcomes," *Econometrica*, 2024, *92* (6), 1777–1800.

Kalkanci, Basak and Erica L. Plambeck, "Managing Supplier Social and Environmental Impacts with Voluntary versus Mandatory Disclosure to Investors," *Management Science*, 2020, *66* (8), 3311–3328.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Do Financial Concerns Make Workers Less Productive?," *Quarterly Journal of Economics*, 2025, *140* (1), 635–689.

KC, Diwas Singh and Christian Terwiesch, "Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations," *Management Science*, 2009, *55* (9), 1486–1498.

Kitzmüeller, Markus and Jay Shimshack, "Economic Perspectives on Corporate Social Responsibility," *Journal of Economic Literature*, 2012, *50* (1), 51–84.

Lafontaine, Francine and Margaret Slade, "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 2007, *45* (3), 629–685.

Lin, Winston, "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *Annals of Applied Statistics*, 2013, *7* (1), 295–318.

Locke, Richard M., *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy*, Cambridge: Cambridge University Press, 2013.

Lys, Thomas, James P. Naughton, and Clare Wang, "Signaling through Corporate Accountability Reporting," *Journal of Accounting and Economics*, 2015, *60* (1), 56–72.

Macchiavello, Rocco and Ameet Morjaria, "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *American Economic Review*, 2015, *105* (9), 2911–2945.

_ **and** _ , "Competition and Relational Contracts in the Rwanda Coffee Chain," *Quarterly Journal of Economics*, 2021, *136* (2), 1089–1143.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science*, 2013, *341* (6149), 976–980.

Manthei, Kathrin and Dirk Sliwka, "Multitasking and Subjective Performance Evaluations: Theory and Evidence from a Field Experiment in a Bank," *Management Science*, 2019, *65* (12), 5861–5883.

McKenzie, David and Christopher Woodruff, "Business Practices in Small Firms in Developing Countries," *Management Science*, 2017, *63* (9), 2967–2981.

Milgrom, Paul and John Roberts, "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, 1990, *80* (3), 511–528.

Mookherjee, Dilip, "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 2006, *44* (2), 367–390.

Ocasio, William, "Towards an Attention-Based View of the Firm," *Strategic Management Journal*, 1997, *18* (S1), 187–206.

Plambeck, Erica L. and Terry A. Taylor, "Supplier Evasion of a Buyer's Audit: Implications for Motivating Supplier Social and Environmental Responsibility," *Manufacturing & Service Operations Management*, 2016, *18* (2), 184–197.

Rasul, Imran and Daniel Rogger, "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal*, 2018, *128* (608), 413–446.

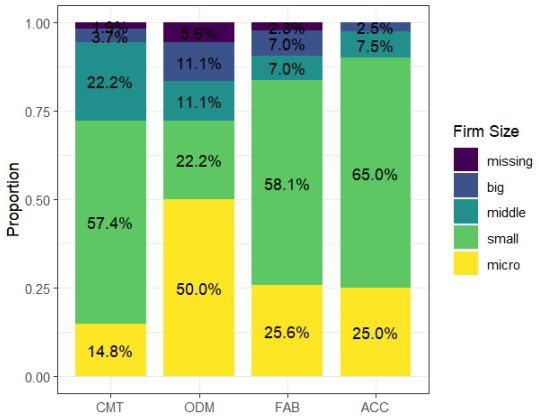
Short, Jodi L., Michael W. Toffel, and Andrea R. Hugill, "Monitoring Global Supply Chains," *Strategic Management Journal*, 2016, *37* (9), 1878–1897.

Tirole, Jean, "Cognition and Incomplete Contracts," *American Economic Review*, 2009, *99* (1), 265–294.

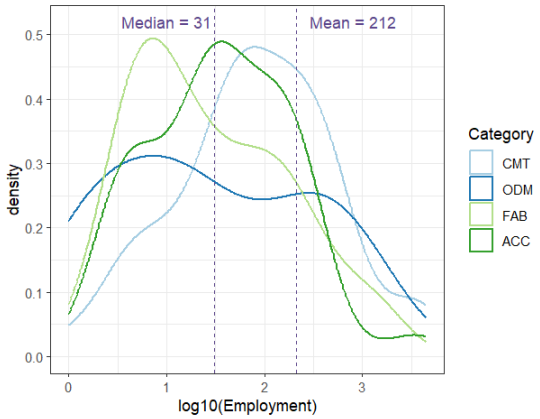
Tomar, Sorabh, "Greenhouse Gas Disclosure and Emissions Benchmarking," *Journal of Accounting Research*, 2023, *61* (2), 451–492.

Verhoogen, Eric, "Firm-Level Upgrading in Developing Countries," *Journal of Economic Literature*, 2023, *61* (4), 1410–1464.

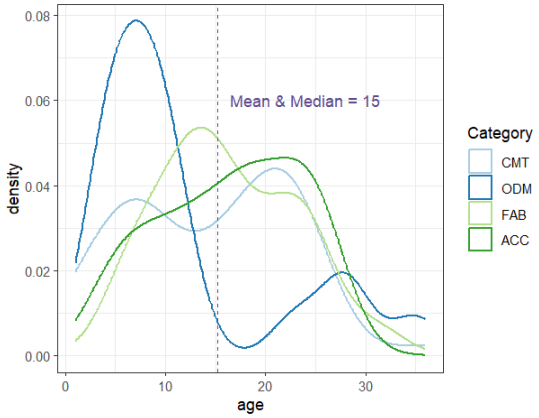
Figures and Tables



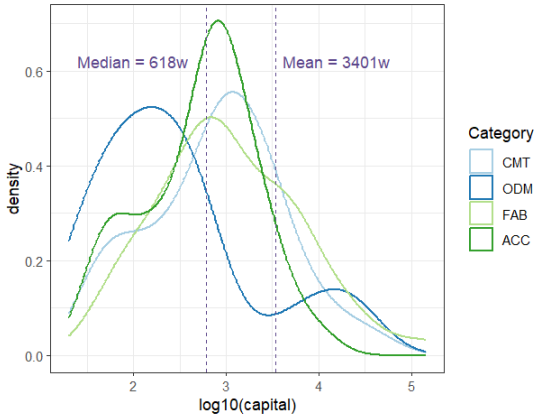
Panel A: Supplier Size



Panel B: Supplier Employment



Panel C: Supplier Age



Panel D: Supplier Registered Capital

Figure 1: Characteristics of the Supplier Network

Notes: This figure presents the distribution of key characteristics for the 155 suppliers in the experimental sample. Panel A displays firm size by supplier type: Cut-Make-Trim (CMT), Original Design Manufacturer (ODM), fabric and material supplier (FAB), and ancillary material supplier (ACC). Panel B reports total employment. Panel C reports firm age in years since establishment. Panel D reports registered capital in CNY ten thousands. The sample is concentrated among micro and small enterprises, with a median employment of 31 workers and a median firm age of 15 years. These characteristics illustrate the limited administrative capacity that motivates the experimental design.



Figure 2: Digital Platform Interface: Welcome Page

Notes: This figure displays the platform Welcome Page, with English translation, presented to suppliers upon login at the start of each reporting cycle. The text introduces the ESG reporting requirements and the scope of third-party verification assigned to the supplier. The page layout is identical across users, while the substantive content varies with the supplier's randomly assigned treatment condition. The Welcome Page is therefore the primary channel through which suppliers learn whether KPI selection is mandated or delegated and whether monitoring is comprehensive or targeted toward carbon-related metrics.

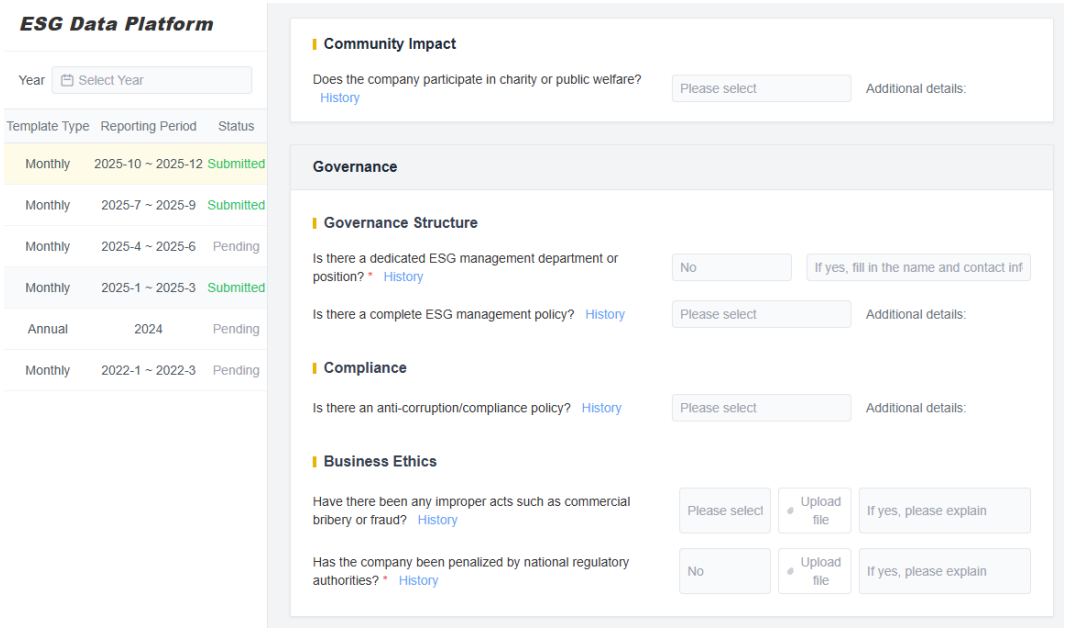


Figure 3: Digital Platform Interface: Main Reporting Dashboard

Notes: This figure shows the main reporting dashboard, with English translation, where suppliers enter ESG data. The interface organizes reporting requirements into environmental, social, governance, financial, and certification tabs. For each active indicator, suppliers may enter data, add qualitative remarks, and upload supporting documentation. Historical submissions are displayed alongside current-period fields. The set of required ESG indicators changes with the supplier's authority-allocation assignment, while the financial and certification fields remain common to all suppliers.

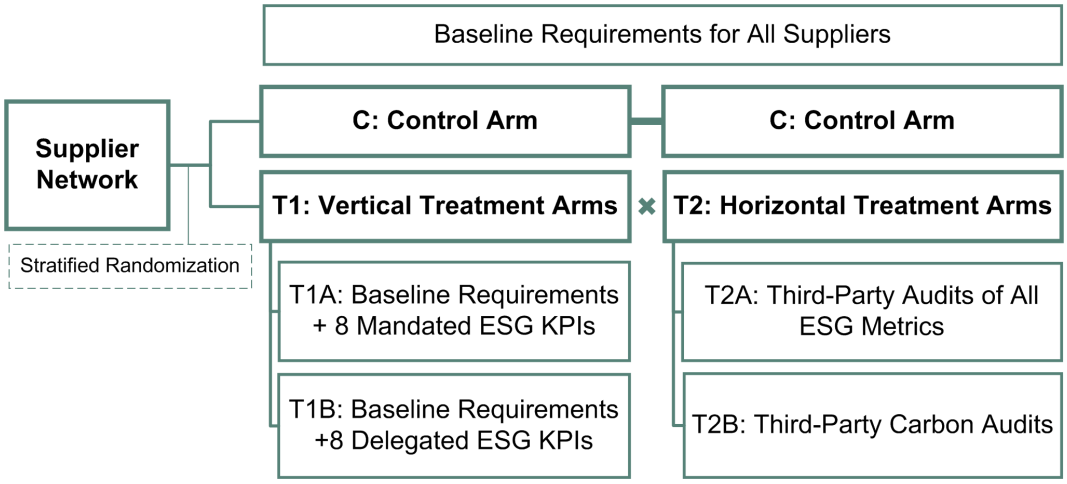


Figure 4: The 2×2 Experimental Design

Notes: This figure illustrates the experimental assignment. All suppliers face common baseline reporting requirements covering financial and certification indicators. Treatment suppliers are cross-randomized along two dimensions: vertical authority allocation over KPI selection, comparing Mandate with Delegation, and horizontal monitoring scope, comparing Comprehensive Monitoring with Targeted Carbon Monitoring. Crossing these two dimensions yields four treatment cells alongside a control group assigned to baseline reporting only.

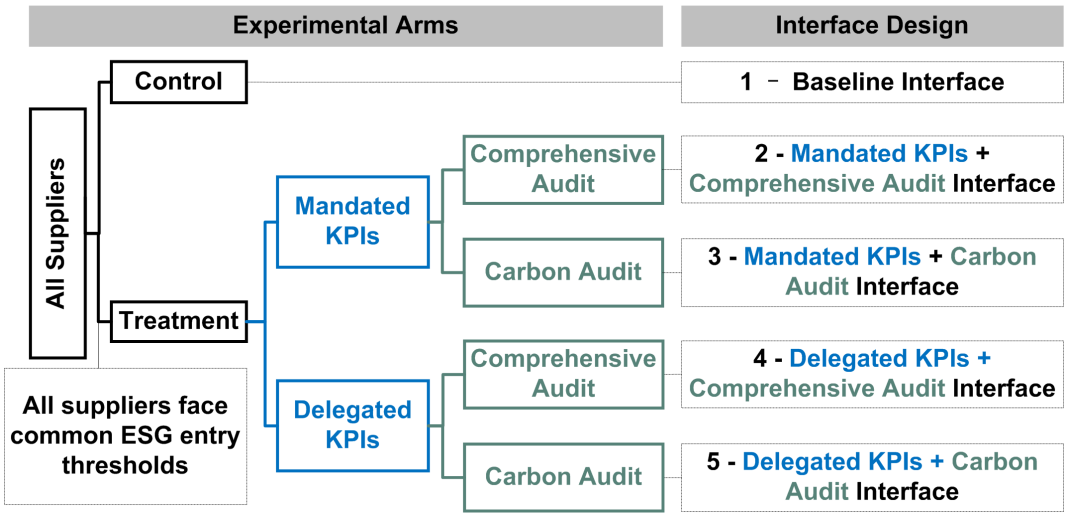


Figure 5: Interface Design by Experimental Arm

Notes: This figure maps the five experimental conditions to their platform implementations. Interface 1 is assigned to the control group. Interfaces 2 through 5 correspond to the four treatment cells formed by crossing authority allocation and monitoring scope. The assigned interface determines both the instructions on the Welcome Page and the data-entry fields marked as required on the reporting dashboard. Group-specific screenshots and English translations are reproduced in Appendix B.

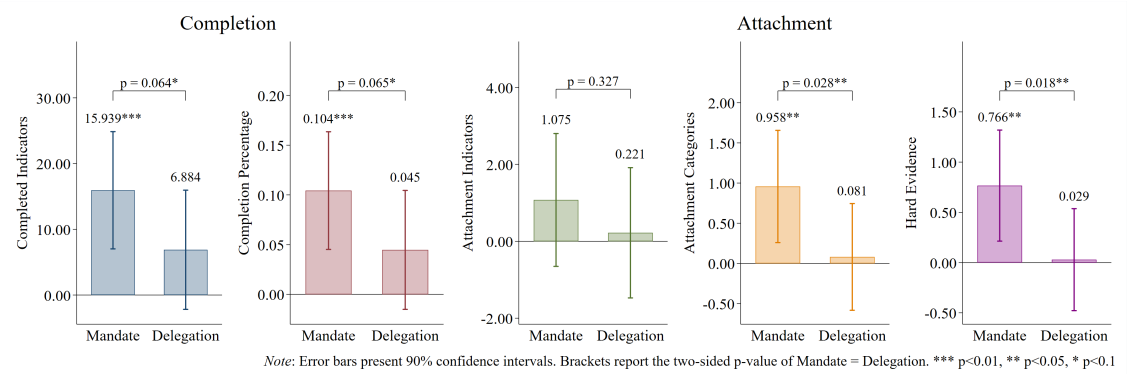


Figure 6: Authority Allocation and the Cost of KPI Choice

Notes: This figure reports the effects of mandated and delegated KPI selection on supplier engagement relative to the control group. Completeness is measured by the number and share of platform indicators the supplier completed. Documentation is measured by three breadth outcomes, namely indicators documented by at least one attachment, out of 75, attachment categories covered, out of 12, and categories backed by hard, verifiable evidence. Each regression includes reporting-round and randomization-stratum fixed effects, and standard errors are clustered by supplier. Points or bars show coefficient estimates and whiskers show 90 percent confidence intervals. The bracket in each panel reports the two-sided p-value of a test that the Mandate and Delegation coefficients are equal, with stars at the usual levels. The corresponding coefficient estimates, standard errors, constants, fixed effects, observations, and R^2 values are reported in Table F2.

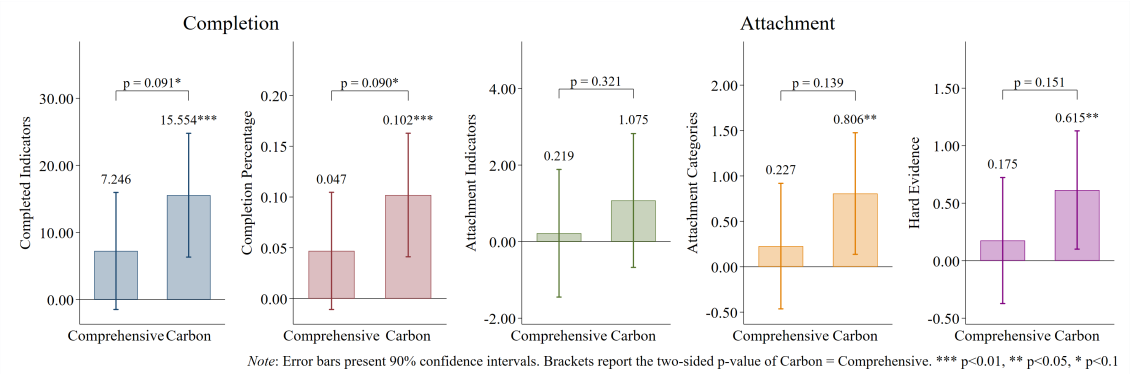


Figure 7: Monitoring Scope and the Value of Focus

Notes: This figure reports the effects of comprehensive and targeted carbon monitoring on supplier engagement relative to the control group. The outcomes cover completeness, measured by completed fields, and documentation, measured by the breadth and verifiability of the supporting material. The three counts are indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Each regression includes reporting-round and randomization-stratum fixed effects, and standard errors are clustered by supplier. Points or bars show coefficient estimates and whiskers show 90 percent confidence intervals. The bracket in each panel reports the two-sided p-value of a test that the Targeted Carbon and Comprehensive coefficients are equal, with stars at the usual levels. The corresponding regression statistics are reported in Table F3.

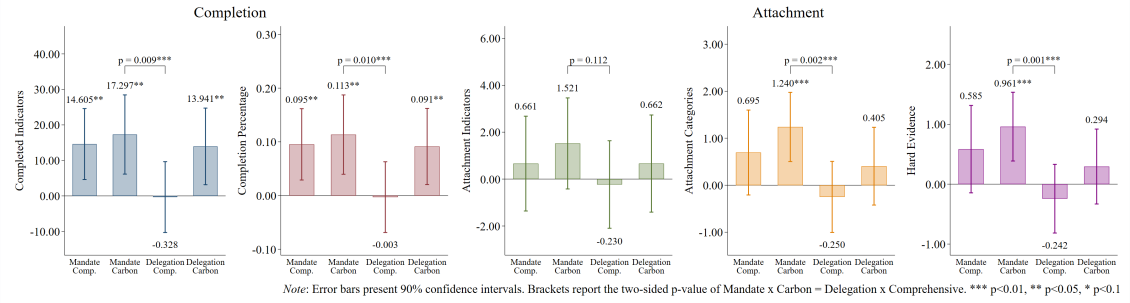
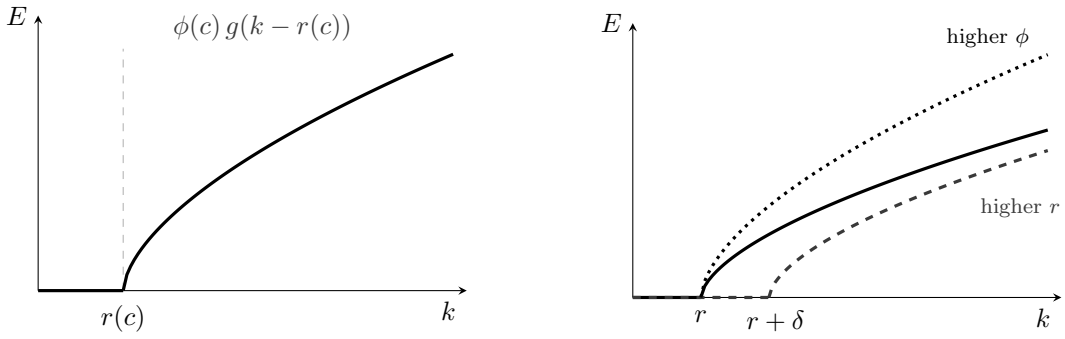


Figure 8: Engagement Across the Four Governance Bundles

Notes: This figure compares each of the four crossed treatment cells with the control group. The cells combine Mandate or Delegation with Comprehensive Monitoring or Targeted Carbon Monitoring. The outcomes are the number and share of completed indicators, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Regressions include reporting-round and randomization-stratum fixed effects with standard errors clustered by supplier. The mandate and targeted-monitoring cell produces the largest joint improvement across the two engagement margins. The bracket in each panel reports the two-sided p-value of a test of equality between the fully simplified and the fully complex cell, the statistic reported in the final row of Table F4. Table F4 reports the full set of regression statistics.



Panel A: Anatomy of the Curve *Panel B: Comparative Statics*

Figure 9: The Engagement Curve and Its Comparative Statics

Notes: This figure illustrates the engagement curve $E(k; c)$ defined in equation (7). The horizontal axis measures a supplier's organizational capacity k, and the vertical axis measures realized engagement. Panel A shows the two segments of the curve. Engagement is zero until capacity reaches the comprehension cost $r(c)$ given in equation (3), which is the capacity the supplier spends on processing the governance structure before any effort reaches the reporting program. Above that entry point, engagement rises in the capacity that remains, at a rate scaled by execution productivity $\phi(c)$ from equation (4), and the rise is concave. Panel B shows the two comparative statics. A higher comprehension cost shifts the curve rightward by the amount of the increase and leaves its shape unchanged, shown by the dashed curve. A higher execution productivity raises and steepens the curve and leaves the entry point unchanged, shown by the dotted curve. Lemma 1 states both properties. The curves are drawn for illustration and are not fitted to data.

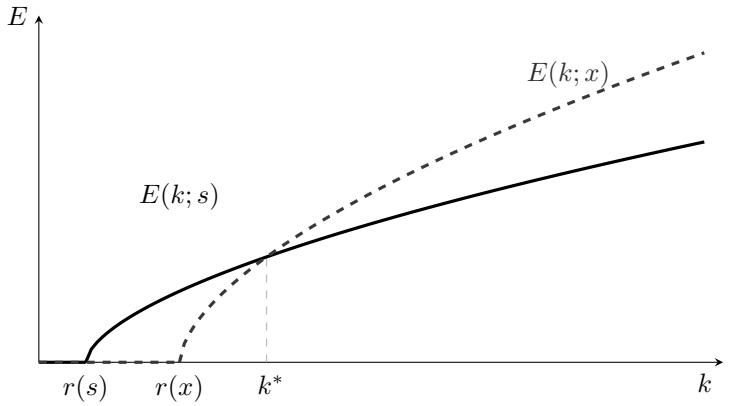


Figure 10: The Single Crossing in Supplier Capacity

Notes: This figure illustrates Proposition 2. It compares two governance structures that differ in complexity. The simple structure s, drawn as the solid curve, has the lower comprehension cost $r(s)$ and the lower execution productivity. The complex structure x, drawn as the dashed curve, has the higher comprehension cost $r(x)$ and the higher execution productivity. The complex structure starts later on the capacity axis and climbs faster once it starts, so the two curves meet at most once. In the case drawn they meet at k^*, and both are still rising there. Suppliers with capacity below k^* engage more under the simple structure, because its head start dominates. Suppliers with capacity above k^* engage more under the complex structure, because its steeper climb dominates. The threshold k^* rises with the complex structure's comprehension cost and falls with its execution productivity. The curves are drawn for illustration and are not fitted to data.

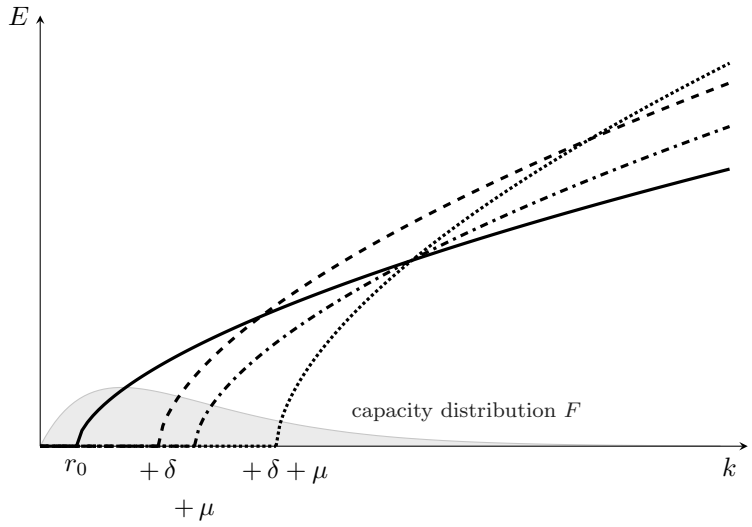


Figure 11: Four Governance Structures and a Constrained Capacity Distribution

Notes: This figure places the four governance structures of the experiment on a common capacity axis and superimposes the capacity distribution F, shown as the shaded density. Entry points order the four cells by total comprehension cost, following equation (3). The mandate and targeted cell (M, T) enters at r_0 and is drawn as the solid curve. The delegation and targeted cell (D, T) adds the choice overhead δ and is drawn as the dashed curve. The mandate and comprehensive cell (M, C) adds the scope overhead μ and is drawn as the dash-dotted curve. The fully complex cell (D, C) carries both overheads and is drawn as the densely dotted curve. The figure is drawn with $\delta < \mu$. Execution productivity from equation (4) orders the same four curves in the opposite direction, so each later entrant climbs faster. Expected engagement in a given experimental arm integrates the corresponding curve against the capacity density. When the network's capacity mass lies to the left of every pairwise crossing, the arm ranking follows the entry order, so the capacity-minimizing bundle (M, T) delivers the highest engagement and the fully complex bundle (D, C) the lowest. Proposition 3 states the result. The curves are drawn for illustration and are not fitted to data.

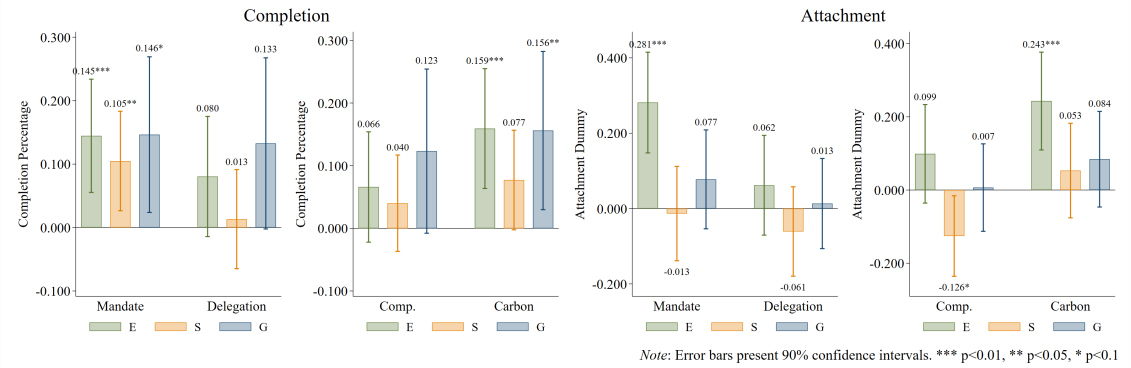


Figure 12: Engagement Responses Across ESG Dimensions

Notes: This figure separates treatment responses across environmental, social, and governance reporting. The completion panels measure the share of fields filled in within each domain. The attachment panels measure whether the supplier uploaded supporting documentation for the corresponding domain. Each coefficient compares a treatment arm with the control group in a regression containing reporting-round and randomization-stratum fixed effects. Standard errors are clustered by supplier, and whiskers show 90 percent confidence intervals. The estimates are largest and most precise for environmental indicators, which have clearer verification protocols and more accessible quantitative records. Table F8 provides the corresponding numerical estimates.

Table 1: Balance Table

	Control (C)	Mandate × Comp. (T1)	Mandate × Targeted (T2)	Delegation × Comp. (T3)	Delegation × Targeted (T4)
Firms	46	27	27	27	28
Panel A. Means, with p-values against control in brackets					
Production	31,232	29,820 [0.93]	111,478 [0.19]	32,663 [0.93]	30,453 [0.96]
Employment	111.1	171.0 [0.37]	220.3 [0.34]	231.2 [0.24]	381.5 [0.09]
Firm age (years)	15.52	15.82 [0.88]	12.81 [0.18]	14.85 [0.73]	16.75 [0.54]
Registered capital	4,417	3,436 [0.82]	2,934 [0.73]	1,813 [0.54]	3,497 [0.83]
ESG certificates (count)	1.00	0.56 [0.24]	1.16 [0.71]	0.92 [0.85]	0.85 [0.71]
CMT	0.33	0.33 [0.95]	0.38 [0.62]	0.33 [0.95]	0.36 [0.79]
ODM	0.11	0.11 [0.98]	0.12 [0.93]	0.11 [0.98]	0.14 [0.67]
FAB	0.33	0.26 [0.55]	0.31 [0.87]	0.26 [0.55]	0.21 [0.31]
ACC	0.24	0.30 [0.60]	0.19 [0.65]	0.30 [0.60]	0.29 [0.66]
S ranking	0.05	0.08 [0.60]	0.17 [0.11]	0.04 [0.88]	0.12 [0.33]
A ranking	0.30	0.25 [0.67]	0.26 [0.75]	0.33 [0.79]	0.27 [0.79]
B ranking	0.65	0.67 [0.89]	0.57 [0.51]	0.63 [0.84]	0.62 [0.78]
Large or medium size	0.16	0.15 [0.95]	0.15 [0.95]	0.22 [0.51]	0.21 [0.56]
Small size	0.55	0.58 [0.80]	0.62 [0.57]	0.56 [0.94]	0.57 [0.83]
Micro size	0.30	0.27 [0.82]	0.23 [0.56]	0.22 [0.51]	0.21 [0.45]
Panel B. The three right-skewed variables: medians, and means after an inverse hyperbolic sine transform					
Production, median	7,593	8,550	7,576	8,130	5,800
Production, IHS mean	9.75	9.80 [0.91]	9.70 [0.92]	9.90 [0.70]	9.60 [0.70]
Employment, median	21.5	59.0	26.0	31.5	37.0
Employment, IHS mean	3.74	4.33 [0.24]	3.95 [0.69]	3.92 [0.74]	4.40 [0.20]
Registered capital, median	599	500	1,000	500	900
Registered capital, IHS mean	7.00	7.02 [0.96]	7.60 [0.17]	6.97 [0.94]	7.44 [0.28]

Notes: Pre-treatment characteristics for the control group and for each cell of the 2×2 design, the level at which the factorial estimates of Section 4.3 are read. Brackets contain p-values from two-sided t-tests against the control mean. Production and registered capital are annual output value and registered capital in ten thousands of CNY, and employment is the registry-filed headcount summarized in Section 2.2. Roster items are missing for a few suppliers, so the observations behind a row range from 137 to 154. One of the sixty mean comparisons in Panel A, employment in cell T4, is significant at the ten percent level. Panel B guards against a single large factory moving a cell mean on a right-skewed variable without moving the distribution. No cell differs from control on the median or on the inverse hyperbolic sine mean of production, employment, or registered capital. Joint orthogonality tests point the same way. Regressing each cell indicator on the full set of Panel A characteristics against control returns p-values of 0.43, 0.68, 0.94, and 0.87 for T1 through T4, and a multinomial logit of arm assignment on the same characteristics returns $\chi^2(48) = 29.53$ with $p = 0.98$.

Appendix A. Complete List of Platform Indicators

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
1	Financial	6001	Financial Indicators	6001001	Monthly factory revenue*
1	Financial	6001	Financial Indicators	6001002	Annual factory output value*
1	Financial	6001	Financial Indicators	6001003	Monthly output value supplied to the Lead Firm*
1	Financial	6002	Operational Indicators	6002001	Overall on-time delivery rate*
1	Financial	6002	Operational Indicators	6002002	On-time delivery rate to the Lead Firm*
1	Financial	6002	Operational Indicators	6002003	Overall return rate*
1	Financial	6002	Operational Indicators	6002004	Return rate to the Lead Firm*
1	Financial	6003	Cooperation	6003001	Number of partner brands*
2	Environmental	1001	Production Information	1001001	Monthly output volume
2	Environmental	1001	Production Information	1001002	Monthly output for the Lead Firm**

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
2	Environmental	1002	Resource Consumption	1002001	Monthly purchased electricity**
2	Environmental	1002	Resource Consumption	1002002	Monthly purchased steam
2	Environmental	1002	Resource Consumption	1002003	Monthly diesel consumption
2	Environmental	1002	Resource Consumption	1002004	Monthly gasoline consumption
2	Environmental	1002	Resource Consumption	1002005	Monthly natural gas consumption
2	Environmental	1002	Resource Consumption	1002006	Monthly energy use by source†
2	Environmental	1002	Resource Consumption	1002007	Monthly coal combustion
2	Environmental	1002	Resource Consumption	1002008	Monthly natural gas combustion

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
2	Environmental	1002	Resource Consumption	1002009	Monthly oil combustion
2	Environmental	1002	Resource Consumption	1002010	Monthly biomass combustion
2	Environmental	1002	Resource Consumption	1002011	Monthly electric boiler use
2	Environmental	1003	Solid Waste Treatment	100301	Solid waste treatment†
2	Environmental	1004	Hazardous Waste Treatment	100701	Hazardous waste treatment†
2	Environmental	1005	Chemical Management	1004001	Chemical inventory list
2	Environmental	1005	Chemical Management	1004002	Chemical use records
2	Environmental	1006	Green Actions	1005001	Use of green electricity**
2	Environmental	1006	Green Actions	1005002	Share of green electricity used**
2	Environmental	1006	Green Actions	1005003	Monthly solar power use†

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
2	Environmental	1006	Green Actions	1005004	Monthly wind power use†
2	Environmental	1006	Green Actions	1005005	Monthly geothermal energy use†
2	Environmental	1006	Green Actions	1005006	Direct purchase of green electricity (e.g., PPA)†
2	Environmental	1006	Green Actions	1005007	Green certificates held (e.g., GEC)†
2	Environmental	1007	Water Resources	1006001	Rainwater recovery system in place**
2	Environmental	1007	Water Resources	1006002	Wet production process in use
2	Environmental	1007	Water Resources	1006003	Monthly wet production volume
2	Environmental	1007	Water Resources	1006004	Monthly total water use
2	Environmental	1007	Water Resources	1006005	Monthly production water use
2	Environmental	1007	Water Resources	1006006	Monthly recycled and reused water
2	Environmental	1007	Water Resources	1006007	Monthly domestic water use
2	Environmental	1007	Water Resources	1006008	Monthly wastewater discharge

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
2	Environmental	1007	Water Resources	1006009	Wastewater treatment method
3	Social	2001	Employee Composition	2001001	Total female employees
3	Social	2001	Employee Composition	2001002	Total male employees
3	Social	2001	Employee Composition	2001003	Female frontline workers
3	Social	2001	Employee Composition	2001004	Male frontline workers
3	Social	2001	Employee Composition	2001008	Total factory headcount (auto-calculated)†
3	Social	2001	Employee Composition	2001005	Female non-frontline employees
3	Social	2001	Employee Composition	2001007	Total non-frontline employees
3	Social	2002	Recruitment and Termination	2002001	Total terminated employees
3	Social	2002	Recruitment and Termination	2002002	Termination rate (auto-calculated)†

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2002	Recruitment and Termination	2002003	Total voluntary resignations
3	Social	2002	Recruitment and Termination	2002004	Resignation rate (auto-calculated)†
3	Social	2002	Recruitment and Termination	2002005	Female non-frontline employees terminated
3	Social	2002	Recruitment and Termination	2002006	Total non-frontline employees terminated
3	Social	2002	Recruitment and Termination	2002007	Total employees terminated
3	Social	2002	Recruitment and Termination	2002008	Female frontline employees resigning voluntarily
3	Social	2002	Recruitment and Termination	2002010	Total frontline resignations
3	Social	2002	Recruitment and Termination	2002011	Female non-frontline employees resigning voluntarily

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2002	Recruitment and Termination	2002013	Total non-frontline resignations
3	Social	2002	Recruitment and Termination	2002014	Total resignations
3	Social	2002	Recruitment and Termination	2002015	Resignation rate
3	Social	2003	Wages and Compensation	2003001	Local minimum wage standard†
3	Social	2003	Wages and Compensation	2003002	Factory minimum wage
3	Social	2003	Wages and Compensation	2003003	Average overtime hourly wage
3	Social	2003	Wages and Compensation	2003004	Average gross wage of full-attendance frontline workers

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2003	Wages and Compensation	2003005	Average gross wage of full-attendance female frontline workers (excl. overtime)
3	Social	2003	Wages and Compensation	2003006	Average gross wage of full-attendance male frontline workers (excl. overtime)†
3	Social	2003	Wages and Compensation	2003007	Average gross wage of full-attendance frontline workers (excl. overtime)
3	Social	2003	Wages and Compensation	2003008	Average gross wage of the lowest-paid 20 percent of full-attendance frontline workers (excl. overtime)
3	Social	2003	Wages and Compensation	2003009	Ratio of the lowest 20 percent average monthly wage to the all-employee average (excl. overtime)
3	Social	2004	Social Insurance	2004001	Employees covered by social insurance**†
3	Social	2004	Social Insurance	2004002	Social insurance coverage rate (auto-calculated)
3	Social	2004	Social Insurance	2004003	Employees covered by work injury insurance

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2004	Social Insurance	2004004	Work injury insurance coverage rate (auto-calculated)†
3	Social	2004	Social Insurance	2004005	Female frontline workers covered by work injury insurance
3	Social	2004	Social Insurance	2004006	Work injury insurance coverage rate, female frontline workers
3	Social	2004	Social Insurance	2004007	Work injury insurance coverage rate, frontline workers
3	Social	2004	Social Insurance	2004010	Female frontline workers covered by medical insurance
3	Social	2004	Social Insurance	2004011	Medical insurance coverage rate, female frontline workers
3	Social	2004	Social Insurance	2004012	Medical insurance coverage rate, frontline workers
3	Social	2004	Social Insurance	2004015	Female frontline workers covered by pension insurance
3	Social	2004	Social Insurance	2004016	Pension insurance coverage rate, female frontline workers
3	Social	2004	Social Insurance	2004017	Pension insurance coverage rate, frontline workers

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2004	Social Insurance	2004020	Female frontline workers covered by unemployment insurance
3	Social	2004	Social Insurance	2004021	Unemployment insurance coverage rate, female frontline workers
3	Social	2004	Social Insurance	2004022	Unemployment insurance coverage rate, frontline workers
3	Social	2001	Working Hours	2001009	Actual factory working days
3	Social	2001	Working Hours	2001010	Statutory working days in the quarter (auto-filled)
3	Social	2005	Working Hours	2005001	Absence days of male frontline workers†
3	Social	2005	Working Hours	2005002	Absence days of female frontline workers
3	Social	2005	Working Hours	2005003	Daily working hours of frontline workers
3	Social	2006	Employee Representation	2006001	Democratically elected female worker representatives

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2006	Employee Representation	2006002	Democratically elected male worker representatives†
3	Social	2006	Employee Representation	2006003	Democratically elected female union representatives
3	Social	2006	Employee Representation	2006004	Democratically elected male union representatives†
3	Social	2007	Employee Grievances	2007001	Grievances / complaints received
3	Social	2007	Employee Grievances	2007002	Grievances / complaints resolved
3	Social	2008	Labor Contract Management	2008001	Female workers with signed labor contracts
3	Social	2008	Labor Contract Management	2008002	Male workers with signed labor contracts†
3	Social	2008	Labor Contract Management	2008003	Labor contract coverage rate (auto-calculated)†
3	Social	4001	Community Impact	4001001	Participation in charitable or public-interest activity

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
4	Governance	5001	Governance Structure	5001001	Dedicated ESG department or position in place**
4	Governance	5001	Governance Structure	5001002	Formal ESG management policy in place
4	Governance	5002	Compliance	5002001	Anti-corruption / compliance policy in place
4	Governance	5003	Business Ethics	5003001	Incidents of commercial bribery, fraud, or other improper conduct
4	Governance	5003	Business Ethics	5003002	Penalties imposed by national regulators**
5	Certification	3001	Environmental	3001001	ISO 14001 (Environmental Management System)*
5	Certification	3001	Environmental	3001002	LCMP (Low Carbon Manufacturing Programme)*
5	Certification	3001	Environmental	3001003	HIGG FEM (Environmental Performance Verification)*
5	Certification	3001	Environmental	3001004	Other environmental certifications*
5	Certification	3002	Social Responsibility	3002001	SA8000 (Social Responsibility Audit)*
5	Certification	3002	Social Responsibility	3002002	SMETA (Social Responsibility Audit)*

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
5	Certification	3002	Social Responsibility	3002003	WCA (Workplace Conditions Assessment)*
5	Certification	3002	Social Responsibility	3002004	SLCP (Social Responsibility Verification)*
5	Certification	3002	Social Responsibility	3002005	BSCI (Social Responsibility Verification)*
5	Certification	3002	Social Responsibility	3002006	ERSA (Social Responsibility Verification)*
5	Certification	3002	Social Responsibility	3002007	Better Work (Labor Rights Certification)*
5	Certification	3002	Social Responsibility	3002008	WRAP (Responsible Production Certification)*
5	Certification	3002	Social Responsibility	3002009	Disney ILS (Labor Standards Certification)*
5	Certification	3002	Social Responsibility	3002010	EcoVadis (Social Responsibility Assessment)*
5	Certification	3002	Social Responsibility	3002011	Other social certifications*
5	Certification	3003	Material Certification	3003001	RDS (SC) (Responsible Down Standard Scope Certificate)*
5	Certification	3003	Material Certification	3003002	RDS (TC) (Responsible Down Standard Transaction Certificate)*

Continued on next page

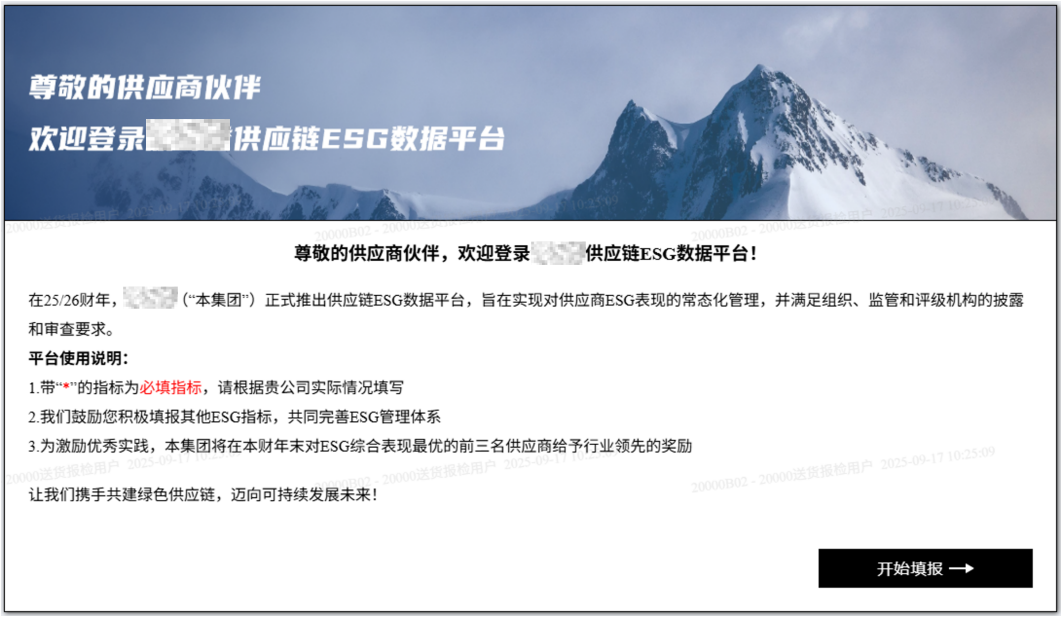
Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
5	Certification	3003	Material Certification	3003003	GRS (SC) (Global Recycled Standard Scope Certificate)*
5	Certification	3003	Material Certification	3003004	GRS (TC) (Global Recycled Standard Transaction Certificate)*
5	Certification	3003	Material Certification	3003005	OEKO-TEX (Ecological Textile Standard)*
5	Certification	3003	Material Certification	3003006	Bluesign certification*
5	Certification	3003	Material Certification	3003007	FSC (Forest Certification)*

Notes: A single asterisk marks the fields required of every supplier under the reporting program, the 8 financial and 22 certification indicators that form the common baseline. A double asterisk marks the eight indicators the buyer designated for suppliers in the Mandate arm, while suppliers in the Delegation arm chose eight of their own. The 78 environmental, social, and governance indicators are the set over which suppliers make reporting decisions. A dagger marks 20 further fields the platform carries that the reporting requirement does not treat as separate items. Five are components of the green-electricity share, which enters through the share itself, and three are heading rows that open a reporting block. The remainder are five male-specific counterparts of items the platform records for female employees, five automatically calculated rates and totals, and two fields the platform supplies from other records. The table therefore lists 128 platform fields, of which the 108 without a dagger are the applicable set of Section efsubsec:platform, summarized by domain in Table eftab:indicators.

Appendix B. Platform Interface Screenshots

This appendix displays screenshots and English translations of the group-specific welcome pages from the lead firm's ESG data platform. Each screen reflects one combination of the two experimental variations, the degree of autonomy in selecting ESG key performance indicators and the scope of third-party auditing. All messages encourage suppliers to provide comprehensive information to enhance supply chain management. For confidentiality reasons, the identity of the lead firm has been anonymized.

Panel A. Interface 1: Baseline

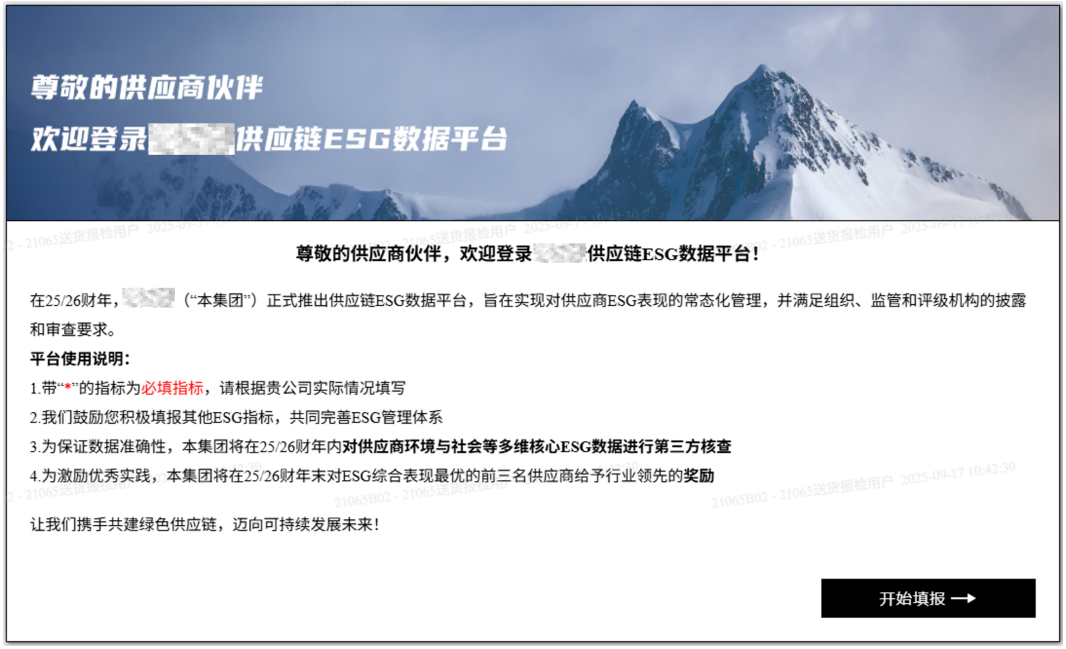


(Original screenshot)



(English translation)

Panel B. Interface 2: Mandated KPIs, Comprehensive Audit

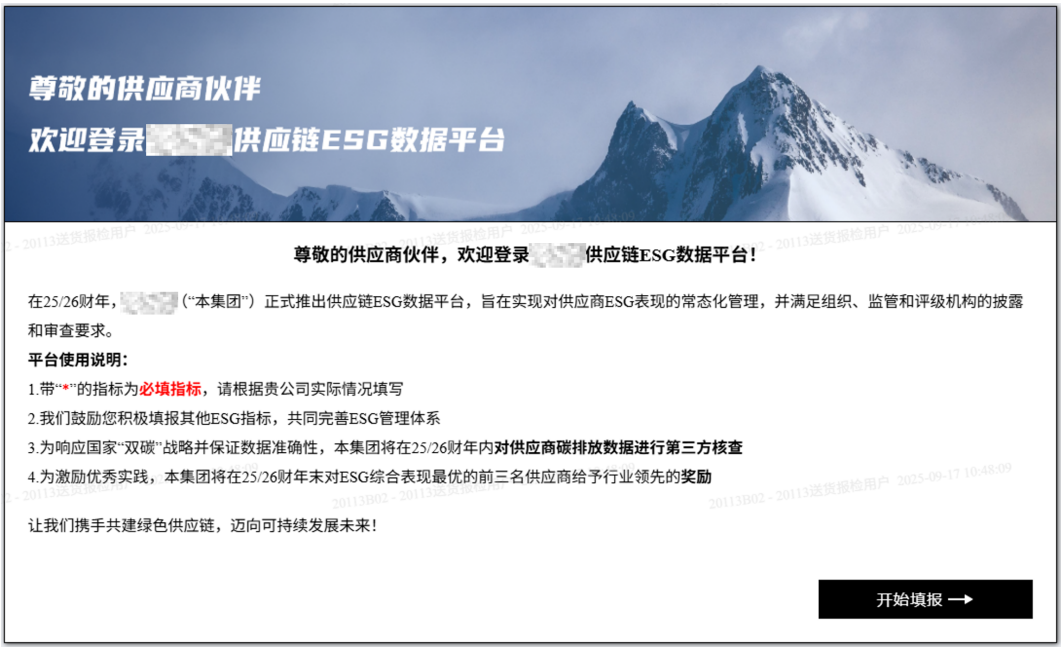


(Original screenshot)



(English translation)

Panel C. Interface 3: Mandated KPIs, Carbon Audit

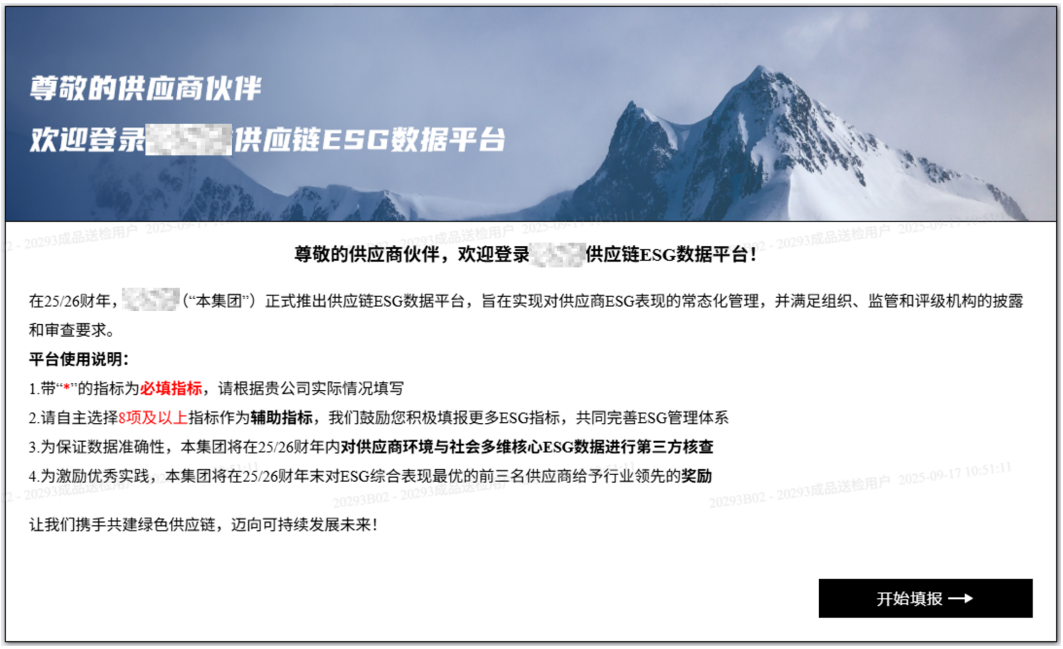


(Original screenshot)

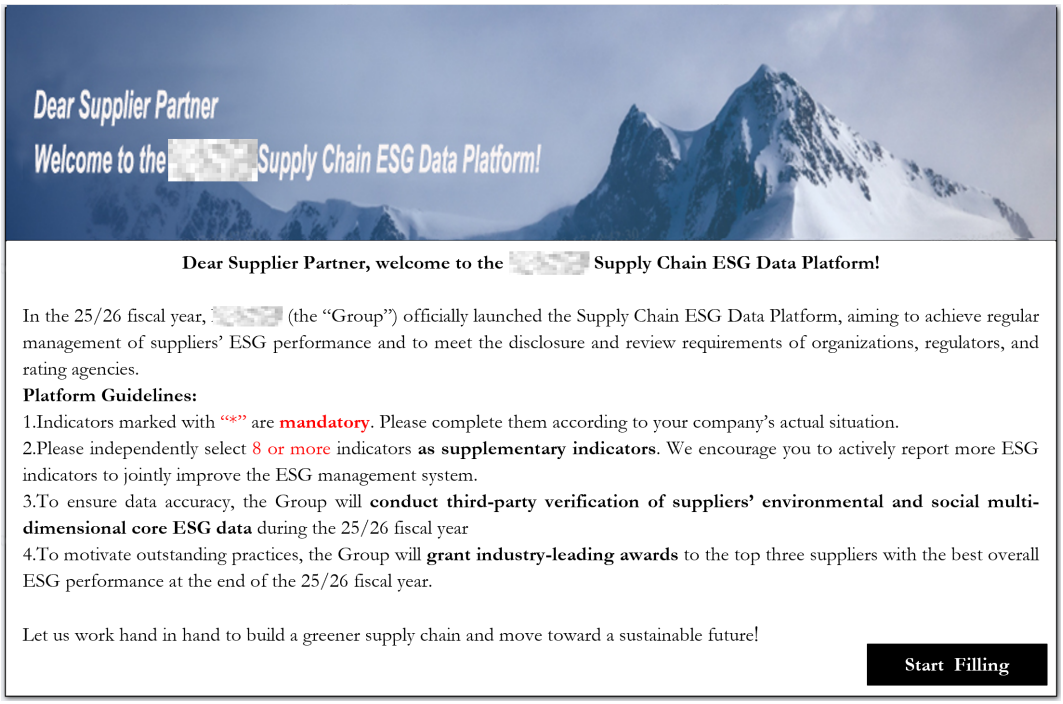


(English translation)

Panel D. Interface 4: Delegated KPIs, Comprehensive Audit

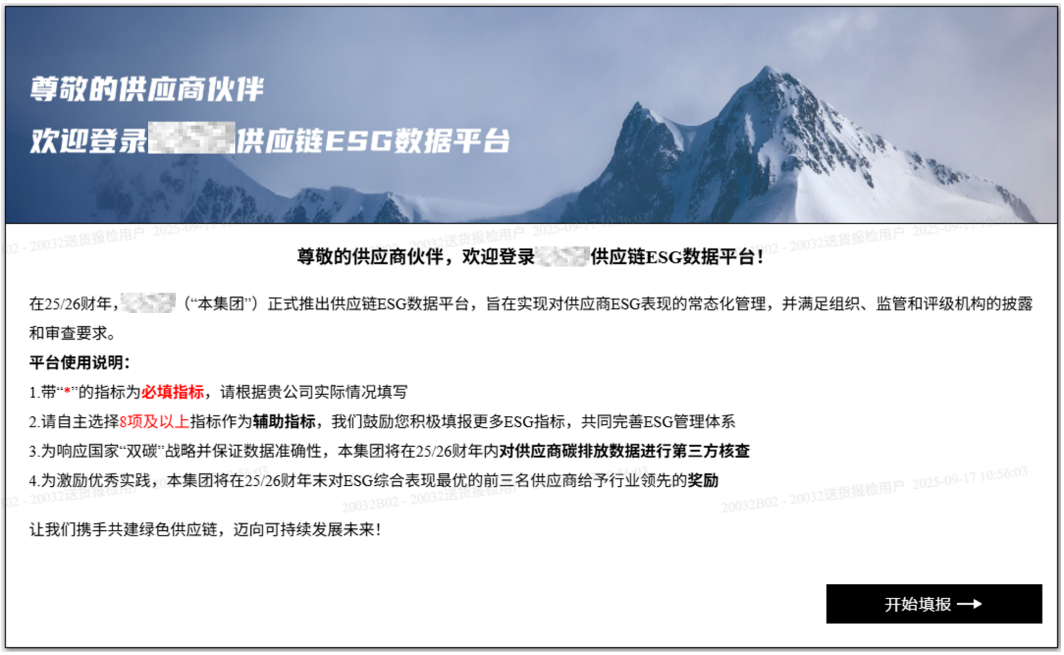


(Original screenshot)



(English translation)

Panel E. Interface 5: Delegated KPIs, Carbon Audit



(Original screenshot)



(English translation)

Appendix C. Processing ESG Documentation Files

This appendix describes the semi-automated procedure used to convert unstructured supplier-uploaded ESG documentation into the three attachment outcomes used in the empirical analysis: *Indicators*, *Categories*, and *Hard evidence*.

The natural starting point is the volume of material a supplier uploads, but volume is a noisy measure of engagement. A single reporting item such as monthly electricity consumption can be documented with one meter reading or with a dozen photographs of the same meter, so a file count rises with filing habits as much as with the information provided. The procedure therefore identifies which reporting item each file documents, and the analysis works with counts of items. *Indicators* is the number of distinct reporting items a supplier documented. Those items are grouped into the twelve ESG topics of the classification taxonomy, which gives *Categories*, the number of topics a supplier covered. Each item is also classified by evidence type, which gives *Hard evidence*, the number of topics supported by quantitative or directly verifiable material, and allows us to ask how much of the documentation gain is verifiable. *Soft evidence* is constructed in parallel on the same definition.

The workflow proceeds in four stages, one supplier and one reporting round at a time. An inventory of every attachment is assembled and each file is converted into a machine-readable form. A large language model (LLM) then reads each file, identifies the evidence items it contains, and assigns each item to the most specific reporting item on the platform, to one of the twelve ESG topics, and to an evidence type. Every machine-generated classification is reviewed against the source file. Duplicate and out-of-scope records are then removed and the survivors are aggregated to the supplier-round level.

C.1. File Parsing and Standardization

Supplier attachments arrive in multiple formats, including PDF files, Word documents, Excel spreadsheets, compressed archives, scanned images, and photographs. The first step is an inventory of every attachment in a supplier's folder for the round, including files held in the platform's duplicates folder and files contained inside compressed archives, so that no upload is overlooked. An archive is unpacked and each component file is reviewed separately, which means a supplier who submits several documents in one archive is treated the same as one who submits them individually.

Each file is then transformed into a standardized machine-readable representation. For text-based files such as PDFs and Word documents, text is extracted directly using format-specific parsing tools. Spreadsheets are read sheet by sheet, including embedded images, and converted into structured text that preserves the row-column relationships needed for interpretation. For scanned files and photographs, the visual content is passed to a multimodal model through encoded image inputs, allowing direct optical recognition of the underlying evidence. This stage standardizes otherwise heterogeneous inputs while preserving the substantive content relevant for classification.

C.2. LLM-Assisted Classification

After parsing, each file is submitted to an LLM-based classifier. The classifier first identifies the evidence items the file contains, such as individual invoices, monthly records, certificates, test results, or policy documents. It then assigns each item to the most specific reporting item on the platform, to the ESG topic that item belongs to, and to an evidence type, either `hard` or `soft`. Items in the same file that support the same reporting item are grouped into one record, and a file that supports several reporting items generates one record for each. A water-fee workbook covering both tap-water charges and wastewater charges, for example, produces two records.

The classification taxonomy is closed and fixed ex ante. The admissible topics are Production, Electricity, Resource Consumption, Solid and Hazardous Waste, Chemicals, Green Initiatives, Water Resources, Employees, Certificates, Community Impact, Corporate Governance, and Other. Labels are assigned from the readable content of a file, and a topic is admitted only when the file directly supports it. Topics that are merely mentioned, implied, or conventionally associated with a document type are not added.

Two families of documents require more specific rules. Invoices are classified by the item billed, so that electricity charges, photovoltaic generation records, and water and wastewater charges map to separate reporting items, and mixed invoices and summary workbooks are split accordingly. Certificates are mapped to the named program alone, such as SLCP, SMETA, ISO 14001, GRS, or RDS.

The evidence-type assignment follows. *Hard* evidence refers to quantitative and directly verifiable material, such as invoices, metering records, production and resource-consumption tables, inventories, chemical-use records, and numerical test results. *Soft* evidence refers to qualitative or descriptive material, such as policy documents, manuals, contracts, reports without quantitative records, and photographs documenting activities or practices. Evidence type does not change the reporting item or the topic a record is assigned to. It determines only whether that topic enters *Hard evidence*, so a topic can appear in *Categories* while being absent from *Hard evidence* when it is supported by descriptive material alone.

C.3. Human Validation and Quality Control

The procedure is intentionally semi-automated. All machine-generated classifications are reviewed before aggregation. Each source file is reopened and compared with the records derived from it, and discrepancies are corrected.

This review serves two purposes. First, it corrects extraction errors arising from low-quality scans, formatting irregularities, or incomplete optical recognition. Second, it enforces the coding rules in borderline cases involving overlapping topics or mixed qualitative and

quantitative content.

Two filters are applied at the same stage. Duplicates are identified by comparing file hashes. A file submitted more than once within a round is counted once, while a file resubmitted in a later round is retained and flagged, so that repeated submission is not read as new documentation. Records are dropped when the evidence falls outside the reporting period or when the file is unrelated to ESG reporting. This hybrid design preserves the scalability of automated coding while maintaining conservative measurement standards for the final analysis dataset.

C.4. Construction of Supplier-Level Variables

Let R_{st} denote the records that survive validation and filtering for supplier s in reporting round t. Each record r carries a reporting item $i(r)$, an ESG topic $c(r)$, and an evidence type. The attachment outcomes are

$$\begin{aligned}
\textit{Indicators}_{st} &= \big|\{\, i(r) : r \in R_{st} \,\}\big|, \\
\textit{Categories}_{st} &= \big|\{\, c(r) : r \in R_{st} \,\}\big|, \\
\textit{Hard evidence}_{st} &= \big|\{\, c(r) : r \in R_{st},\ r \text{ hard} \,\}\big|, \\
\textit{Soft evidence}_{st} &= \big|\{\, c(r) : r \in R_{st},\ r \text{ soft} \,\}\big|.
\end{aligned}$$

Three counting rules follow from these definitions. A reporting item is counted once for a supplier-round however many files support it, so repeated uploads within an item do not raise *Indicators*. A file that documents several reporting items raises *Indicators* by more than one, and several items within the same topic raise *Indicators* separately while raising *Categories* only once. A topic enters *Hard evidence* as soon as one retained record supporting it is coded hard, and further hard files in the same topic add nothing. All three measures are therefore breadth measures. The panel retains only reporting items for which some supplier provided surviving evidence, and a supplier-round with no surviving evidence

takes the value zero on all three measures and remains in the estimation sample.

The empirical analysis uses *Indicators*, *Categories*, and *Hard evidence*, which capture successively demanding margins: the reach of a supplier's documentation, the breadth of ESG topics it spans, and the number of those topics backed by verifiable material. A supplier can raise reach without raising breadth by documenting several items within one topic, and can raise breadth without raising verifiability by covering a topic with descriptive material alone. *Soft evidence* is the analogous count for qualitative material. We construct it on the same definition and do not use it in the body, where the documentation results rest on the hard-evidence margin. The procedure therefore turns evidence submitted in heterogeneous formats into comparable measures of what a supplier documented, which is the response the reporting program is designed to elicit.

C.5. Illustrative Example

Table C1 provides a de-identified example of the record-level coding and the supplier-round aggregation. Most files support a single reporting item. The spreadsheet in row 8 contains several distinct types of ESG information and therefore supports five. *Indicators* exceeds *Categories* for this supplier because three of the thirteen items it documented fall into a topic another file already covers, namely resource consumption, water resources, and solid and hazardous waste, so thirteen items span ten topics.

Table C1: Illustrative Mapping from ESG Evidence Files to Supplier-Level Outcomes

Index	File	Assigned Reporting Item	Assigned Category	Evidence Type
1	Chemical usage screenshot.png	Chemical usage record	Chemicals	Hard
2	2025 SLCP report.pdf	SLCP	Certificates	Soft
3	Electricity report, Oct. 31, 2025.pdf	Monthly grid electricity purchase	Electricity	Hard
4	Tap water data, Oct. 2025.pdf	Monthly water consumption	Water Resources	Hard
5	Steam data, Oct. 2025.pdf	Monthly steam purchase	Resource Consumption	Hard
6	Sludge disposal agreement, 2025.pdf	Solid-waste outsourcing contract	Solid and Hazardous Waste	Soft
7	Code of ethics manual.pdf	Anti-corruption and compliance policy	Corporate Governance	Soft
8	ESH statistics (including green electricity).xls	Solid-waste treatment record	Solid and Hazardous Waste	Hard
8	ESH statistics (including green electricity).xls	Monthly fuel consumption	Resource Consumption	Hard
8	ESH statistics (including green electricity).xls	Monthly wastewater discharge	Water Resources	Hard
8	ESH statistics (including green electricity).xls	Use of green electricity	Green Initiatives	Hard
8	ESH statistics (including green electricity).xls	Monthly production volume	Production	Hard
9	Charity fund agreement.pdf	Charitable or public-interest activity	Community Impact	Soft
Supplier-round aggregation			*Indicators*	13
			Categories	10
			Hard evidence	7
			Soft evidence	4

Notes. This table provides a de-identified example of the coding procedure. File names are translated and simplified for exposition. Nine uploaded files yield thirteen records, because the spreadsheet in row 8 documents production, fuel use, water, green electricity, and waste separately. *Indicators* counts distinct reporting items, *Categories* counts distinct ESG topics, and *Hard evidence* and *Soft evidence* count the topics supported by hard and by soft evidence. Repeated uploads within one reporting item or one topic do not raise the corresponding count.

Appendix D. Formal Results and Proofs

This appendix proves the results stated in Section 5 and develops the two microfoundations the body refers to: the delegated-selection subgame behind Lemma 2, and the supplier's decision to process the contract.

D.1 Proofs for the Baseline Model

Proof of Lemma 1. For $k < r(c)$ the feasible set in (5) is empty and engagement is zero. For $k \geq r(c)$ the objective is strictly concave in e, with derivative $\alpha\, w(c)\, \phi(c)\, g'(e) - \rho$. That derivative is strictly positive on $[0,\, k - r(c)]$: Assumption 2 makes it positive at the upper end, and $g'' < 0$ makes $g'(e) \geq g'(k - r(c))$ everywhere below it. The optimum is therefore the upper bound $e = k - r(c)$, which gives (7). Continuity at $k = r(c)$ follows from $g(0) = 0$, and strict monotonicity and strict concavity above $r(c)$ follow from $g' > 0$ and $g'' < 0$ in Assumption 1. For the comparative statics, $r(c)$ enters only through the difference $k - r(c)$, so an increase in $r(c)$ translates the curve rightward by that amount. The factor $\phi(c)$ multiplies both the level and the slope, $\partial E / \partial k = \phi(c)\, g'(k - r(c))$, and leaves the entry point $r(c)$ unchanged. □

Proof of Proposition 1. With $\delta = \mu = 0$, (3) gives $r(c) = r_0$ for every structure, so $E(k; c) = \phi(c)\, g(k - r_0)$ and the four curves differ only by the factor $\phi(c)$. By (4), $\phi(D, C) \geq \phi(c)$ for every c, with strict inequality whenever $\pi_D + \pi_C > 0$, so the engagement ranking holds without any restriction on w. By Assumption 3, $w(D, C) > w(c)$ for every $c \neq (D, C)$, so the value ranking follows from multiplying a weakly larger engagement level by a strictly larger weight. The population statement follows by integrating a pointwise ranking against any F. □

Proof of Proposition 2. (i) On $[\, r(x),\, \bar{k}\,]$ both curves are given by (7), so $\Delta_{xs}(k) = \phi(x)\, g(k -$

$r(x)) - \phi(s)\, g(k - r(s))$, with derivative

$$\Delta'_{xs}(k) \;=\; \phi(x)\, g'\big(k - r(x)\big) \;-\; \phi(s)\, g'\big(k - r(s)\big) \;>\; 0.$$

The inequality holds because $k - r(x) < k - r(s)$ and $g'' < 0$ imply $g'(k - r(x)) > g'(k - r(s))$, and $\phi(x) \geq \phi(s)$. At $k = r(x)$, $\Delta_{xs}(r(x)) = -\phi(s)\, g(r(x) - r(s)) < 0$. A strictly increasing function has at most one zero, and it crosses from below.

(ii) If $\phi(x) = \phi(s)$ then $\Delta_{xs}(k) = \phi(s)\,[\,g(k - r(x)) - g(k - r(s))\,] < 0$ at every k, so no crossing exists. If $\phi(x) > \phi(s)$ and $\Delta_{xs}(\bar{k}) > 0$, continuity and strict monotonicity give a unique zero $k^* \in (r(x), \bar{k})$. Both structures are climbing throughout $[\,r(x),\, \bar{k}\,]$ by Assumption 2, so the crossing occurs while both curves are rising. Differentiating $\phi(x)\, g(k^* - r(x)) = \phi(s)\, g(k^* - r(s))$ gives

$$\frac{\partial k^*}{\partial r(x)} \;=\; \frac{\phi(x)\, g'(k^* - r(x))}{\Delta'_{xs}(k^*)} \;>\; 0, \qquad \frac{\partial k^*}{\partial \phi(x)} \;=\; -\,\frac{g(k^* - r(x))}{\Delta'_{xs}(k^*)} \;<\; 0.$$

(iii) Define $\varphi(k) = w(x)\, E(k; x) - w(s)\, E(k; s)$. Then $\varphi(r(x)) < 0$, and $\varphi(k^*) = (w(x) - w(s))\, E(k^*; s) > 0$ because the two engagement levels coincide at k^* and $w(x) > w(s)$ by Assumption 3. Let $k^*_\Pi = \inf\{k : \varphi(k) \geq 0\} \leq k^*$. Continuity gives $\varphi < 0$ on $(r(s),\, k^*_\Pi)$, and for $k > k^*$ both $E(k; x) > E(k; s)$ and $w(x) > w(s)$ give $\varphi(k) > 0$. □

Proof of Proposition 3. (i) Take any two of the four structures. If the one with the lower comprehension cost also has the weakly higher execution productivity, it delivers strictly greater engagement at every $k > r(M, T)$, and the comparison holds on the whole support. Otherwise the pair satisfies the hypotheses of Proposition 2, and below its crossing the structure with the lower comprehension cost delivers strictly greater engagement. Suppliers with $k \leq r(c)$ contribute zero under c and preserve the comparison weakly. Fix $k \in (r(M, T),\, \underline{k}^*)$: every pairwise comparison then ranks the structures by total overhead. The interval $(r(M, T),\, \underline{k}^*)$ carries positive mass because F has strictly positive density on $(0, \bar{k})$, so each compared pair is strictly separated on a set of positive F-measure. Integrating

against F with $F(\underline{k}^*) = 1$ yields the stated strict rankings.

(ii) The same argument applied to φ from Proposition 2(iii), using $F(\underline{k}^*_\Pi) = 1$, ranks $\Pi(c)$ identically, and the buyer chooses the highest-ranked structure. □

D.2 The Delegated-Selection Subgame

Under delegation the supplier chooses an indicator set before allocating capacity to execution. Two sets are available. The informative set matches indicators to the buyer's information needs and carries per-unit value $\bar{w}$. The convenient set minimizes the supplier's own selection effort and carries the mandated value $\underline{w} < \bar{w}$. Selecting the informative set requires additional evaluation effort $\tau > 0$, whose capacity value is $\rho\,\tau$.

Proof of Lemma 2 and Corollary 1. Given a selection with per-unit value w, the supplier's continuation payoff at residual capacity $e = k - r(D, H)$ is $\alpha\, w\, \phi(D, H)\, g(e)$, so the gain from informative selection is

$$G(\alpha, e) \;=\; \alpha \left[\, \bar{w} - \underline{w} \,\right] \phi(D, H)\, g(e),$$

and the supplier selects the informative set whenever $G(\alpha, e) \geq \rho\,\tau$. The function G is continuous, strictly increasing in α, equal to zero at $\alpha = 0$, and strictly increasing in e and hence in k. Let $\underline{e}$ denote residual capacity at the lowest capacity level in the support that processes the contract, and define $\underline{\alpha}$ as the value of α solving $G(\alpha, \underline{e}) = \rho\,\tau$, which lies in $(0, 1)$ under the maintained condition $G(1, \underline{e}) > \rho\,\tau$. For $\alpha \geq \underline{\alpha}$ the inequality holds at every capacity level in the support, so selection is informative and $w(D, H) = \bar{w} > \underline{w} = w(M, H)$. For $\alpha < \underline{\alpha}$ it fails at $\underline{e}$, the convenient set is selected, and $w(D, H) = w(M, H)$, which is Corollary 1. □

The productivity gain π_D in (4) does not depend on this comparison. A supplier that selects for convenience still selects indicators its own records support, so it still converts effort into reported content at the higher rate.

D.3 The Processing Decision

The baseline treats processing as part of participation. Suppose the supplier may decline to process the contract, retaining the payoff ρk. Processing is chosen when $\alpha\, w(c)\, \phi(c)\, g(e) \geq \rho\,[\, r(c) + e\,]$ at $e = k - r(c)$. The left side is zero at $k = r(c)$ while the right side is positive, and the difference is strictly increasing in k because $\alpha\, w(c)\, \phi(c)\, g'(e) > \rho$ by Assumption 2. Processing therefore occurs above a unique threshold $k^e(c) > r(c)$ whenever the inequality holds at $\bar{k}$, and $k^e(c)$ is increasing in $r(c)$ and decreasing in $\alpha\, w(c)\, \phi(c)$. Engagement is zero on $[\,0,\, k^e(c))$ and follows (7) above it. Entry points shift rightward, and further so for structures with larger $r(c)$, so the single crossing of Proposition 2 and the population ranking of Proposition 3 obtain with the thresholds evaluated at $k^e(c)$ in place of $r(c)$.

D.4 Proofs for the Extensions

Proof of Proposition 4. (i) Fix $k > r(D, C)$ and write $y = k - r_0$ for residual capacity under the capacity-minimizing bundle. Residual capacity under the other three structures is $y - \delta$ under (D, T), $y - \mu$ under (M, C), and $y - \delta - \mu - \gamma$ under (D, C), all strictly positive. With ϕ as in (4), (8) becomes

$$I(k) \;=\; g(y) \;+\; (1 + \pi_D + \pi_C)\, g(y - \delta - \mu - \gamma) \;-\; (1 + \pi_C)\, g(y - \mu) \;-\; (1 + \pi_D)\, g(y - \delta),$$

which decomposes as

$$\begin{aligned} I(k) \;=\; & \underbrace{\big[\, g(y) + g(y - \delta - \mu - \gamma) - g(y - \delta) - g(y - \mu) \,\big]}_{\text{(a)}} \\ & + \underbrace{\pi_D \big[\, g(y - \delta - \mu - \gamma) - g(y - \delta) \,\big]}_{\text{(b)}} \\ & + \underbrace{\pi_C \big[\, g(y - \delta - \mu - \gamma) - g(y - \mu) \,\big]}_{\text{(c)}}. \end{aligned}$$

Term (a) is negative. At $\gamma = 0$, integrating g'' over the rectangle $[0, \delta] \times [0, \mu]$ gives

$$\int_0^{\delta} \int_0^{\mu} g''\big(y - a - b\big)\, db\, da \;=\; g(y) + g(y - \delta - \mu) - g(y - \delta) - g(y - \mu) \;<\; 0$$

by the fundamental theorem of calculus and $g'' < 0$. Any $\gamma > 0$ lowers $g(y - \delta - \mu - \gamma)$ further and so lowers (a). Terms (b) and (c) are nonpositive because g is increasing and the argument in each bracket is smaller in the first term than in the second, and they are strictly negative under Assumption 4. Hence $I(k) < 0$.

(ii) The interaction γ enters only through the (D, C) term, and $\partial I / \partial \gamma = -(1 + \pi_D + \pi_C)\, g'(y - \delta - \mu - \gamma) < 0$.

(iii) Rearranging (8),

$$\big[\, E(k; M, C) - E(k; D, C) \,\big] + \big[\, E(k; D, T) - E(k; D, C) \,\big] - \big[\, E(k; M, T) - E(k; D, C) \,\big] \;=\; -\, I(k),$$

so the two single-relief gains sum to more than the joint gain exactly when $I(k) < 0$.

(iv) Integrate part (i) against any F supported on $(\, r(D, C),\, \bar{k}\,]$, which gives $\mathbb{E}_F\, I(k) < 0$. □

Part (i) holds at $\gamma = 0$, so compounding does not require the two clauses to interact in comprehension. Part (iv) restricts the support because a supplier with $k \le r(D, C)$ produces zero engagement under the fully complex structure, which removes the term that makes I negative. Proposition 3 and Proposition 4 therefore hold together on populations supported on $(\, r(D, C),\, \underline{k}^*\,)$, an interval that is non-empty when the fully complex structure is feasible below the smallest pairwise crossing, $r(D, C) < \underline{k}^*$.

Proof of Proposition 5. The supplier allocates residual capacity $e = k - r(c)$ across the N

tasks and the core business:

$$\max_{\{s_j\}\geq 0} \sum_{j=1}^{N} \omega_j \, \psi(\lambda_j s_j) \;+\; \rho\Big[\, e - \sum_{j=1}^{N} s_j \,\Big], \qquad \omega_j \;=\; \alpha \, w_j(c) \, \phi(c),$$

subject to the capacity term in brackets remaining nonnegative, with multiplier $\chi \geq \rho$, where $w_j(c)$ is the per-unit value of content on task j. The first-order condition is $\omega_j \, \lambda_j \, \psi'(\lambda_j s_j) \leq \chi$, with equality when $s_j > 0$. Since $\psi'(0) = \bar{\psi} < \infty$, task j is undertaken if and only if $\omega_j \lambda_j \bar{\psi} > \chi$. With the value of content equal across tasks, $\omega_j = \omega$ for every j, so setting $\eta_j = \omega \, \lambda_j \, \bar{\psi}$ gives thresholds strictly ordered by tractability, $\eta_1 > \eta_2 > \cdots > \eta_N$. The multiplier $\chi(k; c)$ is continuous and strictly decreasing in $e = k - r(c)$ while the constraint binds, so tasks activate in the order $\eta_1 > \eta_2 > \cdots$ as e grows. For an active task, implicit differentiation of the first-order condition gives

$$\frac{\partial s_j}{\partial \chi} \;=\; \big[\, \omega \, \lambda_j^2 \, \psi''(\lambda_j s_j) \,\big]^{-1} \;<\; 0.$$

A governance change that lowers $r(c)$ raises e one for one, lowers χ, raises s_j on every active task, and brings the most tractable inactive task above its threshold. □

The ordering of treatment effects across tasks comes from the activation margin. A task below its threshold responds not at all, and the tasks below threshold are the least tractable ones, so a governance change that frees capacity registers on the tractable dimensions and fades as task complexity rises. The size of the response on tasks already active depends on the curvature of ψ at the prevailing level of content and is not ordered by tractability without further restrictions. When the value of content differs across tasks, the thresholds are $\eta_j = \omega_j \lambda_j \bar{\psi}$ and the activation order follows the product $w_j(c) \, \lambda_j$, so a structure that raises the value of content on one task raises that task's threshold.

Proof of Proposition 6. Part (i) is Proposition 2(i). For part (ii), take any F' supported on $(\, k^*_{\max}, \, \bar{k} \,]$. All of its mass lies where $w(D, C) \, E(k; D, C) > w(c) \, E(k; c)$ pointwise for every

$c \neq (D, C)$, by Proposition 2(iii) applied to each pair, and integration preserves the strict ranking. □

Proof of Proposition 7. Continuity follows from Lemma 1. For $k \leq r(M, h)$ both terms are zero. On $[\, r(M, h),\, r(D, h)\,]$, $E(k;\, D, h) = 0$ while $E(k;\, M, h) = \phi(M, h)\, g(k - r(M, h))$ is strictly increasing, so Δ_V is strictly increasing on that interval and $\Delta_V(k) \to 0$ as $k \downarrow r(M, h)$. Attenuation follows: as the filer's competing load ℓ rises toward $K - r(M, h)$, task capacity $k = K - \ell$ falls to the entry point of the simpler structure in the comparison, and the contrast converges to zero continuously. The horizontal contrast replaces δ with μ throughout, so the same argument applies with entry points $r(V, T) < r(V, C)$. □

Appendix E. Survey and Reminder-Call Variable Construction

The endline survey was administered in June 2026 through the lead firm's supplier management channels. The instrument was designed to measure the constructs of the model directly: the supplier's organizational capacity endowment, the burdens experienced during reporting, and the reasons why intended reporting did not materialize. The survey was fielded in Chinese. This appendix reproduces in English translation the six items from which the paper's survey variables are built, and then documents the construction of each variable used in Section 6. The fielded instrument also carried items on overall satisfaction with the platform, perceived filing difficulty, and the effort spent selecting indicators and inventorying records. It further recorded person-days of labor input, data availability and cross-functional coordination by ESG domain, the mode of preparing supporting materials, and a five-point scale for unmet reporting intent. None of these enters the analysis, and we list them so that the items we do use can be read against the full set. Respondents were assured that answers would be used only to improve the platform and would not affect any cooperation or evaluation, which limits the incentive to misreport constraints strategically.

E.1 Survey Instrument (English Translation)

Supplier Survey on the Use and Improvement of the Supply Chain ESG Digital Management Platform

To our supplier partners: Thank you for building a sustainable ESG future with us. To further improve the supplier ESG digital management platform and realize shared value across the chain, we kindly ask you to complete the following questionnaire (approximately 5–10 minutes). There are no right or wrong answers. Responses are used only to improve the platform and do not affect any cooperation or evaluation.

Part I. Basic Information

Q1. Please enter your company's supplier code. *[Open response; required]*

(Please use the supplier code assigned by the lead firm. It links this feedback to your cooperation records so that we can provide targeted support.)

Supplier code: _____

Part II. Your Company's Current ESG Work Arrangements

Q2. How is responsibility for ESG and compliance matters currently arranged in your company? *[Single choice]*

1. No designated person; handled ad hoc by whoever is available
2. Handled part-time by administrative, human-resources, or finance staff
3. A dedicated person is assigned to this responsibility
4. A dedicated team or department is in place

Q3. Which of the following best describes how your company currently manages routine operational data, such as production, employment, and energy use? *[Single choice]*

1. Scattered across departments, each keeping its own records, located when needed
2. Partly centralized; some data still collected ad hoc
3. Largely centralized in a unified system and retrievable at any time

Q4. During this reporting round, was the time that the person mainly responsible could devote to the task sufficient? *[Single choice]*

1 Very insufficient 2 Somewhat insufficient 3 Neutral 4 Fairly sufficient 5 Very sufficient

Part III. Situations Encountered During Reporting

Q5. If there were items your company wanted to report but ultimately did not, what were the main reasons? (Select all that apply) *[Multiple choice]*

1. The data cannot be located or does not exist
2. It would require departments or staff that we do not currently have
3. Compiling it would take too much time
4. The specific requirements or definitions were unclear
5. We considered it unnecessary for us
6. It is unrelated to our business
7. There were no items we wanted to report but did not

Part IV. Support You Would Like Us to Provide

Q6. To help your company complete ESG reporting, which forms of support from the group would you most welcome? (Select up to 3) *[Multiple choice; at most three]*

1. Ready-to-use reporting templates
2. Training or explanation sessions
3. Reducing the number of fields to be reported
4. Extending the reporting deadline
5. System or data-interface integration
6. Dedicated one-on-one assistance
7. Other (please specify)

This concludes the questionnaire. Thank you for your cooperation.

E.2 Variable Construction

Table E1 maps each survey item to the variables used in the mechanism analysis. Three conventions apply throughout. First, responses are matched to the experimental sample through the supplier code in Q1; when a supplier submits more than once, the latest submission is retained. Second, constraint measures are coded so that higher values indicate more friction, while capacity endowments (dedicated staff, centralized data, sufficient time) are reported as constraint indicators, so that the figures read as what suppliers lack. Third, binary indicators collapse the underlying scales at the cuts listed below, which separate the burdened tail from the rest of the distribution.

Table E1: Survey items and derived variables

Item	Variable (figure label)	Definition
Q1	Matching key	Supplier code; matched to platform and randomization records to assign treatment status. Latest submission kept per code.
Q2	No dedicated ESG staff	Indicator equal to one if option 1 or 2 is selected (no dedicated person or team).
Q3	No centralized data system	Indicator equal to one if option 1 or 2 is selected (data not held in a unified system).
Q4	Insufficient time	Indicator equal to one if the response is 3 or below (time not clearly sufficient).

Table E1 (continued)

Item	Variable (figure label)	Definition
Q5	No data available; No dept/staff; Too time-consuming; Lack of understanding; Deemed unnecessary; Irrelevant to business	Indicators for each cited reason (options 1–6). Options 1–3 are execution constraints and enter the capacity-barrier count below. Option 4 is the survey counterpart of the comprehension cost $r(c)$ in equation (3) and is reported on its own. Options 5–6 are preference-based reasons.
Q5	Capacity barriers cited (0–3)	Count of options 1–3 selected (no data, no staff, too time-consuming). Option 4, unclear requirements, is excluded from the count and reported separately.
Q5	Any unreported item	Indicator equal to one if option 7 is *not* selected (at least one wanted-but-unreported item exists).
Q6	Demands fewer fields	Indicator equal to one if option 3 (reducing the number of fields) is among the selected forms of support.

Notes: The survey yielded 135 responses from suppliers in the experimental sample, and response rates are balanced across arms (Figure F10). Survey response itself is recorded as an indicator for whether a supplier in the experimental sample completed the endline questionnaire. Treatment contrasts in Section 6 regress each derived variable on the same arm indicators used for the platform outcomes, with randomization-stratum fixed effects and stratum-clustered standard errors.

E.3 Survey Evidence on the Capacity Channel

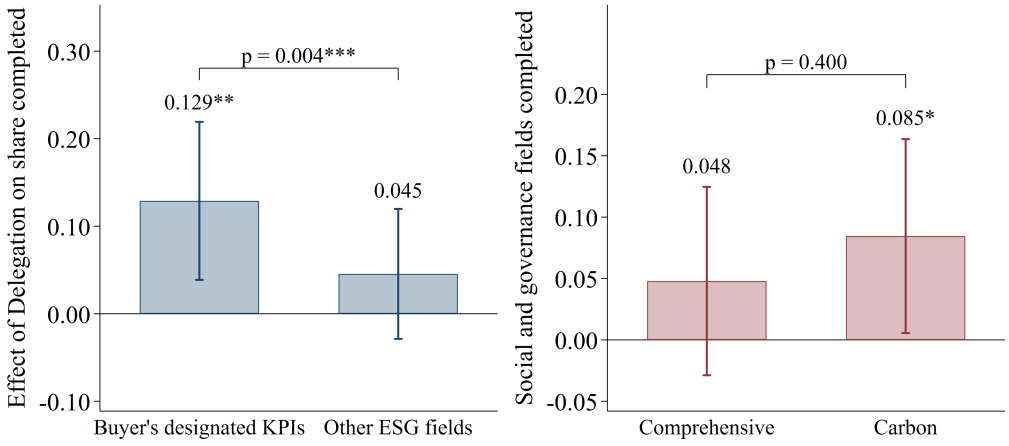
Section 6.6 summarizes the survey-based constraint measures. This subsection gives the figure-by-figure detail.

The two figures support the governance rankings by showing how the arms differ in perceived capacity. In the vertical contrast, delegation increases the number of capacity barriers suppliers cite, raises reported time shortages, and generates stronger demand for a smaller set of fields, while mandate leaves these frictions near control levels (Figure F14). In the horizontal contrast, comprehensive monitoring raises the same constraint measures relative to targeted monitoring (Figure F15). The arms that produce weaker platform engagement are exactly the arms in which suppliers describe tighter capacity, which is the ordering the channel requires.

E.4 Reminder-Call Engagement Measures

The September 2025 briefings described in Section 3.3 were structured reminder calls placed ahead of the upcoming filing. The team recorded and transcribed 150 conversations, of which 146 match a supplier in the experimental sample and enter the estimation sample. From the transcripts we construct four measures of how a supplier engages with the request outside the platform. *Speech share* is the fraction of the total conversation spoken by the supplier. *Question count* is the number of substantive supplier questions about the platform, the reporting cycle, or data requirements. The *written follow-up* indicator equals one if the supplier requests materials, links, or instructions after the call. The *engagement index* is the equal-weighted mean of the standardized values of these three measures, placing the continuous and binary components on a common scale. Suppliers were assigned to the two callers at random and caller assignment is balanced across arms. Treatment contrasts are estimated in the call-level cross-section with randomization-stratum and caller fixed effects, with standard errors clustered by stratum. The caller fixed effect absorbs systematic differences in conversational style across the two callers and improves precision.

Appendix F. Additional Figures and Tables



Note: Left, effect of Delegation on the share completed of the buyer's seven designated KPIs and of the other 71 ESG fields. Right, effects of the monitoring arms on the share of social and governance fields completed, which the carbon audit does not cover. Error bars present 90% confidence intervals. Brackets report the two-sided p-value of equality. *** p<0.01, ** p<0.05, * p<0.1

Figure F1: Supplier Choices and the Buyer's Interests

Notes: This figure asks whether the governance arms shifted reporting away from what the buyer values. The left panel reports the effect of Delegation, relative to control, on the share of fields completed, separately for the seven of the buyer's eight designated KPIs that lie in the applicable field set and for the other 71 ESG fields. The two effects come from one regression stacked by field type, with reporting-round and randomization-stratum fixed effects interacted with field type and standard errors clustered by supplier, and the bracket reports the p-value of a test that the two effects are equal. The right panel reports the effects of Comprehensive Monitoring and Targeted Carbon Monitoring, relative to control, on the share of social and governance fields completed, which the carbon audit does not cover. The regression includes reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Bars show coefficient estimates and whiskers show 90 percent confidence intervals. Suppliers free to choose leaned toward the indicators the buyer would have mandated, and a carbon-only audit did not reduce completion of the unaudited fields.

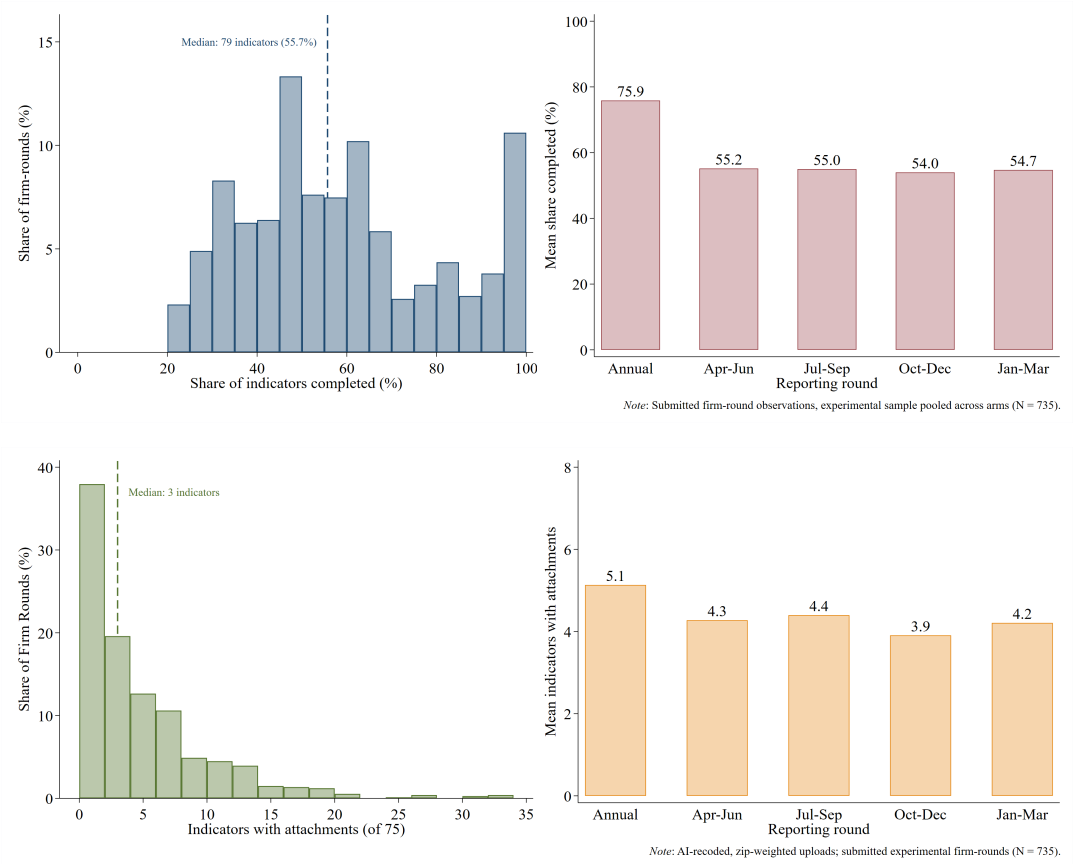


Figure F2: Reporting Completion and Supporting Attachments Across Filing Rounds

Notes: This figure presents four descriptive panels for the 735 submitted supplier-round observations in the experimental sample. Panel A shows the distribution of the share of indicators completed. The median supplier-round completes 79 indicators, equivalent to 55.7 percent of the indicator set. Panel B reports mean completion by filing round. Panel C shows the distribution of the number of indicators documented by at least one attachment, out of 75. The median supplier-round documents three. Panel D reports the mean number of documented indicators by filing round. The annual filing documents more indicators and completes more of the form than the quarterly filings, while the four quarterly rounds display similar completion and documentation patterns.

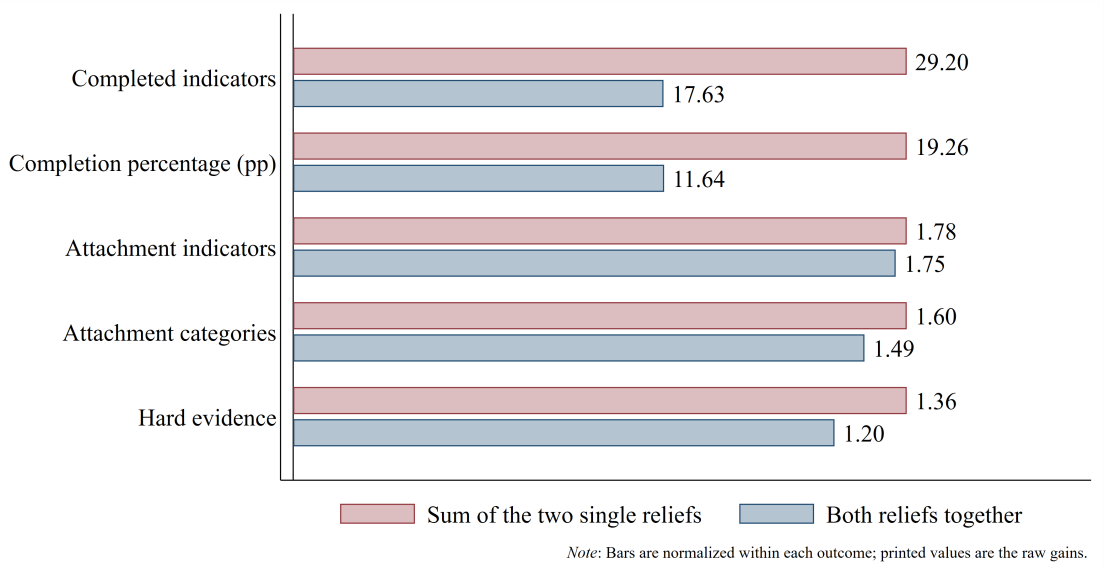


Figure F3: Compounding of the Two Reporting Burdens

Notes: This figure compares the sum of the two single reliefs with the relief delivered by both together, for each of the five engagement outcomes. All quantities are measured from the fully complex cell, which combines delegated KPI selection with comprehensive monitoring, and are signed so that a positive value means better reporting. Every outcome is signed so that a positive value means better reporting, and the completion share is expressed in percentage points. The upper bar in each pair is the sum of the gain from mandating the indicator set and the gain from narrowing the audit to carbon, each measured on its own. The lower bar is the gain from the cell that does both. A joint gain below the sum means the two burdens compound, and a joint gain above the sum means they overlap. Bars are normalized within each outcome, and the printed values are the raw gains. All quantities come from the four-cell regression reported in Table F4, estimated on 735 supplier-round observations with reporting-round and randomization-stratum fixed effects and standard errors clustered by supplier. None of the five reported second differences is statistically significant at the 10 percent level. Table F7 reports the underlying gains, the second differences, their standard errors, and their p-values.

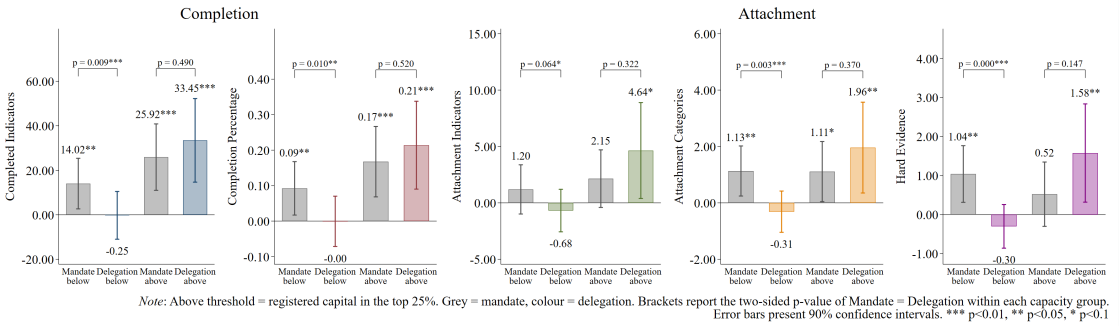


Figure F4: Supplier Capacity and the Returns to Authority Allocation

Notes: This figure reports the effects of Mandate and Delegation separately for suppliers below and above a capacity threshold, defined as registered capital in the top quartile of the sample (38 of the 150 suppliers with reported capital: 9 control, 18 mandate, and 11 delegation). The five columns report completed indicators, the completion share, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Estimates come from equation (2), which interacts both arms with indicators for below and above the threshold and enters the threshold indicator on its own. Control suppliers are the omitted category within each capacity group, so every coefficient is an arm effect measured against controls of the same capacity group. Bars are grouped by capacity group and shaded by arm. The bracket over each pair reports the two-sided p-value of a test that the Mandate and Delegation coefficients are equal within that capacity group. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Delegation leaves engagement unchanged below the threshold and raises it on all five outcomes above the threshold, where it also exceeds Mandate. The change in the simple-minus-complex gap across the two capacity groups carries the predicted sign on all five outcomes and is significant at the ten percent level on four of them with supplier-clustered standard errors ($p = 0.080$, 0.090, 0.122, 0.039, and 0.005). Clustering on the thirteen strata makes all five significant ($p = 0.084$, 0.098, 0.094, 0.023, and 0.002), and under a wild-cluster bootstrap with Webb weights over those strata documented indicators, attachment categories, and hard-evidence categories remain significant ($p = 0.097$, 0.022, and 0.002).

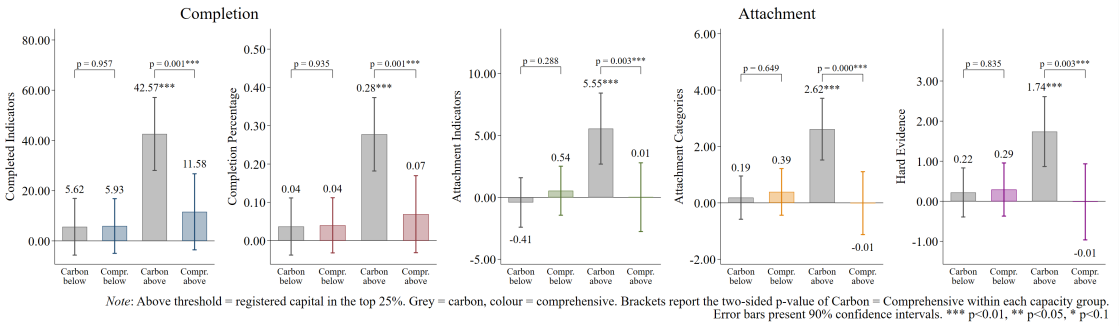


Figure F5: Supplier Capacity and the Returns to Monitoring Scope

Notes: This figure repeats the design of Figure F4 for Targeted Carbon Monitoring and Comprehensive Monitoring. The capacity threshold is registered capital in the top quartile of the sample, and each coefficient is an arm effect measured against control suppliers of the same capacity group. The five columns report completed indicators, the completion share, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. The bracket over each pair reports the two-sided p-value of a test that the Targeted Carbon and Comprehensive coefficients are equal within that capacity group. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Both monitoring arms respond more above the threshold, and the entire advantage of targeted monitoring lies above it. Below the threshold the two arms are statistically indistinguishable on every outcome.

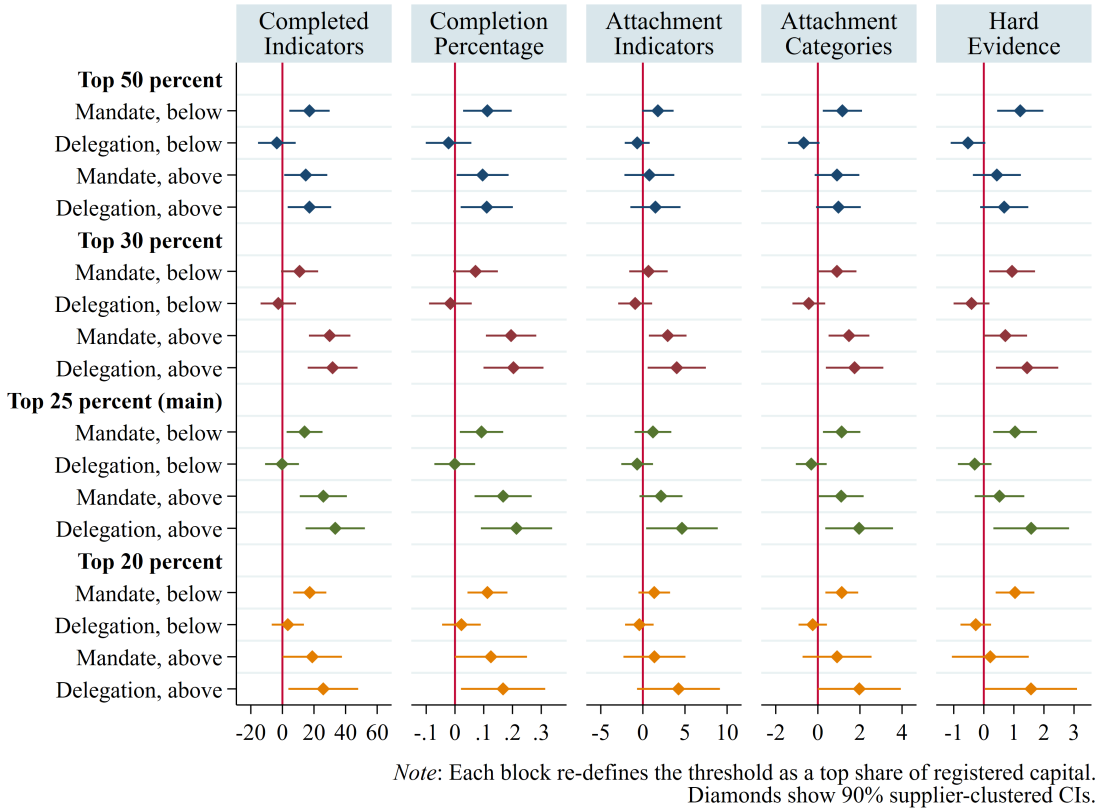


Figure F6: Authority Allocation and the Definition of the Capacity Threshold

Notes: This figure re-estimates Figure F4 at four definitions of the capacity threshold: registered capital in the top 50, 30, 25, and 20 percent of the sample. The top 25 percent block is the specification used elsewhere in the paper. Each block reports the four coefficients of the fully interacted regression, so every estimate is an arm effect measured against control suppliers of the same capacity group. Columns report completed indicators, the completion share, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. The pattern is the same at every threshold. Delegation is flat below the cut and positive above it, and the gap between the two arms narrows and changes sign as the threshold rises.

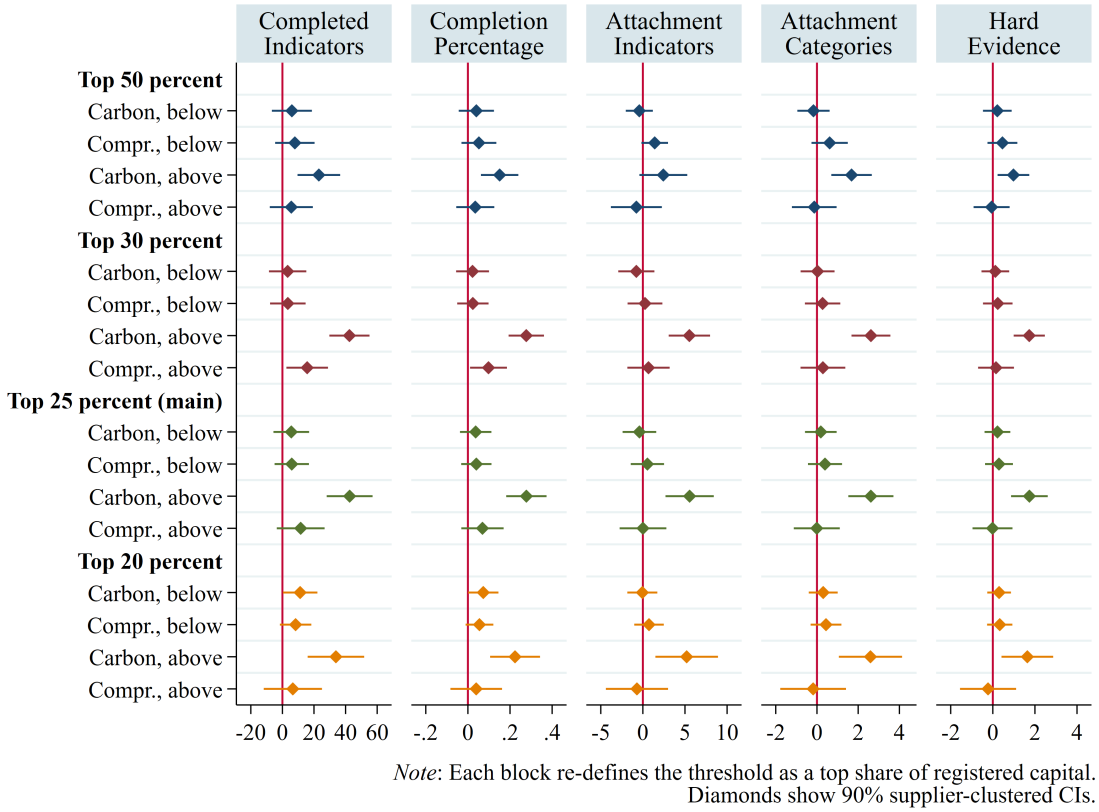
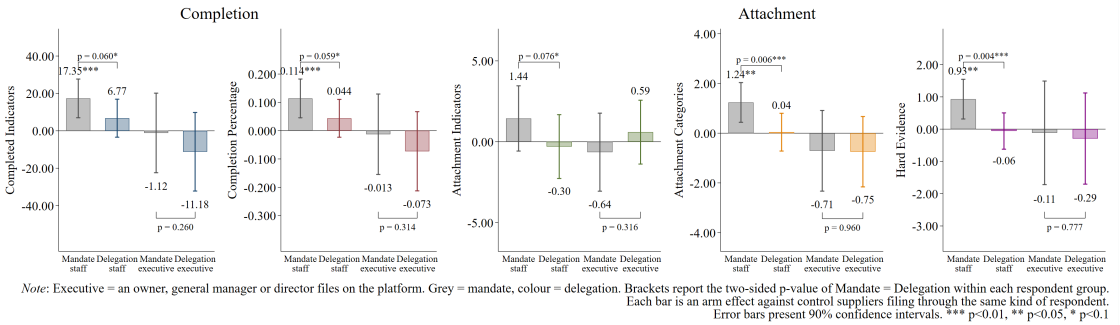


Figure F7: Monitoring Scope and the Definition of the Capacity Threshold

Notes: This figure repeats Figure F6 for Targeted Carbon Monitoring and Comprehensive Monitoring at the same four threshold definitions. Comprehensive Monitoring does not overtake targeted monitoring at any of them, and the response to targeted monitoring above the threshold grows as the threshold rises. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier.



Note: Executive = an owner, general manager or director files on the platform. Grey = mandate, colour = delegation. Brackets report the two-sided p-value of Mandate = Delegation within each respondent group. Each bar is an arm effect against control suppliers filing through the same kind of respondent. Error bars present 90% confidence intervals. *** p<0.01, ** p<0.05, * p<0.1

Figure F8: Respondent Role and the Returns to Authority Allocation

Notes: This figure reports the effects of Mandate and Delegation separately for suppliers whose platform filing is handled by a staff member and for those whose filing is handled by an owner, general manager, or other top executive. The five columns report completed indicators, the completion share, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Estimates come from equation (2) with the executive indicator in place of the capacity indicator, so control suppliers are the omitted category within each respondent group and every coefficient is an arm effect measured against controls filing through the same kind of respondent. Bars are grouped by respondent type and shaded by arm. The bracket over each pair reports the two-sided p-value of a test that the Mandate and Delegation coefficients are equal within that respondent group. The analysis uses the 702 supplier-round observations for which respondent role can be assigned. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Mandate raises engagement where a staff member files and moves nothing where an owner or general manager files, and no arm in the executive group differs from the control group.

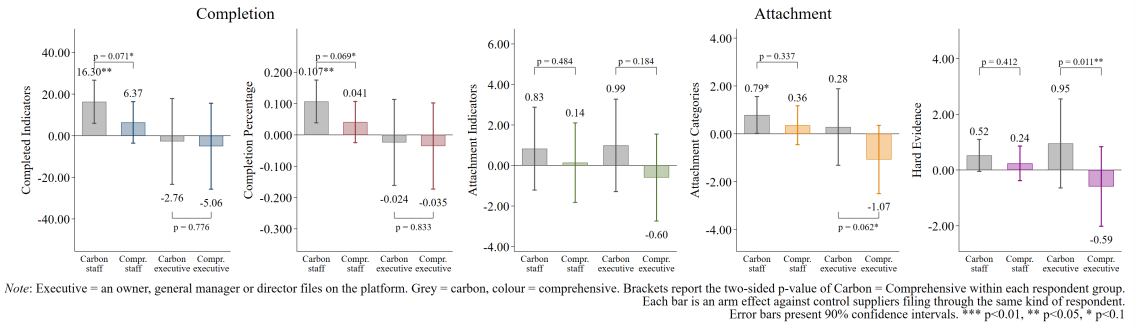


Figure F9: Respondent Role and the Returns to Monitoring Scope

Notes: This figure repeats the design of Figure F8 for Targeted Carbon Monitoring and Comprehensive Monitoring. Each coefficient is an arm effect measured against control suppliers filing through the same kind of respondent. The five columns and the specification follow Figure F8, and the bracket over each pair reports the two-sided p-value of a test that the Targeted Carbon and Comprehensive coefficients are equal within that respondent group. Targeted monitoring raises completed indicators and the completion share where a staff member files, and neither monitoring arm moves any outcome where an owner, general manager, or director files.

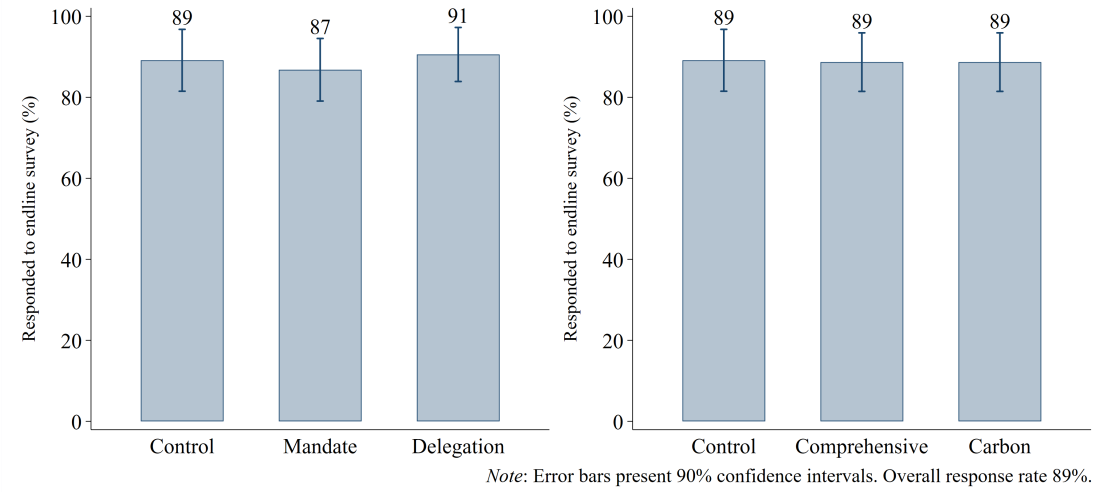
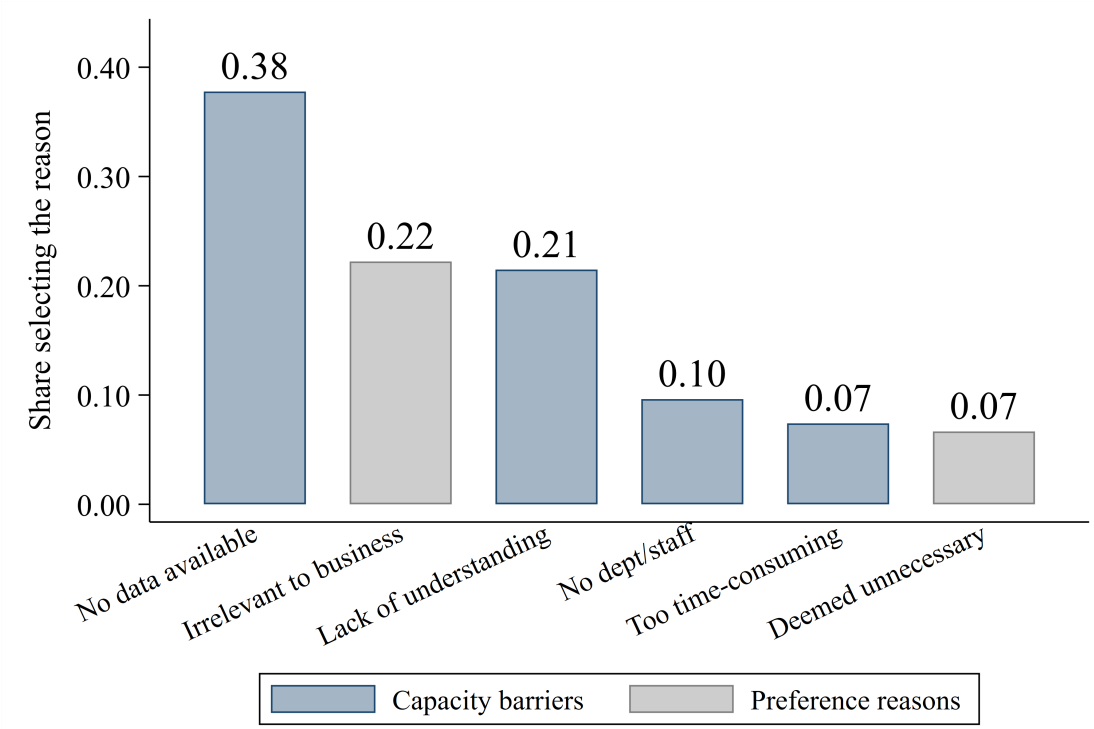


Figure F10: Endline Survey Response and Experimental Balance

Notes: This figure reports survey participation by experimental arm and balance among respondents. The endline survey contains 135 responses from suppliers in the experimental sample. Response rates and observed supplier characteristics are similar across treatment arms, supporting comparisons of reported barriers and organizational infrastructure.



Note: Reasons for wanted-but-unreported items (F1, multi-select), sorted by frequency. Navy = capacity barriers; gray = preference reasons. Matched sample.

Figure F11: Reasons for Desired but Unreported ESG Items

Notes: This figure summarizes the reasons suppliers give for ESG fields that they wanted to report but left unreported. The survey separates execution constraints, namely unavailable data, staffing gaps, and time costs, from unclear requirements, which the model treats as a comprehension cost, and from preference-based reasons. Execution constraints account for 74 of the 142 citations and unclear requirements for a further 29, connecting the observed shortfall in completion to the administrative resources required to assemble and submit information.

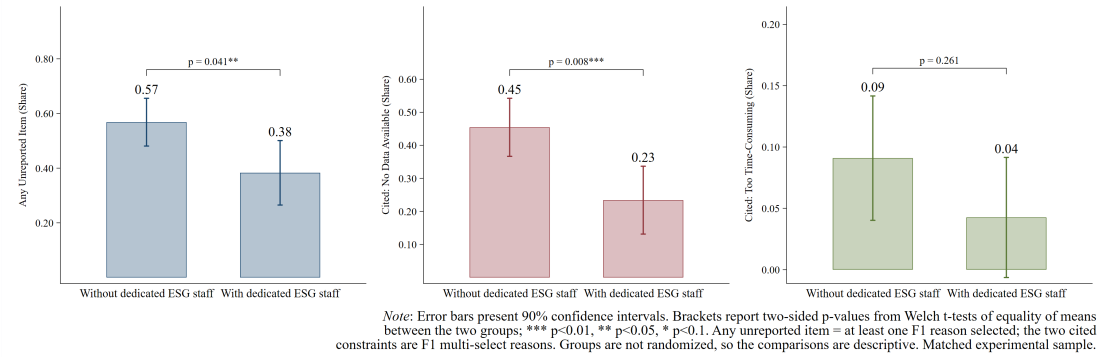
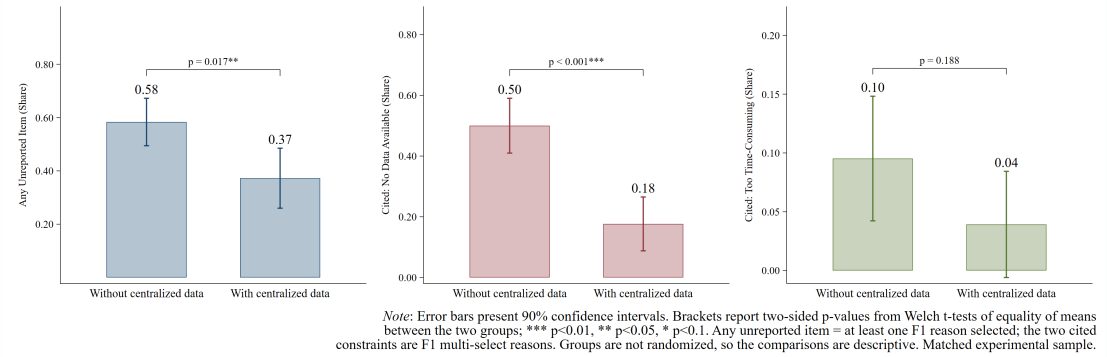


Figure F12: Reporting Barriers and Dedicated ESG Personnel

Notes: This figure relates survey-reported barriers to the presence of dedicated ESG personnel. Suppliers without dedicated staff leave more desired items unreported and are more likely to cite unavailable data, limited staffing, and time costs. The bracket above each pair of bars reports the two-sided p-value from a Welch test of equality of means between the two groups, with $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Dedicated staffing is not randomly assigned, so the comparisons are descriptive and are used to show that an observable feature of organizational capacity tracks the reporting frictions emphasized in the model.



Note: Error bars present 90% confidence intervals. Brackets report two-sided p-values from Welch t-tests of equality of means between the two groups; *** p<0.01, ** p<0.05, * p<0.1. Any unreported item = at least one F1 reason selected; the two cited constraints are F1 multi-select reasons. Groups are not randomized, so the comparisons are descriptive. Matched experimental sample.

Figure F13: Reporting Barriers and Centralized Data Systems

Notes: This figure relates survey-reported barriers to whether the supplier maintains ESG reporting data in a centralized system. Suppliers without centralized data report more desired but unreported fields and more often cite missing information and time costs. The bracket above each pair of bars reports the two-sided p-value from a Welch test of equality of means between the two groups, with $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. These descriptive relationships show that fragmented internal records are associated with the reporting gaps observed on the platform.

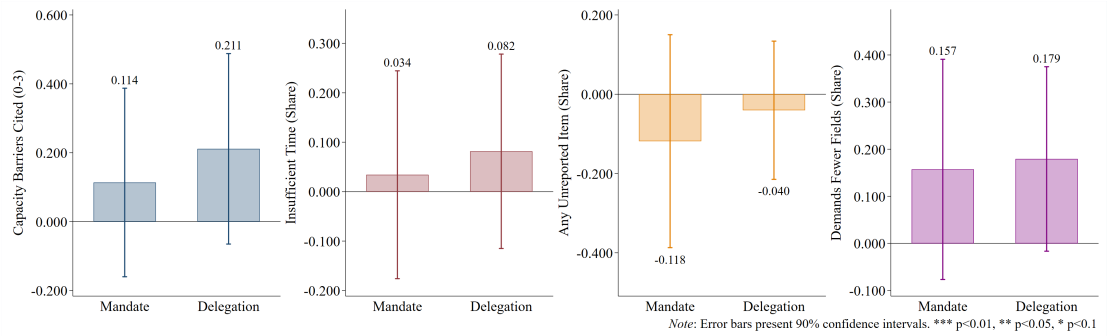


Figure F14: Delegated KPI Choice and Reported Friction

Notes: This figure reports endline-survey treatment contrasts related to the cost of choosing KPIs. The outcomes are the count of capacity barriers cited, reported time shortage, the presence of any wanted-but-unreported item, and demand for a smaller number of reporting fields, each defined in Appendix E. The delegation estimates are compared with the corresponding mandate and control conditions. The pattern indicates that selecting from a broader menu can use administrative resources before data collection begins.

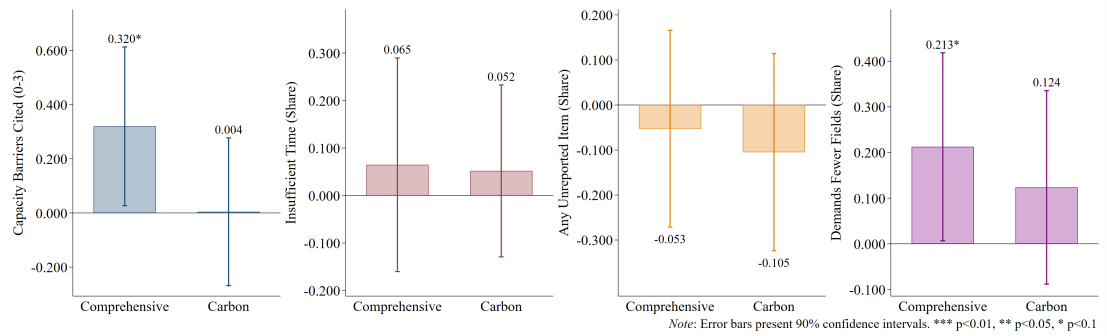


Figure F15: Monitoring Focus and Reported Capacity Barriers

Notes: This figure reports endline-survey contrasts between Comprehensive Monitoring and Targeted Carbon Monitoring. The outcomes measure the number and type of capacity barriers suppliers encounter during reporting. The comprehensive assignment produces a larger reported burden on several measures, consistent with the need to coordinate information across a broader verification scope.

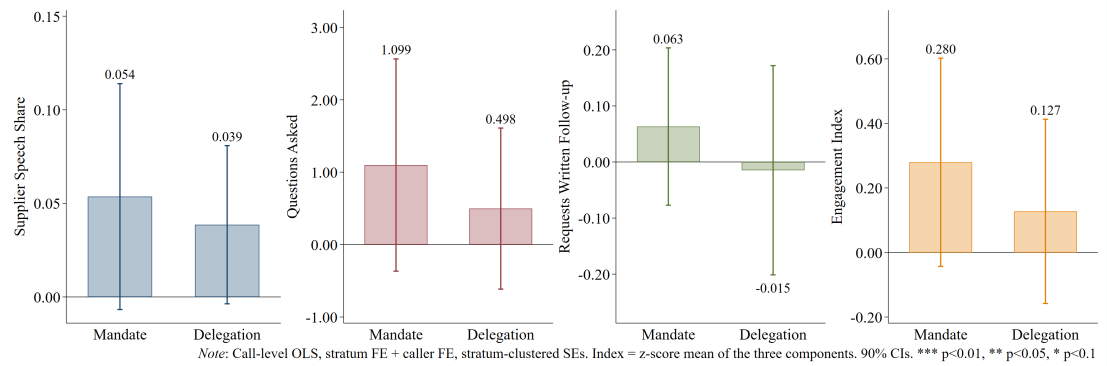


Figure F16: Authority Allocation and Engagement During Reminder Calls

Notes: This figure reports the effects of Mandate and Delegation on supplier engagement during reminder calls. Outcomes include the supplier's share of total speech, the number of questions asked, whether the supplier requested written follow-up, and a standardized index combining these measures. Regressions include caller and randomization-stratum fixed effects, with standard errors clustered by stratum. Mandate estimates are positive across the four engagement measures.

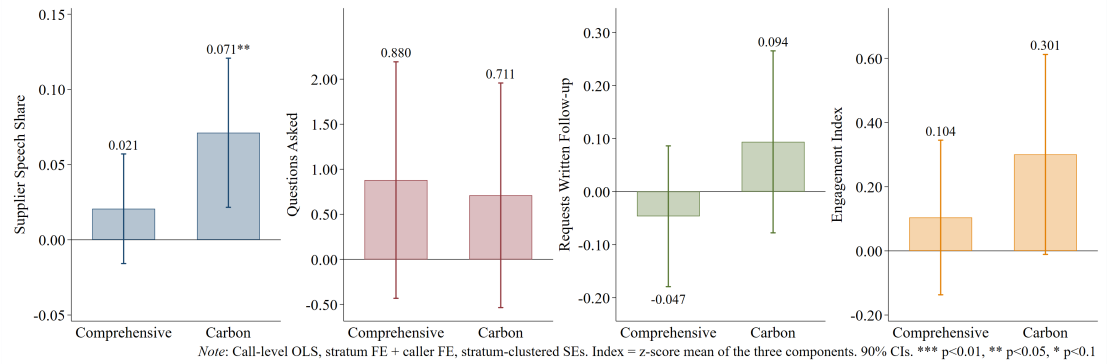


Figure F17: Monitoring Scope and Engagement During Reminder Calls

Notes: This figure reports the effects of Comprehensive Monitoring and Targeted Carbon Monitoring on reminder-call engagement. The outcomes and specifications follow Figure F16. Suppliers assigned to targeted monitoring speak for a larger share of the conversation and have positive estimates for questions, written follow-up requests, and the standardized engagement index. Table F10 reports the full regression statistics.

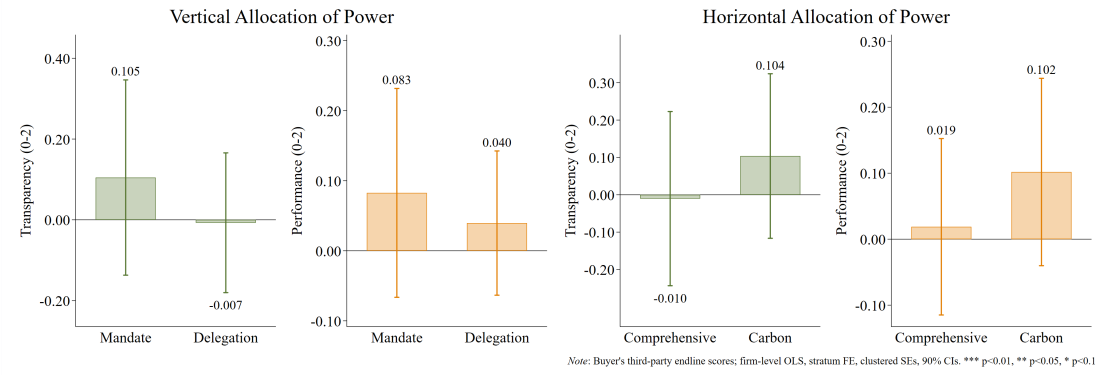
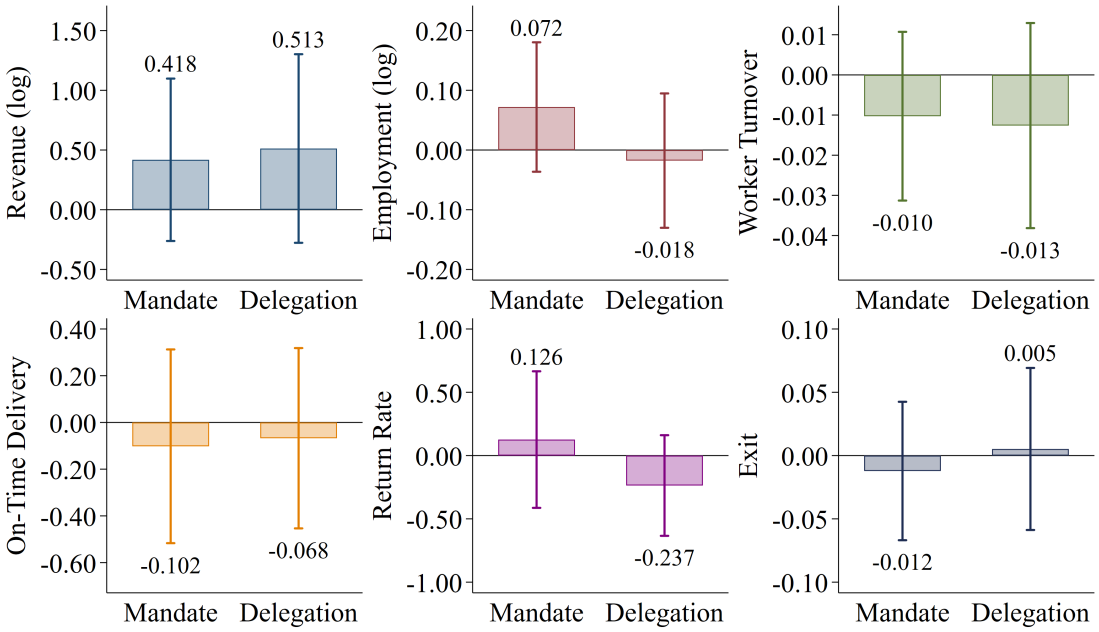


Figure F18: Supplier ESG Performance Rated by the Lead Firm

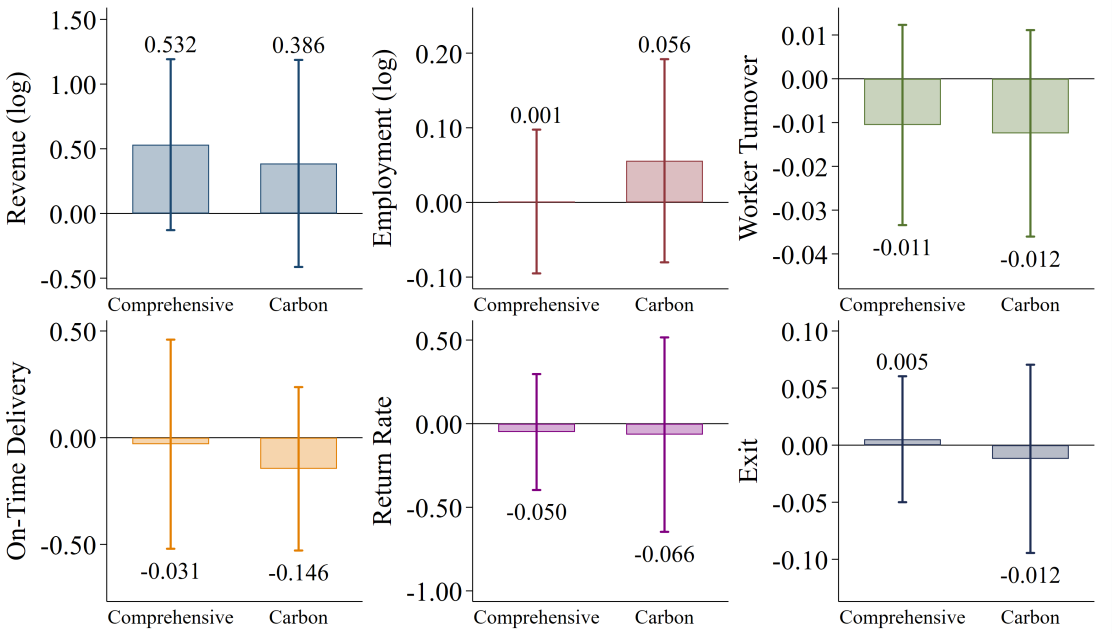
Notes: This figure reports treatment effects on two endline supplier scores assigned by the lead firm: transparency of optional information and a performance bonus. The scoring team did not observe treatment status. Each cross-sectional regression includes randomization-stratum fixed effects. The mandate and targeted-monitoring coefficients are positive for both outcomes, while the confidence intervals include zero. Table F11 reports the coefficient estimates, constants, observations, and R^2 values.



Note: Vertical allocation of authority. Five outcomes internal to the supplier, plus exit from the buyer network. Revenue and employment in logs, the rest in levels. ANCOVA on FY2024 where a baseline exists (Exit: none). Stratum FE, stratum-clustered SEs, 90% CIs. *** p<0.01, ** p<0.05, * p<0.1

Figure F19: Authority Allocation and Supplier Business Outcomes

Notes: This figure reports the effects of Mandate and Delegation on annual revenue, total employment, the voluntary worker turnover rate, the overall on-time delivery rate, the overall return rate, and exit from the buyer network. The first five outcomes are internal to the supplier's own business and the sixth records continued participation in the buyer's network. Revenue and employment enter in logs and the remaining outcomes as they are measured. The five outcomes with a fiscal-2024 counterpart are estimated by endline OLS controlling for that value, and exit is estimated in an endline cross-section. The figure shows coefficient estimates with 90 percent confidence intervals. Table F12 reports the full numerical results and specification statistics.



Note: Horizontal allocation of monitoring scope. Five outcomes internal to the supplier, plus exit from the buyer network. Revenue and employment in logs, the rest in levels. ANCOVA on FY2024 where a baseline exists (Exit: none). Stratum FE, stratum-clustered SEs, 90% CIs. *** p<0.01, ** p<0.05, * p<0.1

Figure F20: Monitoring Scope and Supplier Business Outcomes

Notes: This figure reports the effects of Comprehensive Monitoring and Targeted Carbon Monitoring on the same six outcomes used in Figure F19, under the same specifications. No arm moves any of the six, and every confidence interval covers zero.

Table F1: ESG Indicators on the Digital Platform

Category	Count	Required	Example indicators
Environmental	25	Varies	Renewable-electricity share; chemical-usage inventory; rainwater-harvesting system; monthly energy consumption by source
Social	48	Varies	Involuntary and voluntary turnover; labor-contract coverage; wage-equity metrics; social-insurance coverage
Governance	5	Varies	Dedicated ESG management department; comprehensive ESG policy; anti-corruption and compliance policy
Financial	8	All suppliers	Annual output value; on-time delivery rate; number of partner brands; return rates
Certifications	22	All suppliers	ISO 14001; SA8000 Social Accountability; Bluesign Certification; HIGG FEM
Total	108		

Notes: The platform tracks 108 supply chain indicators organized by reporting domain. Financial and certification metrics are required for all suppliers and form the common baseline. Requirements for environmental, social, and governance indicators vary by experimental assignment. Suppliers in the Mandate arm are required to report a fixed set of eight ESG KPIs spanning these domains, whereas suppliers in the Delegation arm select eight KPIs from the available menu. The complete indicator list is reproduced in Appendix A.

Table F2: Authority Allocation and Supplier Engagement

	(1) Completed	(2) Completion share	(3) Indicators	(4) Categories	(5) Hard evidence
Mandate	15.939***	0.104***	1.075	0.958**	0.766**
	(5.389)	(0.036)	(1.044)	(0.422)	(0.334)
Delegation	6.884	0.045	0.221	0.081	0.029
	(5.479)	(0.036)	(1.024)	(0.401)	(0.307)
Constant	77.675***	0.537***	3.934***	1.892***	1.065***
	(4.211)	(0.028)	(0.821)	(0.302)	(0.237)
Observations	735	735	735	735	735
R^2	0.202	0.323	0.172	0.172	0.171
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Supplier	Supplier	Supplier	Supplier
p-value, arm equality	0.064	0.065	0.327	0.029	0.018

Notes: This table reports OLS estimates of the effects of mandated and delegated KPI selection relative to the control group. Completed is the number of applicable indicator fields the supplier filled in and Completion share is the corresponding proportion. Indicators is the number of distinct reporting items, out of the 75 in the evidence taxonomy of Appendix C, for which the supplier uploaded at least one supporting document. Categories is the number of attachment categories covered out of 12, and Hard evidence is the number of categories supported by verifiable records. All specifications include reporting-round and randomization-stratum fixed effects. Robust standard errors clustered by supplier are reported in parentheses. The estimand throughout is each governance structure measured against no ESG program, which is the comparison the buyer faces. The final row reports, for completeness, the p-value from a test of equality between the two arm coefficients. That test compares treated suppliers with one another and pools suppliers of differing capacity, whose responses Section 6.3 shows to differ. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F3: Monitoring Scope and Supplier Engagement

	(1) Completed	(2) Completion share	(3) Indicators	(4) Categories	(5) Hard evidence
Comprehensive monitoring	7.246	0.047	0.219	0.227	0.175
	(5.270)	(0.035)	(1.008)	(0.418)	(0.331)
Targeted carbon monitoring	15.554***	0.102***	1.075	0.806**	0.615**
	(5.572)	(0.037)	(1.057)	(0.405)	(0.311)
Constant	77.692***	0.537***	3.936***	1.894***	1.067***
	(4.193)	(0.028)	(0.820)	(0.301)	(0.236)
Observations	735	735	735	735	735
R^2	0.200	0.321	0.172	0.159	0.155
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Supplier	Supplier	Supplier	Supplier
p-value, arm equality	0.091	0.090	0.321	0.139	0.151

Notes: This table reports OLS estimates of the effects of Comprehensive Monitoring and Targeted Carbon Monitoring relative to the control group. Outcome definitions follow Table F2. Every specification contains reporting-round and randomization-stratum fixed effects. Robust standard errors clustered by supplier are reported in parentheses. The estimand throughout is each governance structure measured against no ESG program, which is the comparison the buyer faces. The final row reports, for completeness, the p-value from a test of equality between the two arm coefficients. That test compares treated suppliers with one another and pools suppliers of differing capacity, whose responses Section 6.3 shows to differ. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F4: Engagement Across the Four Governance Bundles

	(1) Completed	(2) Completion share	(3) Indicators	(4) Categories	(5) Hard evidence
Mandate, comprehensive	14.605**	0.095**	0.661	0.695	0.585
	(6.038)	(0.040)	(1.221)	(0.546)	(0.441)
Mandate, targeted carbon	17.297**	0.113**	1.521	1.240***	0.961***
	(6.736)	(0.045)	(1.172)	(0.445)	(0.347)
Delegation, comprehensive	-0.328	-0.003	-0.230	-0.250	-0.242
	(6.024)	(0.040)	(1.127)	(0.457)	(0.347)
Delegation, targeted carbon	13.941**	0.091**	0.662	0.405	0.294
	(6.518)	(0.043)	(1.252)	(0.500)	(0.378)
Constant	77.670***	0.537***	3.933***	1.892***	1.065***
	(4.203)	(0.028)	(0.822)	(0.302)	(0.237)
Observations	735	735	735	735	735
R^2	0.221	0.339	0.177	0.183	0.181
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Supplier	Supplier	Supplier	Supplier
p-value, cell equality	0.009	0.010	0.112	0.002	0.001

Notes: Each coefficient compares one crossed treatment cell with the control group. The four cells combine mandated or delegated KPI selection with comprehensive or targeted carbon monitoring. Outcome definitions follow Table F2. All regressions include reporting-round and randomization-stratum fixed effects, and robust standard errors are clustered by supplier. The estimand throughout is each cell measured against no ESG program. The final row reports, for completeness, the p-value from a test of equality between the fully simplified and the fully complex cell. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F5: Robustness of the Authority-Allocation Results

	Completion share				Attachment categories			
	(1) Main	(2) Stratum cluster	(3) Stratum bootstrap	(4) Collapsed	(5) Main	(6) Stratum cluster	(7) Stratum bootstrap	(8) Collapsed
Mandate	0.104***	0.104**	0.104**	0.108**	0.958**	0.958**	0.958**	0.980**
	(0.036)	(0.045)	(0.045)	(0.046)	(0.422)	(0.430)	(0.430)	(0.430)
Delegation	0.045	0.045	0.045	0.048	0.081	0.081	0.081	0.137
	(0.036)	(0.037)	(0.037)	(0.040)	(0.401)	(0.348)	(0.348)	(0.360)
Constant	0.537***	0.537***	0.537***	0.533***	1.892***	1.892***	1.892***	1.856***
	(0.028)	(0.026)	(0.026)	(0.027)	(0.302)	(0.218)	(0.218)	(0.220)
Observations	735	735	735	152	735	735	735	152
R^2	0.323	0.323	0.323	0.214	0.172	0.172	0.172	0.169
Round fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Stratum	Stratum	Stratum	Supplier	Stratum	Stratum	Stratum

Notes: This table reports four alternative treatments of the authority-allocation estimates. Main is the repeated supplier-round specification of Table F2, which clusters standard errors on the supplier. Stratum cluster retains that coefficient specification and clusters on the stratum instead, the coarser and more conservative level. Stratum bootstrap repeats the stratum-clustered column and differs only in its stars. Collapsed averages the repeated outcomes to the supplier level and omits round fixed effects. All specifications retain randomization-stratum fixed effects. The stars in the stratum-bootstrap column are wild-cluster bootstrap p-values computed with Webb weights and 9,999 replications over the thirteen randomization strata. Bootstrapping over suppliers instead moves no p-value in the table by more than 0.025. Parentheses report standard errors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F6: Robustness of the Monitoring-Scope Results

	Completion share				Attachment categories			
	(1) Main	(2) Stratum cluster	(3) Stratum bootstrap	(4) Collapsed	(5) Main	(6) Stratum cluster	(7) Stratum bootstrap	(8) Collapsed
Comprehensive monitoring	0.047	0.047	0.047	0.054	0.227	0.227	0.227	0.267
	(0.035)	(0.044)	(0.044)	(0.045)	(0.418)	(0.438)	(0.438)	(0.435)
Targeted carbon monitoring	0.102***	0.102**	0.102***	0.102**	0.806**	0.806**	0.806**	0.843**
	(0.037)	(0.036)	(0.036)	(0.037)	(0.405)	(0.315)	(0.315)	(0.318)
Constant	0.537***	0.537***	0.537***	0.533***	1.894***	1.894***	1.894***	1.858***
	(0.028)	(0.025)	(0.025)	(0.027)	(0.301)	(0.207)	(0.207)	(0.209)
Observations	735	735	735	152	735	735	735	152
R^2	0.321	0.321	0.321	0.208	0.159	0.159	0.159	0.156
Round fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Stratum	Stratum	Stratum	Supplier	Stratum	Stratum	Stratum

Notes: This table reports four alternative treatments of the monitoring-scope estimates. The four specifications are the repeated supplier-round model, stratum-level clustered standard errors, the same fit under a wild bootstrap over the strata, and a supplier-level collapsed sample. All models include randomization-stratum fixed effects, and the repeated specifications include reporting-round fixed effects. The stars in the stratum-bootstrap column are wild-cluster bootstrap p-values computed with Webb weights and 9,999 replications over the thirteen randomization strata. Bootstrapping over suppliers instead moves no p-value in the table by more than 0.025. Parentheses report standard errors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F7: Compounding of the Two Reporting Burdens

	(1) Mandate relief	(2) Targeted relief	(3) Both reliefs	(4) Sum of singles	(5) Second difference	(6) p-value
Completed indicators	14.933	14.269	17.625	29.202	−11.577 (9.454)	0.223
Completion share (pp)	9.831	9.429	11.643	19.260	−7.617 (6.249)	0.225
Indicators	0.890	0.892	1.750	1.783	−0.032 (1.694)	0.985
Categories	0.945	0.655	1.490	1.600	−0.110 (0.761)	0.885
Hard evidence	0.827	0.537	1.203	1.363	−0.160 (0.590)	0.786

Notes: This table re-expresses the four-cell regression of Table F4 against the fully complex cell, which combines delegated KPI selection with comprehensive monitoring. Column (1) reports the gain from mandating the indicator set while monitoring remains comprehensive. Column (2) reports the gain from narrowing the audit to carbon while KPI selection remains delegated. Column (3) reports the gain from the cell that does both. Column (4) is the sum of Columns (1) and (2). Column (5) is Column (3) minus Column (4), the second difference, with its standard error in parentheses. Column (6) reports the two-sided p-value for the second difference. All quantities are signed so that a positive value means better reporting, and the completion share is expressed in percentage points. Outcome definitions follow Table F2. Every quantity comes from one regression per outcome, estimated on 735 supplier-round observations with reporting-round and randomization-stratum fixed effects and robust standard errors clustered by supplier. Columns (1) to (5) are linear combinations of the coefficients of that single regression, so observation counts and specification rows are common to the whole table. Entries are computed at full precision and rounded for display, so Column (4) need not equal the sum of Columns (1) and (2) in the final digit. None of the five reported second differences is statistically significant at conventional levels.

Table F8: Engagement Responses Across ESG Dimensions

	Completion share			Any attachment		
	(1) Environmental	(2) Social	(3) Governance	(4) Environmental	(5) Social	(6) Governance
Panel A: Authority allocation						
Mandate	0.145***	0.105**	0.146*	0.281***	-0.013	0.077
	(0.054)	(0.047)	(0.074)	(0.081)	(0.076)	(0.079)
Delegation	0.080	0.013	0.133	0.062	-0.061	0.013
	(0.057)	(0.047)	(0.082)	(0.080)	(0.072)	(0.072)
Constant	0.254***	0.624***	0.600***	0.317***	0.195***	0.155***
	(0.043)	(0.035)	(0.063)	(0.057)	(0.056)	(0.056)
R^2	0.216	0.181	0.104	0.158	0.095	0.089
Panel B: Monitoring scope						
Comprehensive monitoring	0.066	0.040	0.123	0.099	-0.126*	0.007
	(0.053)	(0.046)	(0.079)	(0.081)	(0.066)	(0.072)
Targeted carbon monitoring	0.159***	0.077	0.156**	0.243***	0.053	0.084
	(0.058)	(0.048)	(0.076)	(0.081)	(0.078)	(0.079)
Constant	0.254***	0.624***	0.600***	0.317***	0.195***	0.155***
	(0.043)	(0.035)	(0.063)	(0.056)	(0.056)	(0.056)
R^2	0.223	0.166	0.105	0.139	0.131	0.092
Observations	735	735	735	735	735	735
Round fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Supplier	Supplier	Supplier	Supplier	Supplier

Notes: This table reports treatment effects separately for environmental, social, and governance reporting. Completion-share outcomes are the proportions of indicators filled in within each domain. Attachment outcomes equal one when a supplier uploads at least one supporting document in the relevant domain. Panel A compares the authority treatments with control, and Panel B compares the monitoring treatments with control. Regressions include reporting-round and randomization-stratum fixed effects. Robust standard errors clustered by supplier are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F9: Respondent Role and the Dilution of Treatment Effects

	(1) Completed	(2) Completion share	(3) Indicators	(4) Categories	(5) Hard evidence
Panel A: Authority allocation					
Mandate, staff respondent	17.355***	0.114***	1.442	1.236**	0.930**
	(6.248)	(0.041)	(1.218)	(0.482)	(0.371)
Mandate, executive respondent	-1.116	-0.013	-0.643	-0.709	-0.114
	(12.853)	(0.086)	(1.457)	(0.980)	(0.969)
Delegation, staff respondent	6.765	0.044	-0.297	0.040	-0.058
	(6.132)	(0.040)	(1.191)	(0.459)	(0.339)
Delegation, executive respondent	-11.179	-0.073	0.594	-0.746	-0.291
	(12.680)	(0.084)	(1.192)	(0.854)	(0.853)
Executive respondent	-0.891	-0.005	-2.077	0.276	-0.147
	(12.254)	(0.081)	(1.345)	(0.833)	(0.843)
Constant	79.778***	0.551***	4.399***	1.934***	1.137***
	(4.737)	(0.031)	(0.968)	(0.348)	(0.260)
R^2	0.239	0.352	0.222	0.213	0.199
Panel B: Monitoring scope					
Comprehensive, staff respondent	6.369	0.041	0.138	0.358	0.237
	(6.017)	(0.040)	(1.186)	(0.491)	(0.376)
Comprehensive, executive respondent	-5.063	-0.035	-0.596	-1.075	-0.592
	(12.470)	(0.083)	(1.296)	(0.862)	(0.861)
Targeted, staff respondent	16.298**	0.107**	0.832	0.787*	0.524
	(6.265)	(0.041)	(1.235)	(0.465)	(0.347)
Targeted, executive respondent	-2.762	-0.024	0.992	0.284	0.952
	(12.462)	(0.083)	(1.378)	(0.967)	(0.964)
Executive respondent	-1.225	-0.008	-2.173	0.204	-0.219
	(11.993)	(0.080)	(1.333)	(0.813)	(0.833)
Constant	79.872***	0.552***	4.413***	1.946***	1.150***
	(4.716)	(0.031)	(0.968)	(0.345)	(0.257)
R^2	0.235	0.349	0.210	0.187	0.177
Observations	702	702	702	702	702
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes
Cluster level	Supplier	Supplier	Supplier	Supplier	Supplier

Notes: Each coefficient is an arm effect measured against control suppliers whose platform filing is handled by the same kind of respondent, so control suppliers are the omitted category within each respondent group. The executive-respondent row is the level difference between the two groups of control suppliers. An executive respondent is an owner, general manager, or other top executive. Outcome definitions follow Table F2. All specifications include reporting-round and randomization-stratum fixed effects. Robust standard errors clustered by supplier are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F10: Engagement During Reminder Calls

	(1) Speech share	(2) Questions	(3) Written follow-up	(4) Engagement index
Panel A: Authority allocation				
Mandate	0.054	1.099	0.063	0.280
	(0.034)	(0.822)	(0.079)	(0.181)
Delegation	0.039	0.498	-0.015	0.127
	(0.024)	(0.624)	(0.105)	(0.160)
Caller indicator	0.148***	1.949***	0.333***	0.774***
	(0.025)	(0.501)	(0.065)	(0.141)
Constant	0.056**	0.574	0.125	-0.530***
	(0.021)	(0.474)	(0.073)	(0.115)
R^2	0.313	0.166	0.198	0.271
Panel B: Monitoring scope				
Comprehensive monitoring	0.021	0.880	-0.047	0.104
	(0.020)	(0.736)	(0.074)	(0.135)
Targeted carbon monitoring	0.071**	0.711	0.094	0.301
	(0.028)	(0.699)	(0.096)	(0.175)
Caller indicator	0.142***	1.950***	0.315***	0.748***
	(0.022)	(0.510)	(0.063)	(0.132)
Constant	0.059***	0.574	0.134*	-0.517***
	(0.018)	(0.493)	(0.070)	(0.106)
R^2	0.331	0.160	0.209	0.275
Observations	146	146	146	146
Caller fixed effects	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum

Notes: Speech share is the supplier's share of total speaking time during the reminder call. Questions is the number of questions asked by the supplier. Written follow-up equals one when the supplier requests written information after the call. The engagement index is the standardized mean of speech share, questions, and written follow-up. Panel A reports authority-allocation estimates and Panel B reports monitoring-scope estimates. All regressions include caller and randomization-stratum fixed effects, with robust standard errors clustered by stratum in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F11: Supplier ESG Performance Rated by the Lead Firm

	Transparency		Performance	
	(1)	(2)	(3)	(4)
Mandate	0.105		0.083	
	(0.136)		(0.084)	
Delegation	-0.007		0.040	
	(0.097)		(0.058)	
Comprehensive monitoring		-0.010		0.019
		(0.131)		(0.075)
Targeted carbon monitoring		0.104		0.102
		(0.124)		(0.080)
Constant	0.770***	0.770***	0.455***	0.455***
	(0.077)	(0.076)	(0.032)	(0.031)
Observations	155	155	155	155
R^2	0.182	0.182	0.224	0.227
Stratum fixed effects	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum

Notes: This table reports cross-sectional estimates for two endline scores assigned by the lead firm on a zero-to-two scale. Transparency evaluates optional information, and Performance captures the buyer's assessment of ESG performance. The scoring team did not observe supplier treatment status. All specifications include randomization-stratum fixed effects. Robust standard errors clustered by stratum are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table F12: Financial and Operational Outcomes

	(1) Revenue	(2) Employment	(3) Worker turnover	(4) On-time delivery	(5) Return rate	(6) Exit
Panel A: Authority allocation						
Mandate	0.418	0.072	-0.010	-0.102	0.126	-0.012
	(0.382)	(0.061)	(0.012)	(0.233)	(0.303)	(0.031)
Delegation	0.513	-0.018	-0.013	-0.068	-0.237	0.005
	(0.443)	(0.063)	(0.014)	(0.216)	(0.223)	(0.036)
Baseline outcome	0.833***	0.950***	1.207***	0.626***	0.015	n.a.
	(0.104)	(0.037)	(0.237)	(0.099)	(0.010)	
Constant	0.193	0.358*	0.008	37.151***	0.619***	0.055**
	(0.898)	(0.181)	(0.006)	(9.645)	(0.154)	(0.020)
R^2	0.667	0.952	0.650	0.687	0.144	0.145
Panel B: Monitoring scope						
Comprehensive monitoring	0.532	0.001	-0.011	-0.031	-0.050	0.005
	(0.371)	(0.054)	(0.013)	(0.275)	(0.194)	(0.031)
Targeted carbon monitoring	0.386	0.056	-0.012	-0.146	-0.066	-0.012
	(0.449)	(0.076)	(0.013)	(0.215)	(0.326)	(0.046)
Baseline outcome	0.832***	0.953***	1.209***	0.628***	0.016	n.a.
	(0.104)	(0.037)	(0.232)	(0.098)	(0.009)	
Constant	0.199	0.345*	0.008	36.993***	0.618***	0.055**
	(0.900)	(0.180)	(0.006)	(9.603)	(0.153)	(0.020)
R^2	0.668	0.952	0.650	0.687	0.131	0.144
Observations	131	133	130	131	130	152
Specification	Endline OLS	Endline OLS	Endline OLS	Endline OLS	Endline OLS	Endline CS
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table reports effects on annual revenue, total employment, the voluntary worker turnover rate, the overall on-time delivery rate, the overall return rate, and exit from the buyer network. The first five outcomes are internal to the supplier's own business and the sixth records continued participation in the buyer's network. Revenue and employment enter in logs and the remaining outcomes as they are measured, with the two delivery rates in percentage points. The five outcomes with a fiscal-2024 counterpart are estimated by OLS controlling for that value. Exit is an endline indicator and has no corresponding baseline outcome. Panel A reports authority-allocation estimates and Panel B reports monitoring-scope estimates. Robust standard errors clustered by randomization stratum are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.