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 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 audits) to understand how contract design shapes supplier engagement. Mandates and targeted audits lead suppliers to report more information and provide more supporting documentation, while delegation and comprehensive audits produce smaller, imprecisely estimated responses. We interpret these patterns through the trade-off between the informational and productivity benefits of richer contracts and the capacity required to implement them. We develop a model in which comprehension and execution share a capacity budget, making simpler contracts more effective when capacity is limited. Heterogeneity by task tractability, supplier capacity, and reporting role supports this capacity interpretation.

JEL Codes: C93, D23, D82, L14, M14, Q56

Keywords: supply chain governance, field experiment, delegation, multitasking, monitoring, organizational capacity, ESG reporting, private regulation

*We are grateful to Kevin Bauer, Jordi Blanes i Vidal, Jing Cai, Wouter Dessein, Florian Englmaier, Guido Friebel, Ricard Gil, Gianmarco León-Ciliotta, Bingjing Li, Jin Li, Catherine Thomas, Shaoda Wang, and Yang Xie for their insightful comments and suggestions. We also thank participants at the Betty Sadoulet Memorial Conference, CCER Summer Institute, CEPR IMO-ESF Workshop, and Guanghua-Organization Science Paper Development Workshop for valuable feedback, as well as the partner firms' 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 because the analysis uses only firm-level data without personally identifiable information. 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). Governing these relationships requires allocating decision authority along the supply chain and setting the scope of monitoring across performance dimensions. These choices have long been central to the economics of organization, and buyers settle them in practice each time they design 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 independently vary authority over KPI selection and the scope of monitoring. 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 ESG data or targeted verification of carbon metrics alone. A separate control group faces the baseline arrangement, under which ESG reporting is 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. These records make supplier practices visible to the buyer and provide a basis for corrective action and the allocation of future business (Bernstein, 2015; Andrews and Barron, 2016).

These two experimental dimensions map onto classical trade-offs in organizational economics. On authority, delegation can exploit an agent's local information and strengthen incentives to acquire and use that information, while retaining or restricting authority preserves the principal's control when objectives differ (Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008). In our setting, a supplier may know which records it can readily assemble, while the buyer has its own priorities over what information should be reported. On monitoring, broader performance measurement can sustain effort across multiple valued dimensions, while a narrower focus can sharpen incentives and coordination when multiple objectives dilute accountability or compete for attention (Holmström and Milgrom, 1991; Dewatripont et al., 1999; Dessein et al., 2016). Which arrangement performs better therefore depends on the information structure, task measurability, incentives, and attention. These forces are difficult to separate in observed contracts because buyers tailor governance to suppliers and often bundle authority with monitoring. Our experiment isolates both dimensions within the same commercial relationship.

The main effects favor mandate on the authority dimension and targeted monitoring on the monitoring dimension. Mandating a fixed set of indicators raises completeness by 10.5 percentage points relative to the voluntary-reporting control group, or 19.7 percent of the control mean, and broadens suppliers' documentation. Delegation produces smaller estimated changes on both margins. Targeted carbon monitoring generates gains of similar size, raising completeness by 10.3 percentage points and expanding documentation coverage and hard-evidence coverage, whereas comprehensive

monitoring produces smaller responses. In the crossed design, mandate combined with targeted monitoring yields the largest point estimates on every engagement outcome, while delegation combined with comprehensive monitoring yields the smallest.

We interpret these results as reflecting limits on the organizational capacity suppliers can devote to a new administrative task. Understanding a governance requirement and organizing the information needed to meet it require resources before execution begins (Arrow, 1974). In our setting, choosing among unfamiliar indicators or preparing for a broad audit requires staff time, records, and coordination. We develop a model in which these implementation costs draw on the same capacity budget as execution, while delegation and broader monitoring can generate informational or productivity benefits once those costs are absorbed. The balance between these costs and benefits therefore depends on the supplier's capacity to implement the arrangement.

The remaining evidence points in a similar direction. The interaction estimates are negative across all five engagement outcomes, providing suggestive evidence that the two implementation burdens compound. The documentation gains concentrate in environmental reporting, where routine quantitative records and established verification protocols make the information easier to substantiate. Among suppliers classified using pre-treatment capacity proxies, delegation has little effect in the lower-capacity group but raises engagement in the higher-capacity group, where its point estimates exceed mandate's on all five outcomes. Engagement gains also vary by reporting role, appearing where the supplier's standing platform contact is a staff member and fading where the contact is an owner or top executive. The endline survey and telephone briefing transcripts offer suggestive support for the capacity interpretation. Blinded buyer assessments yield higher point estimates for mandate and targeted monitoring, and we detect no statistically significant changes in the measured business outcomes.

Our paper makes three contributions. The first is to provide experimental evidence

on how two central problems of organizational design operate across firm boundaries. A growing share of world production is coordinated through contracts between legally independent firms (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 (Bernstein, 2015). Do organizational theories developed largely for the interior of the firm also govern these arrangements? This connection is developed in Gibbons (2005), but commercial contract terms are rarely available for experimental manipulation. Evidence on delegation and monitoring comes largely from laboratory settings, within-firm relationships, and public bureaucracies (Fehr et al., 2013; Rasul and Rogger, 2018; Bandiera et al., 2021; Kala, 2024). In these settings, the principal typically hires the agent and can shape its administrative environment. Our experiment instead studies governance across firms, where buyers must implement requirements through independent suppliers whose capabilities and internal organization are largely given and whose resources buyers do not directly control. Field experiments in supplier networks have varied relational support (Ghosh and Singh, 2026), procurement incentives (Casaburi and Reed, 2022), and compliance enforcement (Boudreau, 2024). We instead vary who decides what must be reported and how widely performance is verified, providing direct evidence on two margins central to organizational theory.

The second contribution is to make implementation capacity explicit in the joint design of authority and monitoring. We bring together research on delegation and managerial autonomy (Acemoglu et al., 2007; Bloom et al., 2012; Aghion et al., 2021; Kala, 2024) with work on performance measurement and monitoring (Holmström and Milgrom, 1991; Rasul and Rogger, 2018). Our framework emphasizes the capacity required to understand and prepare to execute a governance arrangement, drawing on a tradition that treats acquiring and using information as resource-consuming activities

(Simon, 1947; Arrow, 1974; Jensen and Meckling, 1992; Garicano, 2000; Tirole, 2009; Bolton and Faure-Grimaud, 2010; Dessein et al., 2016).[1] A shared capacity budget links these implementation demands to execution, making the relative performance and interaction of authority and monitoring depend on available capacity. This mechanism provides a capacity-based source of interdependence among organizational choices (Milgrom and Roberts, 1990; Brynjolfsson and Milgrom, 2013). Our experiment tests these predictions in a setting where governance requirements must be carried out by independent firms.

The third contribution is to the study of private regulation, where firms increasingly set and enforce standards throughout their supply chains (Kitzmüeller and Shimshack, 2012; Locke, 2013; Boudreau et al., 2023). Existing work examines how buyers use sourcing, auditing, disclosure, and related governance arrangements to improve supplier compliance and make performance observable (Guo et al., 2016; Plambeck and Taylor, 2016; Kalkanci and Plambeck, 2020; Short et al., 2016). We complement this literature by examining how the structure of reporting and verification shapes suppliers' ability to respond. This implementation perspective complements evidence linking supplier capabilities to compliance (Distelhorst et al., 2017; Boudreau, 2024) and broader work on management practices, organizational barriers, and firm upgrading (Bloom et al., 2013; Atkin et al., 2017; McKenzie and Woodruff, 2017; Bruhn et al., 2018; Verhoogen, 2023). It also parallels evidence that informational complexity and application costs affect participation in public programs (Bhargava and Manoli, 2015; Deshpande and Li, 2019; Finkelstein and Notowidigdo, 2019). The design of private regulation therefore depends in part on the organizational demands a standard places on the firms expected to meet it.

[1] For a recent review of the cognitive foundations of decision-making, organized around the acquisition, interpretation, and aggregation of information, see Enke (2026).

The paper proceeds as follows. Sections 2 and 3 describe the setting and experimental design. Section 4 reports the main results, and Section 5 develops the capacity model. Sections 6 and 7 examine mechanisms, performance, and business outcomes. Sections 8 and 9 discuss implications and conclude.

2 Institutional Context

2.1 The Lead Firm

The lead firm in this study is a leading Chinese apparel brand with an outsourced supplier network serving domestic and international markets. It holds top-tier international ESG ratings for 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.[2] The firm's objective was accordingly to move beyond surface-level compliance toward genuine accountability. This required establishing ESG standards and governance that encouraged substantive participation among resource-constrained independent suppliers.

In this setting, comprehensive and verifiable supplier disclosure is itself a valuable ESG outcome. More complete supplier reporting strengthens the information available

[2] More than 270 buyers worldwide used CDP's Supply Chain program to request environmental information from approximately 45,000 suppliers in 2025 (CDP, 2026).

for third-party ESG assessment (Drempetic et al., 2020). It also gives the firm visibility into upstream practices, helping it identify problems, coordinate improvements, and track progress (Bernstein, 2015; Christensen et al., 2021). More broadly, disclosure and verification programs can improve the outcomes they cover, from restaurant hygiene to industrial emissions (Jin and Leslie, 2003; Doshi et al., 2013; Duflo et al., 2013; Chen et al., 2018; Tomar, 2023). Verifiable reporting can further serve as a credible signal of commitment when suppliers with stronger practices find it easier to substantiate their claims (Spence, 1973; Clarkson et al., 2008).

2.2 The Supplier Network

The firm's supplier network comprises 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 accessories.

For this experiment, the firm introduced ESG management across its 155 active core suppliers. As Appendix Figure A1 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 Supply Chain Management platform, which serves as the standard interface for routine supplier reporting. Integrating the experiment into an existing corporate system ensured that suppliers perceived the ESG requirements as regular policy updates. This naturalistic setting limits experimenter demand effects that can arise when subjects know they are being observed (Harrison and List, 2004).

On login, suppliers are presented with a Welcome Page outlining the current reporting requirements (Appendix A). Its format is common across users, but its text varies by treatment. This page is the primary mechanism for introducing the experimental variations, informing suppliers whether they must report fixed indicators or choose their own, and whether verification is comprehensive or targeted to carbon.

Once suppliers click "Start Filing" on the Welcome Page, they are directed to the main reporting dashboard shown in Appendix Figure A2. 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 B and summarized in Table A2. 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.

The lead firm explained these requirements through the platform and direct contact with suppliers. The Welcome Page announced awards for top performers. After baseline

data collection, the 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, asked suppliers to report broadly and substantiate their entries, and explained that weak ESG performance could affect future orders. These communications conveyed the same buyer preferences across experimental arms. Each supplier learned its own obligations without being told about the randomized design. The four treatment cells therefore differed mainly in who selected the performance measures and how widely entries were verified.[3]

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. Holding third-party verification fixed, this variation tests whether broader oversight improves reporting or imposes coordination costs that reduce engagement.

[3] The reporting data are consistent with this interpretation. Delegation increased completion more on the buyer's designated KPIs than on other ESG fields, and targeted carbon monitoring increased completion of social and governance fields outside its audit scope (Appendix Figure A3). Both patterns are consistent with suppliers responding to the buyer's preference for broad reporting.

3 Experimental Design

3.1 The Factorial Design

We independently vary two governance dimensions in a randomized field experiment (Figure 1). The first dimension is the vertical allocation of authority. Under *Mandate* the lead firm specifies a fixed set of eight ESG KPIs as the minimum disclosure requirement. Under *Delegation* the supplier chooses eight indicators from a broader menu of 25 environmental, 48 social, and 5 governance indicators to meet that requirement.[4] 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 briefings in the same format, and compete for the same awards. Operationally, the five arms are implemented through distinct platform interfaces (Figure 1). The Welcome Page explains the assigned KPI-selection rule and monitoring scope, and the Data Reporting Page implements the assigned selection rule through required filing fields. Control suppliers face the baseline arrangement, with voluntary ESG filing and no assigned KPI-selection rule or verification structure. Comparisons with control identify the effect of introducing a governance structure, while comparisons across treatment cells identify the effects of authority allocation and monitoring scope.

[4] Delegation here refers specifically to supplier choice over the indicators satisfying the minimum ESG reporting requirement, not authority over the supplier's underlying business decisions.

3.2 Randomization and Sample Balance

We stratified the 155 suppliers by type and size, proxied by historical production value, and randomly assigned them across five conditions. Stratification improves precision and balances pre-existing 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 $p = 0.98$. The groups are therefore well balanced on the observed pre-treatment characteristics, 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 onward. The output, revenue, employment, and sales reported in this round pertain to fiscal year 2024 and provide the pre-program benchmark for the business outcomes in Section 7.2.

The experiment continued with four quarterly reporting rounds covering fiscal

year 2025, from April 2025 through March 2026.[5] Collection of quarterly financial data through the platform generally began one month after each quarter ended and was completed about three months after quarter-end, allowing time for late-arriving invoices. Each supplier's Welcome Page and subsequent communications specified its reporting requirements and 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 briefings reiterated each supplier's reporting requirements and verification scope without revealing the broader experimental design. 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. Section 6.5 summarizes descriptive patterns from the briefing transcripts, audit reports, and endline survey.

3.4 Outcome Measures and Estimation Sample

We measure supplier engagement in reporting on two margins. The first is *completeness*, measured by the total number and share of applicable platform fields a supplier fills in.[6] Control suppliers complete about 77 applicable fields without a minimum ESG reporting requirement, compared with about 89 among treated suppliers. All treated suppliers complete the eight required ESG indicators, whether chosen by the buyer or the supplier, as required for platform submission. Completeness is therefore high

[5] 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.

[6] The applicable set includes eighteen fields that the platform computes from other entries. Recomputing both completeness measures using the remaining 90 annual and 138 quarterly entries yields estimated effects of 9.8 percentage points for mandate and 10.5 percentage points for targeted monitoring, close to the corresponding estimates of 10.5 and 10.3 percentage points in the main specification.

in both groups and largely reflects voluntary reporting rather than compliance with assigned requirements.

The second is *documentation*, the breadth and verifiability of the supporting material a supplier uploads. Of the 108 platform metrics, seventy-five accept attachments, and we count how many a supplier supports with at least one file. We also count the ESG categories covered by attachments (out of twelve) and those backed by hard evidence. The indicator count captures the extent of documentation, but differences in correlation across indicators make it a noisy measure of evidentiary breadth. We therefore place more weight on the two category-level measures. Category coverage distinguishes documentation across topics from additional files within an already documented topic, while hard evidence distinguishes verifiable records from descriptive material.[7] Appendix C details the attachment measures, and Figure A4 summarizes completeness and documentation across filing rounds.

Supplementary measures support the capacity tests and the analysis of performance and business outcomes. Pre-treatment registered capital and reporting roles capture supplier resources and filer responsibilities. The endline survey, briefing transcripts, and audit reports provide descriptive evidence on reporting capacity and barriers. Blinded buyer ratings assess transparency and ESG performance. Business outcomes cover revenue, employment, worker turnover, on-time delivery, product returns, and exit from the buyer's network.

For the main engagement analysis, the platform estimation sample comprises 735 supplier-round observations from 152 suppliers with valid treatment assignments.[8]

[7] The main results show that gains in completeness are accompanied by wider documentation coverage and more categories backed by hard evidence. The audit reports also show no statistically detectable differences between mandate and delegation in carbon-data reliability or social-record auditability within each monitoring scope. These patterns alleviate the concern that engagement gains reflect only unsupported entries, although they do not establish the accuracy of individual unaudited fields.
[8] Three suppliers from different treatment cells never entered data on the platform. They are excluded from the submitted-report sample and included in the full-panel sensitivity analysis in Table A8.

Supplier participation varies across reporting rounds because of seasonal production cycles and factory operations.[9] Section 4.4 reports sensitivity to reporting coverage, and Section 7.2 analyzes supplier exit separately.

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 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. Each coefficient summarizes the effect of its governance arm relative to voluntary reporting, pooling over the experimental assignments on the other dimension. 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 without imposing additivity, we replace the pooled indicators with four treatment-cell indicators (Muralidharan et al., 2025).

[9] Five rounds for 152 suppliers yield 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 missing rounds are concentrated 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.

$$
\begin{aligned}
Y_{ist} = \alpha &+ \beta_{MC}\,\mathbf{1}\{V_i = M, H_i = C\} + \beta_{MT}\,\mathbf{1}\{V_i = M, H_i = T\} \\
&+ \beta_{DC}\,\mathbf{1}\{V_i = D, H_i = C\} + \beta_{DT}\,\mathbf{1}\{V_i = D, H_i = T\} \\
&+ \delta_s + \lambda_t + \varepsilon_{ist}.
\end{aligned} \tag{2}
$$

Here M and D denote Mandate and Delegation, and C and T denote Comprehensive and Targeted Monitoring. All four indicators equal zero for control suppliers. These cell coefficients also provide the compounding test in the corollary. The contrast $\beta_{MT} + \beta_{DC} - \beta_{MC} - \beta_{DT}$ compares the gain from simplifying both dimensions with the sum of the gains from simplifying each separately.

We next examine heterogeneity across reporting tasks, supplier capacity, and respondent roles (Implications (i)–(iii)). For reporting tasks, we estimate the pooled specifications separately by ESG domain, using linear probability models for attachment submission. For supplier capacity and respondent role, we interact the pooled treatment indicators with group indicators and include the group main effects. Each treatment effect is therefore measured against control suppliers in the same group. We retain stratum and round fixed effects, clustering by supplier.

We use the same pooled treatment indicators to examine endline ESG ratings and business outcomes, including supplier exit. These cross-sectional OLS regressions include stratum fixed effects, with standard errors clustered by stratum. Financial and operational outcomes with baseline measures control for their fiscal-year 2024 values, with revenue and employment in logs.

4 Main Results

This section reports governance effects on completeness and documentation measures defined in Section 3. Three results emerge. Mandate and targeted carbon monitoring

raise engagement, while delegation and comprehensive monitoring produce smaller estimated responses. The largest point estimates arise under mandate combined with targeted carbon monitoring, and the main patterns persist across specifications.

4.1 Mandate versus Delegation

We begin with the vertical allocation of authority. Figure 2 presents the treatment effects relative to the control group, and Table A3 reports the corresponding regressions. Mandate raises the number of completed indicators by 16.037 and the completion rate by 10.5 percentage points, a 19.7 percent increase relative to the control mean of 53.5 percent. Delegation raises the completion rate by 4.5 percentage points, but the estimate is not statistically distinguishable from zero. The two arms differ at the ten percent level on both completeness outcomes.

The documentation outcomes reinforce this pattern. Mandate increases the number of documented indicators by 1.075, attachment coverage by 0.958 categories, and hard-evidence coverage by 0.766 categories. The attachment-coverage effect represents a 50.2 percent increase relative to the control mean of 1.907 categories. The attachment-coverage and hard-evidence estimates are statistically significant at the five percent level, while the documented-indicator estimate is imprecise. Delegation produces small and imprecise estimates on all three documentation outcomes. The two arms differ at the five percent level on attachment coverage and hard evidence.

4.2 Targeted versus Comprehensive Monitoring

We turn to the horizontal scope of monitoring. Figure 3 presents the treatment effects relative to the control group, and Table A4 reports the corresponding regressions. Targeted carbon monitoring raises the number of completed indicators by 15.622 and

the completion rate by 10.3 percentage points, a 19.2 percent increase relative to the control mean of 53.5 percent. Both completeness estimates are significant at the one percent level. Comprehensive monitoring raises the completion rate by 4.7 percentage points, but the estimate is imprecise. The two monitoring arms differ at the ten percent level on both completeness outcomes.

The documentation outcomes show a similar pattern. Targeted carbon monitoring increases attachment coverage by 0.806 categories, a 42.3 percent increase relative to the control mean, and hard-evidence coverage by 0.615 categories. Both estimates are significant at the five percent level. The number of documented indicators rises by 1.075, but the estimate is not statistically distinguishable from zero, with a standard error of 1.057. Comprehensive monitoring increases attachment coverage by 0.227 categories and produces imprecise estimates on all three documentation outcomes. The two monitoring arms are not statistically distinguishable on documentation. As in the authority comparison, the gains under targeted monitoring extend to both completeness and documentation, while comprehensive monitoring produces smaller and less precise estimated responses.

4.3 The Combined Treatment

The two dimensions cross in the factorial design, allowing us to compare the four treatment cells directly. Figure 4 plots the cell-level estimates, and Table A5 reports the corresponding regressions. Mandate with targeted monitoring produces the largest point estimates on all five outcomes. It raises the completion rate by 11.4 percentage points, the number of documented indicators by 1.521, attachment coverage by 1.240 categories, and hard-evidence coverage by 0.961 categories. Mandate with comprehensive monitoring raises the completion rate by 9.6 percentage points and

attachment coverage by 0.695 categories. Delegation with targeted monitoring raises these outcomes by 9.2 percentage points and 0.405 categories, respectively. Both cells generate completeness gains close to those under mandate with targeted monitoring, but smaller documentation gains. Delegation with comprehensive monitoring produces the smallest point estimates, with little change on either margin.

These patterns suggest that completeness and documentation place different demands on supplier capacity. Filling in fields may require less work than assembling supporting evidence, so simplifying either indicator selection or monitoring scope may release enough capacity to increase completion. Documentation requires additional record collection and coordination, which may explain why statistically significant gains in attachment coverage and hard evidence appear only under mandate with targeted monitoring.

Mandate with targeted monitoring and delegation with comprehensive monitoring differ significantly on four of the five outcomes (Table A5). The estimates for mandate with targeted monitoring also remain essentially unchanged under checks for influential suppliers and alternative adjustments.[10] The design is powered primarily to detect the main effects, so we treat the overall ordering of the four cells as descriptive.

4.4 Robustness

The main regressions cluster standard errors by supplier, the unit of treatment assignment. Tables A6 and A7 report stratum-clustered standard errors to allow for correlation within strata and wild-cluster bootstrap inference with thirteen strata. We also average outcomes to the supplier level to check sensitivity to reporting

[10] This cell contains the network's largest suppliers by baseline production, so we checked whether they drive its estimates. 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.

frequency. The estimates remain stable across these alternatives. The completion effects of mandate and targeted monitoring remain near 10.4 and 10.2 percentage points, respectively, while delegation and comprehensive monitoring produce smaller estimates throughout. The completion and attachment-coverage effects of mandate and targeted monitoring remain significant at the five percent level or better.

Table A8 checks sensitivity to the definition of completeness and reporting coverage. Column 1 excludes the 30 financial and certification fields. Column 2 reduces the weight of repeated monthly entries by averaging them within each environmental indicator before calculating completion. Column 3 removes the mechanical contribution of required indicators by deducting eight completed and eight applicable entries from treated reports. Mandate and targeted monitoring retain larger estimated gains and remain significant relative to control. Column 4 adds observed unsubmitted drafts, while Column 5 retains all 155 suppliers over five rounds and assigns records with no filled entries the same-round control minimum. The ordering of point estimates persists, although the difference between monitoring arms is less precisely estimated in the full panel.

5 A Model of Organizational Capacity

Delegation can improve decisions when agents have useful local information and aligned incentives, and broad monitoring can sustain effort across tasks (Aghion and Tirole, 1997; Dessein, 2002; Holmström and Milgrom, 1991). These benefits can come with implementation costs, as understanding the requirements and organizing the work take resources (Simon, 1947; Arrow, 1974). In our model, comprehension draws on the same capacity budget as execution, leaving fewer resources to carry out the contract. We characterize when this implementation burden favors simpler structures and when

the benefits of delegation and broad monitoring prevail. Appendix D contains the formal results, proofs, and extensions.

5.1 Setup

A lead firm transacts with suppliers whose organizational capacity k is distributed according to a continuous distribution F on $[0, \bar{k}]$. Capacity is the managerial and staff time, records infrastructure, and coordination ability available for the reporting program. The buyer chooses $c = (V, H)$, where $V \in \{M, D\}$ denotes mandated or delegated indicator selection and $H \in \{C, T\}$ denotes comprehensive or targeted verification. Processing the governance structure consumes capacity

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

Here δ is the cost of evaluating the indicator menu and μ is the cost of mapping a broad verification scope onto internal records and functions. These costs accompany gains in execution productivity. Effort e produces engagement $\phi(c)g(e)$, where

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

A supplier that has worked through a richer structure converts effort into reported and documented content at a higher rate. We assume g is continuous at zero and twice continuously differentiable for positive effort, with $g(0) = 0$, $g' > 0$, $g'' < 0$, and $g(e) \to \infty$ as $e \to \infty$. Engagement covers both completeness and documentation, which draw on the same capacity.

Each unit of engagement is worth $w(c) > 0$ to the buyer. The supplier values the buyer's assessment because it can affect future business (Baker et al., 2002;

Macchiavello and Morjaria, 2015).[11] Its stake in the relationship is $\alpha \in (0, 1]$, and capacity used in the core business earns $\rho > 0$ per unit. In the baseline, participation includes processing before effort is chosen. Appendix D.2 allows suppliers to decline processing. A supplier with $k \geq r(c)$ solves

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

We focus on suppliers with binding capacity, $\alpha w(c)\phi(c)g'(k - r(c)) > \rho$ for every c and $k \in (r(c), \bar{k}]$. They devote all residual capacity to the program, so

$$E(k; c) = \phi(c)g\big(\max\{k - r(c), 0\}\big). \tag{6}$$

The buyer chooses c to maximize $\Pi(c) = \int_0^{\bar{k}} w(c)E(k; c)\, dF(k)$. Figure 5, Panel A, shows the engagement curve. The entry point is the comprehension cost, and engagement rises concavely thereafter.

5.2 Implementation Costs and Supplier Capacity

Delegation raises the value of the report when the supplier's interests are sufficiently aligned with the buyer's. We maintain this aligned case, $w(D, H) > w(M, H)$, and assume $w(V, C) > w(V, T)$ because broad verification preserves the value of material outside the focal task. Appendix D.4 sets out the selection comparison. In the benchmark without additional comprehension costs, $\delta = \mu = 0$, all structures leave the same residual capacity. Delegation and comprehensive verification then weakly maximize engagement and maximize buyer value whenever reporting occurs

[11] Future orders can reward engagement even when a supplier's overall contribution cannot be fully specified in an enforceable contract. Our model takes this relational incentive as given and examines the capacity required to understand and carry out the buyer's requirements.

(Proposition D.1).

With positive costs and productivity gains, $\delta, \mu, \pi_D, \pi_C > 0$, a more complex structure starts later and climbs faster. Compare a simple structure s with a complex structure x satisfying $r(x) > r(s)$ and $\phi(x) > \phi(s)$. For $k \geq r(x)$, where both arrangements are feasible, concavity makes $E(k; x) - E(k; s)$ strictly increasing. The gap starts below zero and crosses once if it becomes positive within the support, at

$$\phi(x)g(k^* - r(x)) = \phi(s)g(k^* - r(s)).$$

Figure 5, Panel B, shows this crossing. Below k^* the simple structure's head start dominates. Above it the complex structure's productivity gain dominates. The threshold rises with the complex structure's comprehension cost and falls with its productivity. When $w(x) \geq w(s)$, the buyer's value ranking reverses weakly earlier (Proposition D.2).

These rankings translate into arm averages through the capacity distribution. In a network below the relevant crossings, mandate and targeted verification each raise engagement. Their combination ranks first and delegation with comprehensive verification ranks last, provided enough suppliers can enter the intermediate cells for the rankings to be strict (Proposition D.3). The two intermediate cells remain unranked. Appendix Figure A5, Panel A, places all four structures on the same capacity axis.

5.3 Governance Rankings and Heterogeneity

The central result gives three linked comparisons for a network below the relevant capacity crossings, with some suppliers able to implement the intermediate arrangements. Proposition D.3 in Appendix D gives the formal conditions and proof.

Proposition (Governance under limited capacity)**.**

(i) Mandate raises engagement relative to delegation within each monitoring scope.

(ii) Targeted verification raises engagement relative to comprehensive verification within each authority allocation.

(iii) Mandate with targeted verification generates the highest engagement, and delegation with comprehensive verification the lowest. The two intermediate arrangements lie between them and need not be ranked against each other.

Concavity means that a given processing burden costs more engagement once another burden has already reduced residual capacity.

Corollary (Compounding burdens)**.** *For suppliers able to implement all four arrangements, the gains from simplifying authority and monitoring separately, measured from delegation with comprehensive verification, sum to more than the gain from simplifying both together.*

Further implications for heterogeneity. The model also gives conditional implications for differences across tasks, suppliers, and filers.

(i) *Task tractability.* When information is valued equally across tasks, added capacity brings more tractable tasks into reporting first. The size of the response on tasks already reported depends on their execution technology.

(ii) *Supplier capacity.* The relative gains from delegation and comprehensive verification increase with capacity once both structures are feasible. The baseline ranking reverses if capacity passes the relevant crossing.

(iii) *Reporting roles.* When competing demands leave the filer near the minimum capacity needed to start reporting, the contrast between governance structures shrinks toward zero.

Appendix D gives the formal derivations. Appendix Figure A5, Panels C–E, illustrates these implications.

6 Testing the Capacity Channel

The three parts of the proposition correspond to the main results. We now test compounding and heterogeneity across tasks, suppliers, and filers, then use the briefing transcripts, endline survey, and audit reports to provide supplementary, suggestive evidence on the capacity interpretation.

6.1 Compounding Burdens

The corollary concerns compounding. Both burdens draw on the same budget, so concavity makes a given burden costlier once the other is in place. We compare the three simplified cells with delegation and comprehensive monitoring, the fully complex cell. Moving to mandate removes the choice burden, moving to targeted monitoring removes the audit-scope burden, and combining them removes both. 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 6 and Table A9 report the three gains and the second difference for the five engagement outcomes. Relieving the choice burden alone adds 15.052 completed items and 9.941 percentage points of the completion share. Relieving the audit-scope burden alone adds 14.331 items and 9.487 percentage points. The two sum to 29.383

items and 19.428 percentage points. Relieving both together adds 17.743 items and 11.752 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 categories of hard evidence against a sum of 1.363. The second difference is negative for all five outcomes, although the estimates are imprecise. Moreover, the outcomes come from the same filings, so their common sign does not constitute five independent pieces of evidence. We therefore interpret the pattern as suggestive evidence of compounding without drawing conclusions about its magnitude.

Two mechanisms can produce this pattern. First, a given withdrawal from the shared capacity budget reduces engagement more when another burden has already depleted that budget. Concavity alone can therefore generate a negative second difference. Second, delegated choice under comprehensive verification adds a joint planning burden, since suppliers deciding what to report must consider both the availability of records and the evidence required under the assigned verification scope. Both are consistent with what we find and suggest that the gain from simplifying one governance margin may be smaller once the other has already been simplified.

6.2 Heterogeneity by Task Tractability

We examine Implication (i) by comparing treatment effects across ESG domains. Task tractability depends on data accessibility and the clarity of verification protocols. Environmental reporting often draws on quantitative records such as utility bills, metering data, and procurement documents. Social and governance reporting more often requires qualitative material or coordination across HR, legal, and compliance

functions.

Figure 7 and Table A10 report the domain-specific estimates. Mandate raises the environmental completion share by 14.5 percentage points and the probability of uploading environmental documentation by 28.1 percentage points. Targeted carbon monitoring raises these outcomes by 15.9 and 24.3 percentage points, respectively.

Environmental reporting responds consistently to both mandate and targeted monitoring, especially on the documentation margin. Its completion effects exceed the corresponding social effects and are similar in magnitude to the governance effects of 14.6 percentage points under mandate and 15.6 under targeted monitoring. Both environmental completion estimates are significant at the one percent level, compared with the ten and five percent levels for the respective governance estimates. The documentation effects are substantially larger for environmental reporting than for social or governance reporting. Both environmental documentation estimates are significant at the one percent level, while the corresponding social and governance estimates are not statistically distinguishable from zero.

These patterns are consistent with differences in task tractability, particularly in assembling supporting evidence. They align with the multitask insight that effort responds to differences in measurability (Holmström and Milgrom, 1991) and highlights the capacity needed to assemble verifiable evidence. Accessible records and clearer verification protocols may make it easier for suppliers to translate reporting effort into verifiable documentation.

6.3 Heterogeneity by Supplier Capacity

The benefits of delegation and comprehensive monitoring depend on suppliers' capacity to implement these arrangements. Choosing indicators and assembling evidence across

multiple domains draw on the same staff and reporting systems needed to complete the filing. The model therefore predicts that, once both simpler and more complex arrangements are feasible, the relative gains from complexity increase with capacity (Implication (ii)).

We proxy supplier capacity using registered capital from the buyer's pre-treatment supplier roster. This measure captures organizational scale and serves as an indicator of the administrative resources available for reporting.[12] We compare suppliers in the top quartile of registered capital with those below it, distinguishing the relatively larger suppliers in a network composed mainly of micro and small firms.[13] Each treatment effect is measured relative to control suppliers in the same capacity group.

Figure 8 and Panel A of Table A11 show a reversal in the authority ranking in the point estimates. Below the threshold, mandate raises completed indicators by 14.126 and attachment coverage by 1.132 categories, while delegation's effects are close to zero. Above the threshold, delegation raises all five outcomes, including completed indicators by 33.440, attachment coverage by 1.961 categories, and hard-evidence coverage by 1.577 categories. Its point estimates exceed mandate's on every outcome. The change in the mandate–delegation gap across capacity groups is significant at the ten percent level on four of the five outcomes, with documented indicators the exception.

The monitoring results in Figure 9 and Panel B of Table A11 show a different pattern. Targeted monitoring's advantage increases with capacity, but comprehensive

[12] Registered capital is filed on a common legal standard and is available for 150 of the 152 suppliers. It therefore provides a comparable and broadly observed measure. An alternative is paid-in capital, which records capital actually contributed to the firm and is available for 128 suppliers. The two measures have a rank correlation of 0.92 among suppliers reporting both.

[13] The upper quartile contains 38 of the 150 suppliers with reported capital. Assignment does not significantly predict membership in this group (= 0.20 p=0.20 for authority and = 0.36 p=0.36 for monitoring). Within this group, production, firm age, and registered capital do not differ significantly from control. Employment differs only for targeted monitoring at the ten percent level.

monitoring does not overtake it. Above the threshold, targeted monitoring raises all five outcomes at the one percent significance level, including completed indicators by 42.635 and attachment coverage by 2.615 categories. Comprehensive monitoring's effects are not statistically distinguishable from zero on any outcome. Below the threshold, neither arm differs significantly from control, and the two arms are statistically indistinguishable. The pooled gains from targeted monitoring are therefore concentrated among suppliers above the capital threshold.

Using registry employment to proxy capacity yields similar patterns (Appendix Figures A6 and A7). Alternative registered-capital thresholds at the top 50, 30, 25, and 20 percent of the sample preserve the authority reversal in point estimates, while comprehensive monitoring never overtakes targeted monitoring (Appendix Figures A8 and A9).

The authority reversal is consistent with the capacity crossing in Corollary D.3 falling within our sample. It connects our findings to research on management practices, decentralization, and organizational capability (Bloom et al., 2013; Aghion et al., 2021), and to work on the capabilities required for firm upgrading (Verhoogen, 2023). The monitoring results are consistent with suppliers crossing the lower entry barrier for targeted monitoring, while the threshold at which comprehensive monitoring becomes more beneficial may lie beyond the observed range. In sum, these patterns suggest that suppliers can have enough capacity to benefit from discretion while still gaining more from targeted monitoring.

6.4 Heterogeneity by Reporting Role

We test Implication (iii) by examining whether treatment responses vary with who handles reporting. Owners and executives have broad responsibilities that may limit

the capacity they can devote to filing. We identify reporting roles from the supplier's standing platform contact, recorded before the experiment. Roles are classified for 702 of the 735 supplier-rounds, and about one supplier in seven files through an owner, general manager, or director. Executive-filer shares do not differ significantly across the five cells.

Figures 10 and 11, with Table A12, report treatment effects relative to control suppliers with the same reporting role. Among staff filers, mandate raises completed indicators by 17.440 and the completion rate by 11.5 percentage points, while adding 1.236 attachment categories and 0.930 hard-evidence categories. Targeted monitoring raises completed indicators by 16.359 and the completion rate by 10.8 percentage points. Delegation and comprehensive monitoring produce smaller, imprecise estimates, and mandate exceeds delegation at the one percent level on both category measures. Among executive filers, no arm differs significantly from control on any outcome. The treatment effects are jointly significant among staff filers ($p = 0.018$) and jointly indistinguishable from zero among executive filers ($p = 0.454$).

The concentration of gains among staff filers is consistent with Implication (iii). Reporting capacity depends on the responsibilities of the filer as well as the firm's resources. This interpretation aligns with evidence on executive time allocation and firm performance (Bandiera et al., 2020) and with theories of scarce organizational attention (Dessein et al., 2016). Our findings suggest that the allocation of reporting responsibilities within suppliers can shape their response to the buyer's governance requirements.

6.5 Corroborating Qualitative Evidence

The briefing transcripts, endline survey, and audit reports described in Section 3.3 provide descriptive evidence on the capacity channel. The briefings yielded 150 recordings, from which we constructed an engagement index combining suppliers' speech share, substantive questions, and requests for written follow-up. Engagement was higher under mandate than delegation and under targeted than comprehensive monitoring by 0.18 and 0.23 standard deviations, respectively.

The endline survey yielded 135 matched supplier responses. Among respondents, 62 percent lacked a centralized data system, and 65 percent had no dedicated ESG personnel. Suppliers lacking either resource reported greater task burden and more barriers related to data, staffing, or time. These barriers were also cited more frequently under delegation and comprehensive monitoring. The third-party audit reports documented missing records and verification difficulties, consistent with limited reporting capacity. Suppliers providing less extensive documentation on the platform tended to have weaker audited record-keeping and lower environmental and social compliance scores. Detailed breakdowns are available from the authors upon request.

6.6 Alternative Mechanisms

The findings are consistent with a capacity channel, through which governance affects the resources suppliers can devote to reporting. Other mechanisms may contribute to these responses. We consider five alternatives and assess how they fit the observed patterns.

One alternative is that governance signals the lead firm's priorities. Governance choices can convey information as well as change incentives (Bénabou and Tirole, 2003). Mandate identifies the buyer's preferred indicators, while a carbon audit

highlights the outcome it will scrutinize. The platform and briefings described in Section 2.3 nevertheless conveyed a common preference for broad, substantiated reporting. Consistent with this message, delegation raises completion more on the buyer's designated KPIs than on other ESG fields, and targeted monitoring raises completion of unaudited social and governance fields (Appendix Figure A3). Moreover, if signaling is the main driver, we would expect larger treatment effects among suppliers with more at stake in the buyer relationship. Yet treatment effects do not differ significantly by the buyer's share of supplier sales within any governance arm (Appendix Figures A10 and A11).

A second explanation is clarity. Mandate and targeted monitoring clarify what suppliers must report and what the buyer will check. Clearer requirements can reduce the comprehension cost in equation (3), and related evidence shows that agents may overlook aspects of a task they have not learned to notice (Hanna et al., 2014). Clarity alone, however, should generate similar gains regardless of who completes the filing. Instead, the gains appear when dedicated staff handle reporting and fade when executives file personally (Section 6.4). Clarity is one benefit of simpler structures, but does not explain this respondent-role pattern on its own.

A third explanation is selective disclosure. Delegation may lead suppliers to concentrate on indicators for which records are readily available, leaving other fields unreported (Plambeck and Taylor, 2016; Marquis et al., 2016). This could contribute to delegation's smaller completeness gains. However, delegation's point estimates exceed mandate's among higher-capacity suppliers, with gains in completeness, documentation breadth, and hard evidence (Section 6.3). Selective disclosure alone does not readily explain this broader engagement. The benefits of choice emerge where firms have the capacity to use it.

A fourth explanation is defensive reporting. A broad audit may deter suppliers

from entering information they cannot fully defend (Plambeck and Taylor, 2016). This deterrence could prompt strategic avoidance of scrutiny and contribute to its smaller engagement effects. The survey, however, also points to execution constraints. Suppliers described unavailable data, missing staff, and the time needed to compile information they wanted to report (Section 6.5). These accounts support an explanation based on administrative burden beyond strategic withholding.

A fifth explanation is distrust from control. Restricting an agent's discretion can signal distrust and reduce effort (Falk and Kosfeld, 2006). This could explain weaker engagement under comprehensive monitoring relative to carbon monitoring. The authority comparison, however, points in the other direction. Mandate removes suppliers' choice of indicators, yet raises engagement more than delegation, with gains extending to voluntary fields and verifiable documentation (Section 4.1). This pattern weighs against distrust as a common explanation for both dimensions, while allowing different reactions to the two forms of control.

These alternatives may each explain part of the response and need not be mutually exclusive. No single alternative, however, readily accounts for the heterogeneity and survey evidence together. The capacity channel offers a coherent interpretation of these patterns alongside information and incentive channels.

7 Firm Performance and Business Outcomes

7.1 ESG Performance Rated by the Lead Firm

Do the reporting gains translate into higher buyer assessments? Platform entries alone cannot answer this, since suppliers choose which fields to report and use different units. We therefore use endline ESG ratings from the lead firm's supply chain team,

which was blinded to treatment assignment and assessed suppliers on a common basis. The ratings cover production and operations, program engagement, and reporting practices. Although the methodology is undisclosed, the scores draw on submitted information and the team's ongoing interactions with suppliers. They provide a buyer-side assessment distinct from platform entries and do not mechanically overlap with the governance measures.

Figure A12 and Table A13 report the two endline outcomes. Mandate raises the transparency score by 0.105 and the performance score by 0.083, while targeted carbon monitoring raises them by 0.104 and 0.102, respectively. Delegation and comprehensive monitoring yield smaller estimates. With only 155 suppliers, the individual coefficients are imprecisely estimated, so the ordering remains suggestive.

The treatments generating the strongest engagement also receive the highest blinded transparency and performance ratings. This pattern suggests that the reporting gains extend beyond platform completion to aspects of suppliers' ESG engagement and performance that the buyer recognizes in its broader assessment.

7.2 Financial and Operational Outcomes

The model does not predict the direction of ESG reporting's effect on production. A multitasking concern is that administrative requirements divert resources from production, employment, or commercial activity. We therefore examine whether the engagement gains coincide with displacement in suppliers' business activity or exit from the buyer network. Revenue and employment enter in logs, while the remaining outcomes enter as measured. Each outcome is a single post-program observation per supplier, and we cluster by randomization stratum.

Figures A13 and A14, together with Table A14, report annual revenue, total

employment, voluntary worker turnover, on-time delivery, returns, and network exit. None of the twenty-four treatment estimates is statistically distinguishable from zero. Mandate raises log revenue by 0.418 (standard error 0.382) and log employment by 0.072 (standard error 0.061), and changes the probability of exit by -0.012 (standard error 0.031).

We therefore detect no crowding out of the kind predicted by multitasking theory and documented when effort shifts toward newly measured tasks (Holmström and Milgrom, 1991; Hong et al., 2018). Although fixed supplier capacity implies that reporting draws resources from competing uses, benefits within the buyer-supplier relationship could offset these costs. Repeated transactions under incomplete contracts make the relationship's continuation value important for cooperation (Baker et al., 2002; Boudreau et al., 2023). Reporting engagement could signal a supplier's reliability and commitment to the relationship, encouraging the buyer to direct more orders toward engaged suppliers (Macchiavello and Morjaria, 2015; Bernstein, 2015). These estimates capture net commercial effects, so they cannot separate reporting costs from offsetting relationship benefits. Their standard errors also leave room for economically meaningful gains or losses.

8 Discussion

8.1 Implications for Organizational Theory

Our results generate three implications for organizational economics. First, the benefits of delegation depend on the supplier's capacity to use its authority. Models of authority show how delegation can improve decisions when the agent has relevant private information and its interests are sufficiently aligned with the principal's (Aghion and

Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008). Our heterogeneity results suggest that realizing these benefits also requires the capacity to exercise discretion. Suppliers with organizational slack can use that discretion to tailor reporting, while for lower-capacity suppliers the cost of evaluating unfamiliar indicators can dominate the gain from local tailoring. The model captures this tradeoff through the comprehension cost δ in (3) and the productivity gain π_D in (4), and Corollary D.3 resolves it monotonically in capacity. As capacity rises, the fixed comprehension cost absorbs a smaller share of it, while the productivity gain applies to a larger scale of execution.

The relevant capacity is a broader endowment than managerial attention. Attention constrains day-to-day decisions (Ocasio, 1997; Dessein et al., 2016), while capacity also includes dedicated personnel, centralized records, and structured processes for absorbing administrative work (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 work (KC and Terwiesch, 2009).

This result helps explain why the gains from managerial autonomy documented within firms need not extend to small independent suppliers (Kala, 2024). In our setting, suppliers must use limited administrative resources to evaluate unfamiliar indicators and decide what to report. These costs can outweigh the benefits of tailoring reporting to local information. Mandate's advantage among lower-capacity suppliers is therefore consistent with evidence that simplifying information and easing application burdens can raise take-up (Bhargava and Manoli, 2015; Bettinger et al., 2012). Applying within-firm recommendations to supplier relationships requires accounting not only for the information suppliers possess, but also for their capacity

to use the discretion they receive.

Second, the results highlight implementation capacity alongside the effort-allocation trade-offs of multitasking theory. Standard models emphasize effort substitution once an agent is already participating (Holmström and Milgrom, 1991). In our model, monitoring scope affects entry into reporting. Empirically, targeted monitoring generates broader engagement, and its advantage over comprehensive verification widens with supplier capacity. This pattern is consistent with limited responses under either scope near the entry margin, followed by gains from targeted monitoring as suppliers cross its lower entry threshold. Comprehensive verification may overtake it only at a higher capacity level beyond the observed range. 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, governance instruments can combine through a common resource constraint. Mandate and targeted monitoring operate on different margins but both preserve execution capacity. Their combination is the only cell that improves both completeness and documentation by a statistically significant margin. The joint design of incentives, decision rights, and coordination is a classical theme (Holmström and Milgrom, 1991, 1994; Athey and Roberts, 2001; Dessein et al., 2010). Our model also accounts for capacity spent processing governance requirements before allocating effort. A shared capacity budget and concave execution technology can make comprehension costs compound. The gain from simplifying either dimension is larger when the other burden remains in place and smaller after that burden has been removed. Section 6.1 reports negative second differences on every outcome, consistent with Corollary D.2.

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 needed for compliance.[14]

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 familiarity with reporting, making governance a lifecycle choice (Locke, 2013).

Second, the sequencing of requirements matters. Metrics with clear standards and accessible records are a natural entry point, allowing suppliers to 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. Easing either implementation burden can release enough capacity for relatively simple tasks. Tasks that require assembling records and coordinating across functions may need relief on both dimensions. Mandate with targeted monitoring provides clear priorities and a focused verification scope, preserving capacity for more demanding work.

The findings also inform private regulation in global value chains, where lead firms increasingly pass public disclosure and due-diligence requirements down to supplier

[14] Survey and interview evidence from Dutch SMEs also documents limited resources and fragmented information systems in responding to sustainability data requests (Niemann et al., 2025).

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).[15] 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

Four questions remain for further work. The first concerns the mapping from the treatments to theory. The buyer's existing platform and supplier relationships limited the contractual dimensions we could vary. Delegation concerns the choice of indicators that satisfy a fixed minimum reporting requirement, while monitoring changes both verification coverage and the composition of audited tasks. The estimates therefore identify the effects of these specific reporting arrangements. Future work could vary operating decision rights and information access under common reporting requirements, and independently vary audit coverage, signal precision, and rewards across tasks. This would allow closer tests of the information and incentive trade-offs in classical models (Aghion and Tirole, 1997; Holmström and Milgrom, 1991).

The second concerns the range over which the governance ranking applies. We estimate it in one apparel network in China, and Section 6.3 shows how it changes with supplier capacity within that network. Where the ranking reverses depends on the comprehension cost of a governance structure and the capacity suppliers bring to it. Both may differ across industries and with the maturity of reporting standards. An extension could treat comprehension costs as a design choice. Templates, training, or plainer terms could lower these costs and shift the point at which the ranking reverses.

[15] To limit this burden, the EU's 2026 reform of the Corporate Sustainability Reporting Directive provides for a cap on the information that reporting firms may require from suppliers with no more than 1,000 employees for their own reporting obligations (European Parliament and Council, 2026).

Buyers would then weigh investment in implementation support against simplifying the contract itself.

The third concerns the reach of the outcomes. Completeness and documentation measure what suppliers report and can substantiate, rather than emissions or labor conditions themselves. A longer panel could show whether the reporting gains under simpler contracts carry through to certification, sourcing decisions, and measured environmental and labor outcomes, and whether the governance ranking holds for those outcomes.

The fourth concerns the measurement of capacity. We proxy it using baseline firm characteristics, the role of the filer, and endline reports on staffing and data systems, each capturing part of the construct defined in the model. Time-use data and cross-functional coordination records could provide more direct measures of how capacity is used and help test whether governance builds it. If mandate leaves suppliers better able to handle delegation later, phased governance could serve as an investment in capacity (Section 8.2).

9 Conclusion

A growing share of world production is coordinated by contract across firm boundaries, where buyers face two questions a manager faces inside a firm. How much discretion should the agent hold, and how broadly should performance be verified? Randomizing both terms of a live supplier program, we find the strongest engagement under mandate and targeted monitoring. Mandating the reported indicators raised the share of applicable fields a supplier fills in by 10.5 percentage points, 19.7 percent of the control mean, and widened documented evidence from 1.9 to 2.9 categories. Delegating the same choice produced smaller, imprecisely estimated gains on both

margins. Targeted carbon verification produced gains of similar size to mandate, while comprehensive verification produced smaller, imprecisely estimated responses relative to voluntary reporting. The buyer's blinded assessors gave the simpler arms the highest point estimates, and we detected no statistically significant changes in the measured business outcomes.

A binding constraint on organizational capacity offers an interpretation of these patterns. Delegation and comprehensive monitoring absorb capacity in evaluating reporting choices and preparing for broader verification, while mandate and targeted monitoring leave more resources for execution. The heterogeneity results, survey, and telephone briefings all point the same way. The model weighs these implementation costs against the benefits of autonomy and broad measurement, showing how the governance ranking can change as supplier capacity grows. Effective governance therefore depends on the capabilities of the supplier network. As lead firms increasingly serve as private regulators beyond their legal boundaries (Gereffi et al., 2005; Boudreau et al., 2023), arrangements that clarify priorities, focus verification, and expand with supplier capability are 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.

Andrews, Isaiah and Daniel Barron, "The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents," *American Economic Review*, 2016, *106* (9), 2742–2759.

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.

Arrow, Kenneth J., *The Limits of Organization*, New York: W. W. Norton, 1974.

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.

Bénabou, Roland and Jean Tirole, "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 2003, *70* (3), 489–520.

Bernstein, Lisa, "Beyond Relational Contracts: Social Capital and Network Governance in Procurement Contracts," *Journal of Legal Analysis*, 2015, *7* (2), 561–621.

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.

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.

Casaburi, Lorenzo and Tristan Reed, "Using Individual-Level Randomized Treatment to Learn about Market Structure," *American Economic Journal: Applied Economics*, 2022, *14* (4), 58–90.

CDP, "Keeping Pace: Disclosure Data Factsheet 2025," January 28 2026. https://www.cdp.net/en/insights/keeping-pace-disclosure-data-factsheet-2025.

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.

Clarkson, Peter M., Yue Li, Gordon D. Richardson, and Florin P. Vasvari, "Revisiting the Relation between Environmental Performance and Environmental Disclosure: An Empirical Analysis," *Accounting, Organizations and Society*, 2008, *33* (4–5), 303–327.

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.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *Review of Economic Studies*, 1999, *66* (1), 199–217.

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.

Drempetic, Samuel, Christian Klein, and Bernhard Zwergel, "The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review," *Journal of Business Ethics*, 2020, *167* (2), 333–360.

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

Enke, Benjamin, "The Cognitive Turn in Behavioral Economics," 2026. Working paper, version dated May 26, 2026.

European Parliament and Council, "Directive (EU) 2026/470 of 24 February 2026 Amending Directives 2006/43/EC, 2013/34/EU, (EU) 2022/2464 and (EU) 2024/1760 as Regards Certain Corporate Sustainability Reporting Requirements and Certain Corporate Sustainability Due Diligence Requirements," *Official Journal of the European Union*, L, 2026/470, February 26 2026. https://eur-lex.europa.eu/eli/dir/2026/470/oj.

Falk, Armin and Michael Kosfeld, "The Hidden Costs of Control," *American Economic Review*, 2006, *96* (5), 1611–1630.

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* (10), 9010–9031.

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.

Jensen, Michael C. and William H. Meckling, "Specific and General Knowledge, and Organizational Structure," in Lars Werin and Hans Wijkander, eds., *Contract Economics*, Oxford: Blackwell, 1992, pp. 251–274.

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.

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.

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

Marquis, Christopher, Michael W. Toffel, and Yanhua Zhou, "Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing," *Organization Science*, 2016, *27* (2), 483–504.

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.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich, "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments," *Review of Economics and Statistics*, 2025, *107* (3), 589–604.

Niemann, Ludger, Sebastiaan Morssinkhof, Martijn Jeroen van der Linden, and Karl de Vries, "Data Requests in Value Chains: The Effects of Corporate Sustainability Reporting on SMEs in The Netherlands," *Sustainability*, 2025, *17* (17), 8029.

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.

Simon, Herbert A., *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, New York: Macmillan, 1947.

Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

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

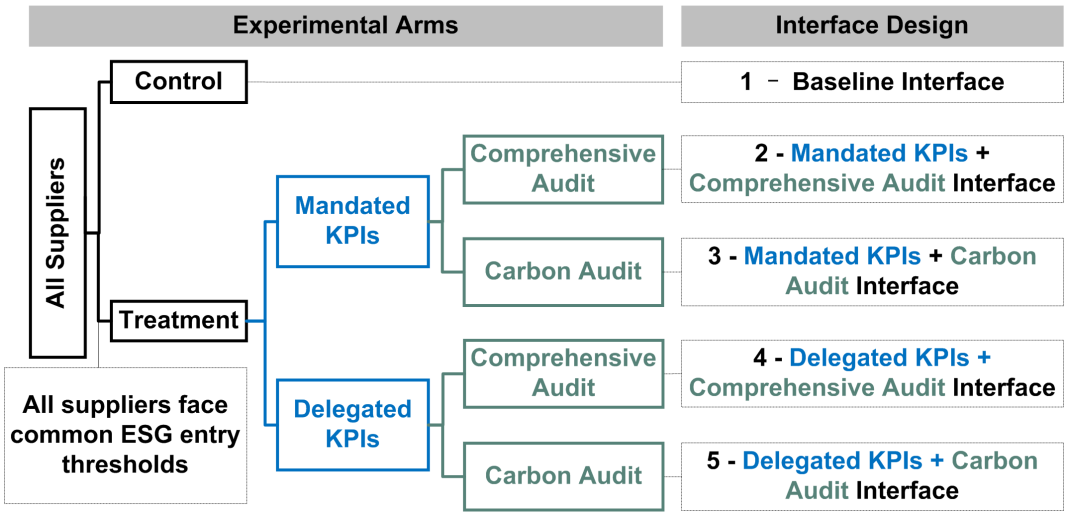


Figure 1: Interface Design by Experimental Arm

Notes: This figure maps the five experimental arms to their platform interfaces. All suppliers face common baseline requirements covering financial and certification indicators. Treatment suppliers are cross-randomized between mandated and delegated KPI selection and between comprehensive and targeted carbon audits, yielding four treatment cells alongside a control group. Interface 1 (Baseline Interface) is assigned to control suppliers, for whom ESG filing remains voluntary. Interfaces 2–5 correspond to the four treatment cells. The assigned interface determines the instructions on the Welcome Page and the required fields on the Data Reporting Page. Group-specific screenshots and English translations are reproduced in Appendix A.

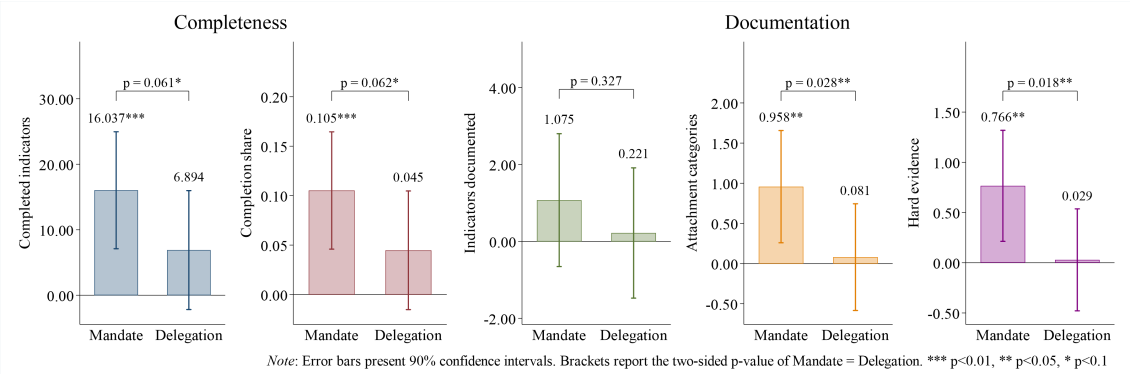


Figure 2: Authority Allocation and the Cost of KPI Choice

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. 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 A3.

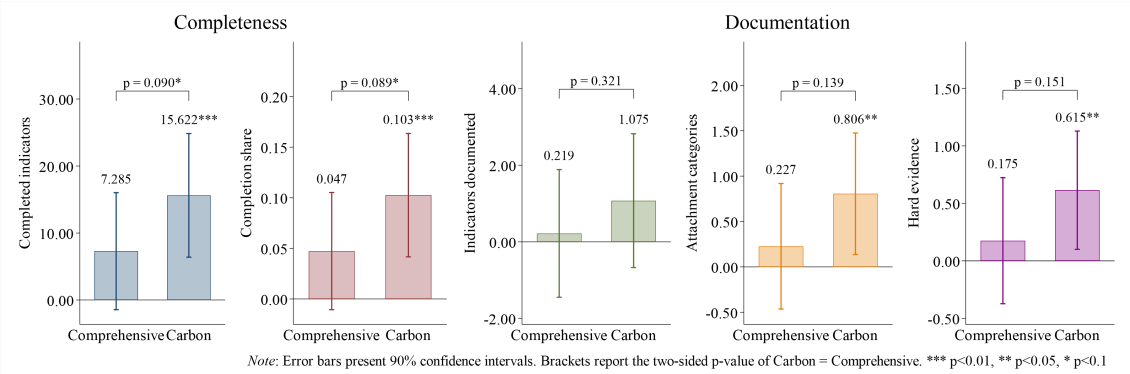


Figure 3: Monitoring Scope and the Value of Focus

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. 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 A4.

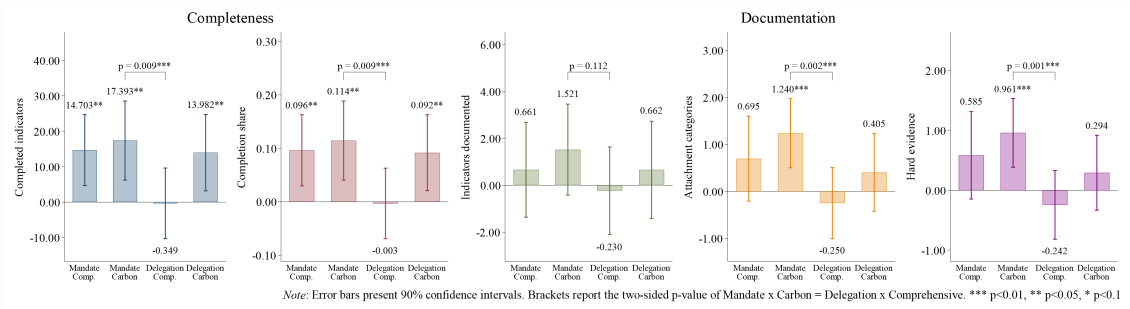
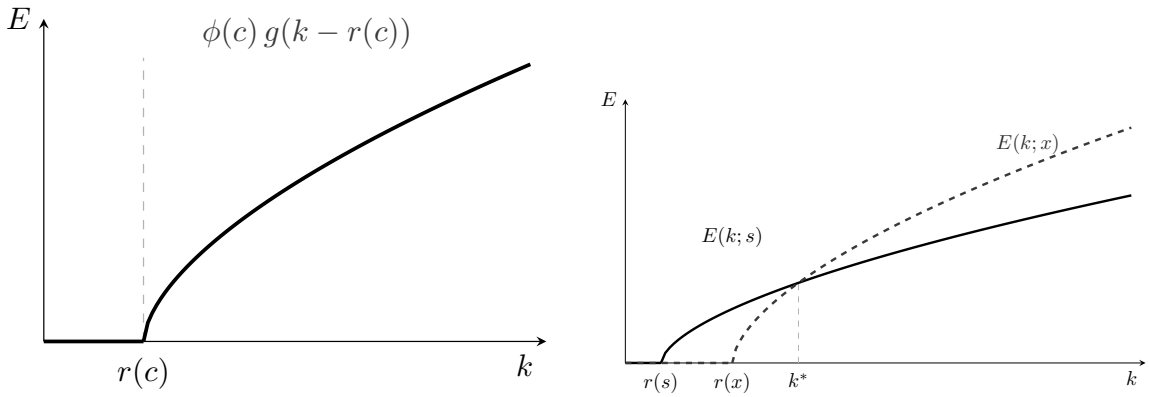


Figure 4: Engagement Across the Four Governance Bundles

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. Each regression includes reporting-round and randomization-stratum fixed effects, and standard errors are clustered by supplier. 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 of equality between the fully simplified and the fully complex cell, the statistic reported in the final row of Table A5. Table A5 reports the full set of regression statistics.



Panel A: The Engagement Curve *Panel B: The Single Crossing*

Figure 5: Organizational Capacity and Supplier Engagement

Notes: This figure shows how comprehension costs and execution productivity shape supplier engagement. Panel A shows equation (6). Engagement is zero until capacity covers the comprehension cost $r(c)$ and rises concavely thereafter. Higher overhead shifts the curve rightward, while higher productivity $\phi(c)$ raises and steepens it without changing the entry point. Panel B compares a simple structure s (solid) with a complex structure x (dashed). The latter has higher overhead and productivity. The curves cross at most once, and an interior crossing exists when the complex structure overtakes the simple structure within the capacity range. The simple structure delivers more engagement below k^* and the complex structure above it. The crossing rises with the complex structure's overhead and falls with its productivity.

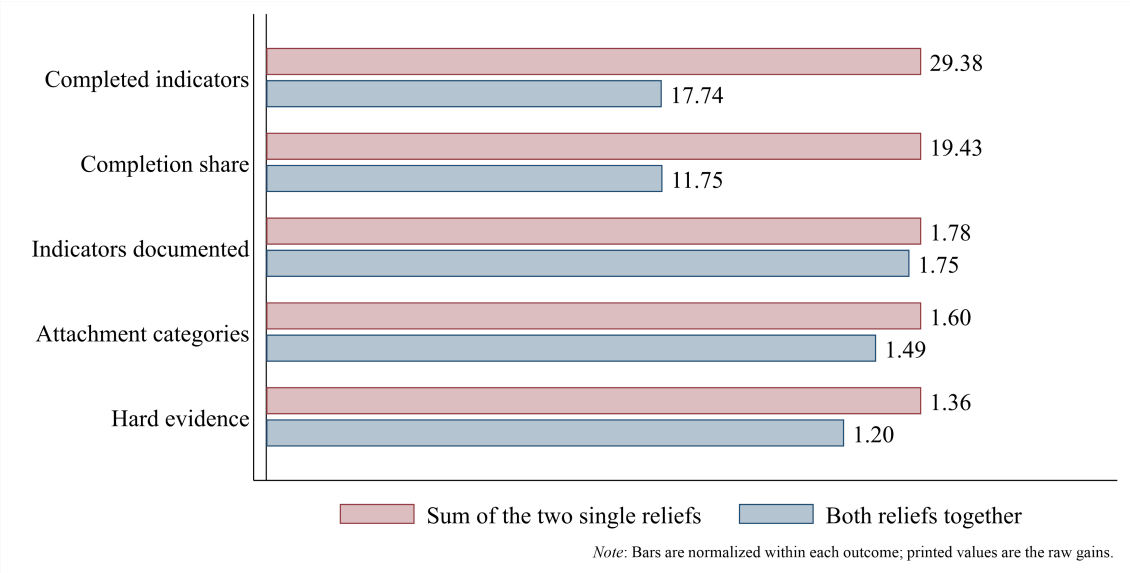


Figure 6: Compounding of the Two Reporting Burdens

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. 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 A5, 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 A9 reports the underlying gains, the second differences, their standard errors, and their p-values.

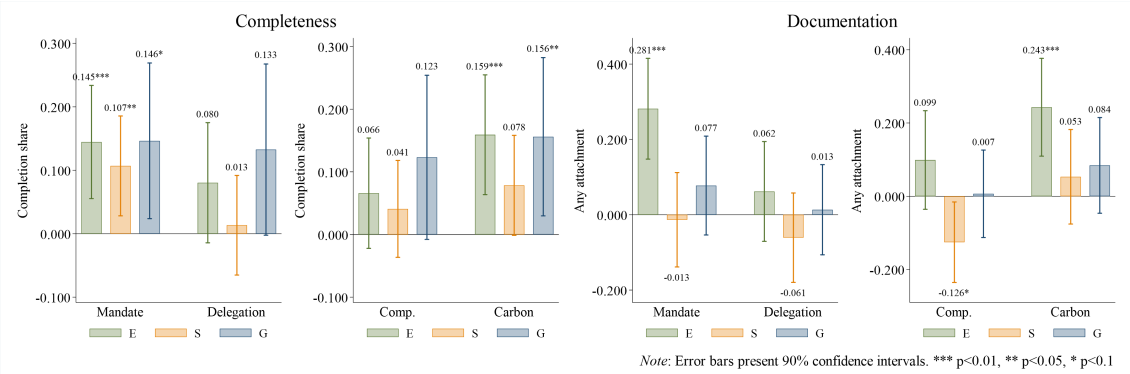


Figure 7: Engagement Responses Across ESG Dimensions

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 A10 provides the corresponding numerical estimates.

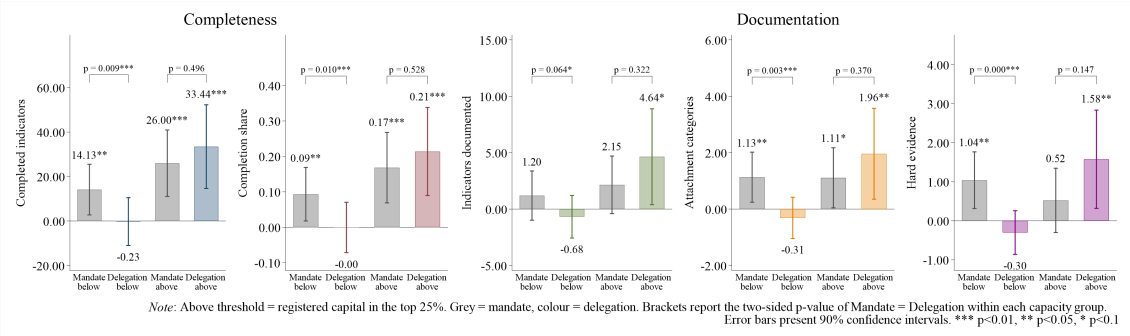


Figure 8: Supplier Capacity and the Returns to Authority Allocation

Notes: This figure reports the effects of Mandate and Delegation below and above a capacity threshold defined by the top quartile of registered capital (38 of 150 suppliers with reported capital: 9 control, 18 mandate, and 11 delegation). Panel A of Table A11 reports the corresponding regressions and control means. The five columns report completed indicators, the completion share, indicators documented, of 75, attachment categories covered, of 12, and categories backed by hard evidence. Regressions include arm-by-capacity-group interactions, a capacity-group indicator, and reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Each coefficient measures an arm's effect relative to controls in the same capacity group. Bars are grouped by capacity and shaded by arm; brackets report two-sided p-values testing equality of the Mandate and Delegation coefficients within each group. Delegation's estimates are small and imprecise below the threshold, but positive on all five outcomes and larger than Mandate's above it. The change in the Mandate-minus-Delegation gap across groups has the predicted sign on all five outcomes and is significant at the ten percent level on four.

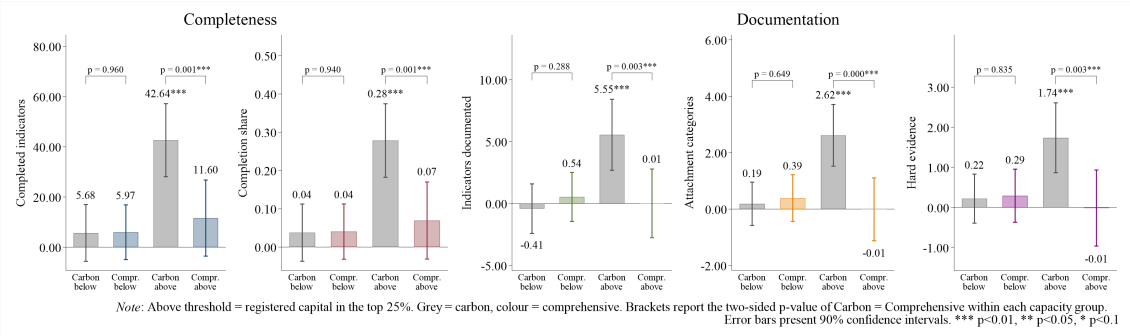
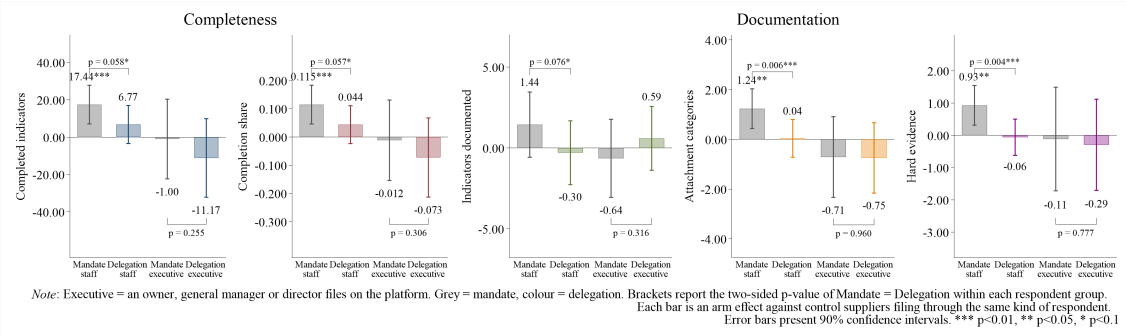


Figure 9: Supplier Capacity and the Returns to Monitoring Scope

Notes: This figure reports the effects of Targeted Carbon Monitoring and Comprehensive Monitoring below and above the same capacity threshold used in Figure 8: the top quartile of registered capital. Panel B of Table A11 reports the corresponding regressions and control means. 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 regression specification and bar layout follow Figure 8, with each coefficient measured relative to controls in the same capacity group. Brackets report two-sided p-values testing equality of the Targeted Carbon and Comprehensive Monitoring coefficients within each group. Regressions include reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. The advantage of targeted over comprehensive monitoring is larger above the threshold on all five outcomes. Below the threshold the two arms are statistically indistinguishable on every outcome.



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 10: Reporting Roles 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. The regressions interact both arms with respondent-group indicators and include the executive-filer indicator on its own. Control suppliers are the omitted category within each respondent group, so 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.

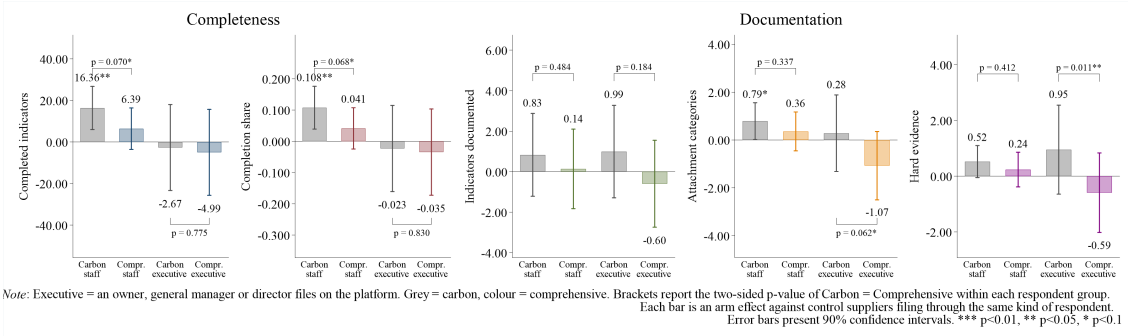


Figure 11: Reporting Roles and the Returns to Monitoring Scope

Notes: This figure repeats the design of Figure 10 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 10, 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.

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: This table presents pre-treatment characteristics for the control group and the four treatment cells compared in Section 4.3. Brackets contain p-values from two-sided t-tests against the control mean. Production (annual output value) and registered capital are in ten thousands of CNY. Employment is the registry-filed headcount summarized in Section 2.2. Missing roster items leave 137 to 154 observations per row. 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 all Panel A characteristics against control returns p-values of 0.43, 0.68, 0.94, and 0.87 for T1 through T4. A multinomial logit of arm assignment on the same characteristics returns $\chi^2(48) = 29.53$ with $p = 0.98$.

Appendix A. Platform Interface Screenshots

This appendix displays the instructions 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, branding has been removed from the panels below. The original instructions and emphasis are preserved.

Panel A. Interface 1: Baseline

尊敬的供应商伙伴，欢迎登录供应链ESG数据平台！

在25/26财年，本集团正式推出供应链ESG数据平台，旨在实现对供应商ESG表现的常态化管理，并满足组织、监管和评级机构的披露和审查要求。

平台使用说明:

1. 带“*”的指标为**必填指标**，请根据贵公司实际情况填写
2. 我们鼓励您积极填报其他ESG指标，共同完善ESG管理体系
3. 为激励优秀实践，本集团将在本财年末对ESG综合表现最优的前三名供应商给予行业领先的奖励

让我们携手共建绿色供应链，迈向可持续发展未来！

开始填报 →

(Original instructions; branding removed)

Dear Supplier Partner, Welcome to the Supply Chain ESG Data Platform!

In the 25/26 fiscal year, the Group officially launched the Supply Chain ESG Data Platform, aiming to achieve regular management of suppliers' ESG performance and to meet the disclosure and review requirements of organizations, regulators, and rating agencies.

Platform Guidelines:

1. Indicators marked with "*" are **mandatory**. Please complete them according to your company's actual situation.
2. We encourage you to provide additional ESG data wherever possible to help strengthen the overall ESG management system.
3. To motivate outstanding practices, the Group will grant **industry-leading awards** to the top three suppliers with the best overall ESG performance at the end of this fiscal year.

Let us work hand in hand to build a greener supply chain and move toward a sustainable future!

Start Filing

(English translation)

Panel B. Interface 2: Mandated KPIs, Comprehensive Audit

尊敬的供应商伙伴，欢迎登录供应链ESG数据平台！

在25/26财年，本集团正式推出供应链ESG数据平台，旨在实现对供应商ESG表现的常态化管理，并满足组织、监管和评级机构的披露和审查要求。

平台使用说明:

1. 带“*”的指标为**必填指标**，请根据贵公司实际情况填写
2. 我们鼓励您积极填报其他ESG指标，共同完善ESG管理体系
3. 为保证数据准确性，本集团将在25/26财年内**对供应商环境与社会等多维核心ESG数据进行第三方核查**
4. 为激励优秀实践，本集团将在25/26财年末对ESG综合表现最优的前三名供应商给予行业领先的**奖励**

让我们携手共建绿色供应链，迈向可持续发展未来！

开始填报 →

(Original instructions; branding removed)

Dear Supplier Partners, Welcome to the Supply Chain ESG Data Platform!

In the 25/26 fiscal year, the Group officially launched the Supply Chain ESG Data Platform, aiming to achieve regular management of suppliers' ESG performance and to meet the disclosure and review requirements of organizations, regulators, and rating agencies.

Platform Guidelines:

1. Indicators marked with "*" are **mandatory**. Please complete them according to your company's actual situation.
2. We encourage you to actively report other ESG indicators to jointly improve the ESG management system.
3. To ensure data accuracy, the Group will conduct **third-party verification of suppliers' multi-dimensional core ESG data,** including environmental and social aspects, during the 25/26 fiscal year.
4. To motivate outstanding practices, the Group will **grant industry-leading awards** to the top three suppliers with the best overall ESG performance at the end of the 25/26 fiscal year.

Let us work hand in hand to build a greener supply chain and move toward a sustainable future!

Start Filing

(English translation)

Panel C. Interface 3: Mandated KPIs, Carbon Audit

尊敬的供应商伙伴，欢迎登录供应链ESG数据平台！

在25/26财年，本集团正式推出供应链ESG数据平台，旨在实现对供应商ESG表现的常态化管理，并满足组织、监管和评级机构的披露和审查要求。

平台使用说明:

1. 带“*”的指标为**必填指标**，请根据贵公司实际情况填写
2. 我们鼓励您积极填报其他ESG指标，共同完善ESG管理体系
3. 为响应国家“双碳”战略并保证数据准确性，本集团将在25/26财年内**对供应商碳排放数据进行第三方核查**
4. 为激励优秀实践，本集团将在25/26财年末对ESG综合表现最优的前三名供应商给予行业领先的**奖励**

让我们携手共建绿色供应链，迈向可持续发展未来！

开始填报 →

(Original instructions; branding removed)

Dear Supplier Partner, welcome to the Supply Chain ESG Data Platform!

In the 25/26 fiscal year, the Group officially launched the Supply Chain ESG Data Platform, aiming to achieve regular management of suppliers' ESG performance and to meet the disclosure and review requirements of organizations, regulators, and rating agencies.

Platform Guidelines:

1. Indicators marked with "*" are **mandatory**. Please complete them according to your company's actual situation.
2. We encourage you to provide additional ESG data wherever possible to help strengthen the overall ESG management system.
3. In response to the national "dual carbon" strategy and to ensure data accuracy, the Group will **conduct third-party verification of suppliers' carbon emission data** in the 25/26 fiscal year.
4. To motivate outstanding practices, the Group will **grant industry-leading awards** to the top three suppliers with the best overall ESG performance at the end of the 25/26 fiscal year.

Let us work hand in hand to build a greener supply chain and move toward a sustainable future!

Start Filing

(English translation)

Panel D. Interface 4: Delegated KPIs, Comprehensive Audit

尊敬的供应商伙伴，欢迎登录供应链ESG数据平台！

在25/26财年，本集团正式推出供应链ESG数据平台，旨在实现对供应商ESG表现的常态化管理，并满足组织、监管和评级机构的披露和审查要求。

平台使用说明:

1. 带“*”的指标为**必填指标**，请根据贵公司实际情况填写
2. 请自主选择8项及以上指标作为**辅助指标**，我们鼓励您积极填报更多ESG指标，共同完善ESG管理体系
3. 为保证数据准确性，本集团将在25/26财年内**对供应商环境与社会等多维核心ESG数据进行第三方核查**
4. 为激励优秀实践，本集团将在25/26财年末对ESG综合表现最优的前三名供应商给予行业领先的**奖励**

让我们携手共建绿色供应链，迈向可持续发展未来！

开始填报 →

(Original instructions; branding removed)

Dear Supplier Partner, welcome to the Supply Chain ESG Data Platform!

In the 25/26 fiscal year, the Group officially launched the Supply Chain ESG Data Platform, aiming to achieve regular management of suppliers' ESG performance and to meet the disclosure and review requirements of organizations, regulators, and rating agencies.

Platform Guidelines:

1. Indicators marked with "*" are **mandatory**. Please complete them according to your company's actual situation.
2. Please independently select 8 or more indicators **as supplementary indicators.** We encourage you to actively report more ESG indicators to jointly improve the ESG management system.
3. To ensure data accuracy, the Group will **conduct third-party verification of suppliers' environmental and social multi-dimensional core ESG data** during the 25/26 fiscal year
4. To motivate outstanding practices, the Group will **grant industry-leading awards** to the top three suppliers with the best overall ESG performance at the end of the 25/26 fiscal year.

Let us work hand in hand to build a greener supply chain and move toward a sustainable future!

Start Filing

(English translation)

Panel E. Interface 5: Delegated KPIs, Carbon Audit

尊敬的供应商伙伴，欢迎登录供应链ESG数据平台！

在25/26财年，本集团正式推出供应链ESG数据平台，旨在实现对供应商ESG表现的常态化管理，并满足组织、监管和评级机构的披露和审查要求。

平台使用说明:

1. 带“*”的指标为**必填指标**，请根据贵公司实际情况填写
2. 请自主选择8项及以上指标作为**辅助指标**，我们鼓励您积极填报更多ESG指标，共同完善ESG管理体系
3. 为响应国家“双碳”战略并保证数据准确性，本集团将在25/26财年内**对供应商碳排放数据进行第三方核查**
4. 为激励优秀实践，本集团将在25/26财年末对ESG综合表现最优的前三名供应商给予行业领先的**奖励**

让我们携手共建绿色供应链，迈向可持续发展未来！

开始填报 →

(Original instructions; branding removed)

Dear Supplier Partner, welcome to the Supply Chain ESG Data Platform!

In the 25/26 fiscal year, the Group officially launched the Supply Chain ESG Data Platform, aiming to achieve regular management of suppliers' ESG performance and to meet the disclosure and review requirements of organizations, regulators, and rating agencies.

Platform Guidelines:

1. Indicators marked with "*" are **mandatory**. Please complete them according to your company's actual situation.
2. Please independently select 8 or more indicators **as supplementary indicators.** We encourage you to actively report more ESG indicators to jointly improve the ESG management system.
3. In response to the national "dual carbon" strategy and to ensure data accuracy, the Group will **conduct third-party verification of suppliers' carbon emission data** in the 25/26 fiscal year
4. To motivate outstanding practices, the Group will **grant industry-leading awards** to the top three suppliers with the best overall ESG performance at the end of the 25/26 fiscal year.

Let us work hand in hand to build a greener supply chain and move toward a sustainable future!

Start Filing

(English translation)

Appendix B. 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: This table presents all 128 platform fields. A single asterisk marks the common baseline required of every supplier, comprising 8 financial and 22 certification indicators. A double asterisk marks the eight indicators designated by the buyer for suppliers in the Mandate arm, while suppliers in the Delegation arm chose eight of their own. The 78 environmental, social, and governance indicators form the set over which suppliers make reporting decisions. A dagger marks 20 additional fields that are not counted as separate reporting 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, six automatically calculated rates and totals, and one field 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 2.3, summarized by domain in Table A2.

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 documented*, *Attachment 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 documented* 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 *Attachment 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 using a fixed topic taxonomy. We use GPT 5.6-Sol Ultra in the Codex interface. The classifier first identifies the evidence items the file contains, such as 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 corresponding ESG topic, 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. A file that supports several reporting items generates one record for each. For example, a water-fee workbook covering both tap-water charges and wastewater charges 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 *Attachment 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. The first identifies duplicates by comparing file hashes. A file submitted more than once within a round is counted once. A file resubmitted in a later round is retained and flagged when it remains valid for that reporting period. The second filter drops records when the evidence falls outside the reporting period or the file is unrelated to ESG reporting. The resulting outcomes measure documentary coverage in each round. 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

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

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 documented*. A file that documents several reporting items raises *Indicators documented* by more than one, and several items within the same topic raise *Indicators documented* separately while raising *Attachment 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 documented*, *Attachment 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 A1 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 documented* exceeds *Attachment 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 A1: 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 documented*	13
			Attachment categories	10
			Hard evidence	7
			Soft evidence	4

Notes. This table shows 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 documented* counts distinct reporting items, *Attachment 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 gives the details of the model in Section 5. The setup is followed by the supplier's processing decision, the benchmark without additional comprehension costs, and the capacity comparisons. Each formal result is presented with its proof. The extensions then use the same capacity budget to study the interaction of burdens, task allocation, the capacity threshold, and the filer's competing load.

D.1 Further Details of the 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 a continuous distribution F on $[0, \bar{k}]$. Appendix D.6.4 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 $r(c)$ units of capacity, as defined in equation (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 $\phi(c)g(e)$, with productivity given by equation (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. Appendix D.6.1 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 D.1 (Execution technology)**.** g is continuous on $[0, \infty)$ and twice continuously differentiable on $(0, \infty)$, with $g(0) = 0$, $g' > 0$, $g'' < 0$, and $g(e) \to \infty$ as $e \to \infty$.

Only values of g on $[0, \bar{k}]$ enter the results.

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 equation (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,

$$\Pi(c) \;=\; \int_0^{\bar{k}} w(c)\, E(k; c)\, dF(k). \tag{7}$$

Timing. The buyer announces c. In the baseline, 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.4. 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 D.2 (Binding capacity)**.** For every structure c and every $k \in (r(c), \bar{k}]$, $\alpha\, w(c)\, \phi(c)\, g'\big(k - r(c)\big) > \rho$.

Under Assumption D.2 the supplier devotes all residual capacity to the program, and equilibrium engagement is given by equation (6).

Lemma D.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.*

Proof of Lemma D.1. For $k < r(c)$ the feasible set in (5) is empty and engagement is zero. At $k = r(c)$ the only feasible effort is zero. For $k > r(c)$ 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 D.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 (6). Continuity at $k = r(c)$ follows from continuity of g and $g(0) = 0$, and strict monotonicity and strict concavity above $r(c)$ follow from $g' > 0$ and $g'' < 0$ in Assumption D.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. □

Figure 5, Panel A, draws the curve. A higher comprehension cost shifts it rightward, and higher productivity raises and steepens it. 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. The comparisons below weigh this cost against the productivity gain. Before making those comparisons, Appendix D.2 considers the initial decision to process the contract and states the conditions needed to extend the baseline results to optional processing.

D.2 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 if

$r(c) > 0$, and the difference is strictly increasing in k because $\alpha\, w(c)\, \phi(c)\, g'(e) > \rho$ for $e > 0$ by Assumption D.2. Processing therefore occurs at or above a unique threshold $k^e(c) > r(c)$ for $r(c) > 0$ 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 (6) from the threshold onward. At $k^e(c)$, engagement jumps to a positive level when $r(c) > 0$.

The single-crossing and population comparisons therefore require additional entry conditions. If the simple structure enters no later than the complex one, there is at most one switch from simple to complex after the simple structure enters, possibly through a jump at the complex entry point. Relative entry times depend on both overheads and returns, so the baseline rankings cannot be obtained by replacing $r(c)$ with $k^e(c)$ alone.

The comparisons that follow retain the baseline treatment of processing as part of participation. Extending them to optional processing requires checking the relative entry times described here.

D.3 Benchmark without Additional Comprehension Costs

The benchmark isolates the information and productivity gains from delegation and broad verification by setting their additional comprehension 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 depends on alignment. 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.4 analyzes this subgame.

Lemma D.2 (Congruent selection)**.** *At a given residual capacity $e > 0$, the supplier selects the informative set if and only if $\alpha \geq \underline{\alpha}(e, H) = \rho\tau/[(\bar{w} - \underline{w})\phi(D, H)g(e)]$, with ties resolved in favor of that set. Informative selection is feasible for $\alpha \leq 1$ only when this threshold is at most one.*

Delegation therefore raises the value of the report only when the supplier's interests are sufficiently aligned with the buyer's. The stake α measures the degree of alignment. In the baseline model, we assume this degree is sufficiently high, consistent with the reporting patterns discussed in Section 2.3. The subgame below shows why a positive selection cost does not imply uniform alignment at arbitrarily small residual capacity.

Assumption D.3 (Value of complexity)**.** $w(D,H) > w(M,H) > 0$ for each H, and $w(V,C) > w(V,T)$ for each V. The value $w(c)$ is then strictly increasing in complexity on each dimension.

Proposition D.1 (Benchmark without additional comprehension costs)**.** *Suppose complexity adds no comprehension cost, $\delta = \mu = 0$. Then for every capacity level $k \in (r_0, \bar{k}]$ and every structure c,*

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

For $c \neq (D,C)$, the engagement inequality is strict when $\phi(D,C) > \phi(c)$, and the value inequality is strict whenever $k > r_0$. The buyer weakly prefers (D,C) for every F and strictly prefers it if F assigns positive mass above r_0.

Proof of Proposition D.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 precisely when $\phi(D,C) > \phi(c)$, so the engagement ranking holds without any restriction on w. By Assumption D.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. Integration gives the weak population ranking. It is strict against every other cell when F assigns positive mass above r_0. □

Proposition D.1 states conditions supporting delegation and comprehensive verification. 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. Eliminating additional comprehension costs is stronger than the result requires. With $\delta, \mu > 0$ the same ranking holds once supplier capacity is large relative to those costs, as Corollary D.3(ii) below shows.

Corollary D.1 (Congruence failure)**.** *At residual capacity $e > 0$, if $\alpha < \underline{\alpha}(e, H)$, the convenient set is selected and delegation delivers no informativeness gain for that supplier. Sufficient congruence is a precondition for the vertical component of the trade-off analyzed below.*

D.4 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 $\underline{w} < \bar{w}$. The comparison holds continuation capacity fixed and represents the additional evaluation cost as $\rho\tau > 0$. If evaluation instead reduces continuation capacity, the selection condition must also include that loss.

Proof of Lemma D.2 and Corollary D.1. At a given residual capacity $e > 0$, the return to informative selection is $G(\alpha, e) = \alpha(\bar{w} - \underline{w})\phi(D, H)g(e)$. Comparing it with the evaluation cost $\rho\tau$ yields the stated threshold $\underline{\alpha}(e, H)$. If α is below this threshold, the convenient set is chosen and its value equals the mandated value. The threshold decreases with e and diverges as $e \downarrow 0$ when $\tau > 0$. Uniform informative selection therefore requires residual capacity bounded away from zero and a stake above the largest threshold on that restricted support. Assumption D.3 maintains the aligned case directly for the baseline comparisons. □

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.5 Capacity Costs and the Reversal

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

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

(i) Δ_{xs} is strictly increasing on $[\, r(x), \bar{k}\,]$, with $\Delta_{xs}(r(x)) < 0$. On this jointly feasible interval the curves therefore meet at most once. Any interior crossing changes the sign of Δ_{xs} from negative to positive.

(ii) An interior crossing exists 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) If $w(x) \geq w(s) > 0$ and the engagement crossing exists, 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 $k \in (r(s), k^*_\Pi)$.

Proof of Proposition D.2. (i) On $[\,r(x),\,\bar{k}\,]$ both curves are given by (6), so $\Delta_{xs}(k) = \phi(x)\,g(k-r(x)) - \phi(s)\,g(k-r(s))$, with derivative for $k > r(x)$

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

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 $k \geq r(x)$, 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 D.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 $\varphi(k^*) = (w(x) - w(s))\,E(k^*;s) \geq 0$ because the two engagement levels coincide at k^* and $w(x) \geq w(s)$ by the stated condition. Moreover, $\varphi'(k) > 0$ for $k > r(x)$ by the same derivative argument as in part (i). Let $k^*_\Pi = \inf\{k \geq r(x) : \varphi(k) \geq 0\} \leq k^*$. Continuity gives $\varphi < 0$ on $(r(s),\,k^*_\Pi)$, and for $k > k^*$ both $E(k;x) > E(k;s)$ and $w(x) \geq w(s) > 0$ give $\varphi(k) > 0$. □

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 be dominated in engagement at every capacity level. Congruence failure in Corollary D.1 removes the informational gain, but need not remove the productivity gain.

Figure 5, Panel B, 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 D.3 (Reversal in constrained populations)**.** *Consider the four comparisons that change one governance dimension while holding the other fixed. Let $\underline{k}^*$ and $\underline{k}^*_\Pi$ be their smallest engagement and value thresholds. A threshold is the first capacity at which the more complex structure weakly catches up after its entry point, and is $+\infty$ if it never does. Suppose F assigns positive mass above* $\max\{r(M,C), r(D,T)\}$.

(i) If $F(\underline{k}^*) = 1$, the three comparisons in the main proposition hold:

(a) For each $h \in \{C,T\}$, $\mathbb{E}_F E(k; M, h) > \mathbb{E}_F E(k; D, h)$.

(b) For each $v \in \{M,D\}$, $\mathbb{E}_F E(k; v, T) > \mathbb{E}_F E(k; v, C)$.

(c) The capacity-minimizing bundle ranks first and the fully complex bundle last:

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

(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).

Proof of Proposition D.3. For each comparison that adds one burden, the simpler structure has lower overhead and lower productivity. It weakly dominates in engagement below that pair's threshold by Proposition D.2, including the region in which one or both structures produce zero. Thus (M,T) weakly dominates each intermediate cell, and each intermediate cell weakly dominates (D,C), at every capacity below $\underline{k}^*$. The positive mass above both intermediate entry points makes each integrated comparison strict. Applying the same argument to $w(c)E(k;c)$ below $\underline{k}^*_\Pi$ gives part (ii). These comparisons do not rank (M,C) against (D,T). □

Appendix Figure A5, Panel A, collects the four curves. Panel D illustrates the role

of the capacity distribution. Parts (i)(a)–(c) of Proposition D.3 formalize the main-text proposition.

The proposition leaves the two single-burden cells unranked against each other. If $\pi_D = \pi_C$, they are ordered by their overheads, so (M, C) delivers more engagement than (D, T) when $\mu < \delta$ and less when $\mu > \delta$, wherever the lower-cost cell is active. With different productivity gains, overheads alone do not determine their ranking. The four-cell regression identifies this contrast, which is why the middle cells carry information about the relative size of the two frictions.

Proposition D.1 and Proposition D.3 describe how the preferred structure depends on supplier capacity. Delegation and comprehensive verification perform best when comprehension is cheap or capacity is abundant. Simpler structures perform best in networks whose capacity lies below the relevant thresholds.

D.6 Extensions and Proofs

Appendix Figure A5, Panels B–E, illustrates four extensions of 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.

D.6.1 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

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

Corollary D.2 (Compounding). *(i) For every $k > r(D,C)$, $I(k) < 0$.*

(ii) On the jointly feasible region, $I(k)$ 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 $(r(D,C), \bar{k}\,]$.

Proof of Corollary D.2. (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 D.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),

$$
\begin{aligned}
\big[\, E(k;M,C) - E(k;D,C) \,\big] + \big[\, E(k;D,T) - E(k;D,C) \,\big& \\
- \;\big[\, E(k;M,T) - E(k;D,C) \,\big] \;=\; -\, I(k),&
\end{aligned}
$$

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. 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.

Part (iv) restricts the support to suppliers able to process every structure. Below $r(D, C)$, the fully complex cell produces zero and the sign of $I(k)$ also depends on which other cells are active. Proposition D.3 and Corollary D.2 hold together on populations supported on $(r(D, C), \underline{k}^*)$, with thresholds evaluated under the same overhead specification. This interval is nonempty when $r(D, C) < \underline{k}^*$. It gives the joint restriction under which the main proposition and the corollary both apply.

D.6.2 Task Tractability

The task-allocation result below derives Implication (i). 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 $\psi(\lambda_j s_j)$. The function ψ is continuous on $[0, \infty)$ and twice continuously differentiable on $(0, \infty)$, with $\psi' > 0$, $\psi'' < 0$, and a finite right 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. The allocation problem and proof follow the proposition.

Proposition D.4 (Task tractability)**.** *For positive residual capacity, let $\chi(k; c)$ denote its shadow value and let the value of content be equal across tasks. The shadow value is nonincreasing in $k - r(c)$ and strictly decreasing wherever $\chi > \rho$. 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 $\chi(k; c) < \eta_j$. Tasks that become active therefore do so in order of tractability*

as capacity grows. Holding content values and productivity fixed while $\chi > \rho$, a reduction in $r(c)$ raises residual capacity, lowers χ, and raises engagement on every active task. The next task activated, if a threshold is crossed, is the most tractable inactive task.

Proof of Proposition D.4. 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, where $w_j(c) > 0$ is the per-unit value of content on task j. Let $\nu \geq 0$ be the multiplier on this constraint and define the total shadow value $\chi = \rho + \nu$. 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, $\eta_1 > \eta_2 > \cdots > \eta_N$. When $\chi > \rho$, the allocations sum to $e = k - r(c)$. Each active allocation is continuous and strictly decreasing in χ, so this sum determines a continuous shadow value that strictly decreases with e. Once capacity is slack, $\nu = 0$ and $\chi = \rho$. Thus tasks that become active do so in the order $\eta_1 > \eta_2 > \cdots$ as e grows. For an active task, implicit differentiation of the first-order condition gives

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

While $\chi > \rho$, a governance change that lowers $r(c)$ raises e one for one, lowers χ, raises s_j on every active task, and activates the most tractable inactive task if its threshold is crossed. □

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.

D.6.3 Supplier Capacity

The crossing in Proposition D.2 gives the formal comparison behind Implication (ii).

Corollary D.3 (Supplier capacity)**.** *Let s and x be as in Proposition D.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 once both structures' comprehension costs are covered.

(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 benchmark ranking of Proposition D.1 holds in population: $\Pi(D, C) > \Pi(c)$ for all $c \neq (D, C)$.

Proof of Corollary D.3. Part (i) is Proposition D.2(i). For part (ii), take any F' supported on $(\,k^*_{\max},\,\bar{k}\,]$. For each $c \neq (D, C)$, the weighted gap $w(D, C)E(k; D, C) - w(c)E(k; c)$ is strictly increasing above $r(D, C)$ by the derivative argument in Proposition D.2, since both the value weight and productivity are higher under (D, C). It is positive above its value crossing. All mass of F' lies above every such crossing, and integration preserves the strict ranking. □

Proposition D.3 and Corollary D.3 together delimit the two regimes. Delegation and comprehensive verification perform best when comprehension costs vanish or the network's capacity lies above the relevant thresholds. Simpler structures perform best below those thresholds. Which regime applies is an empirical property of the network.

D.6.4 Reporting Roles

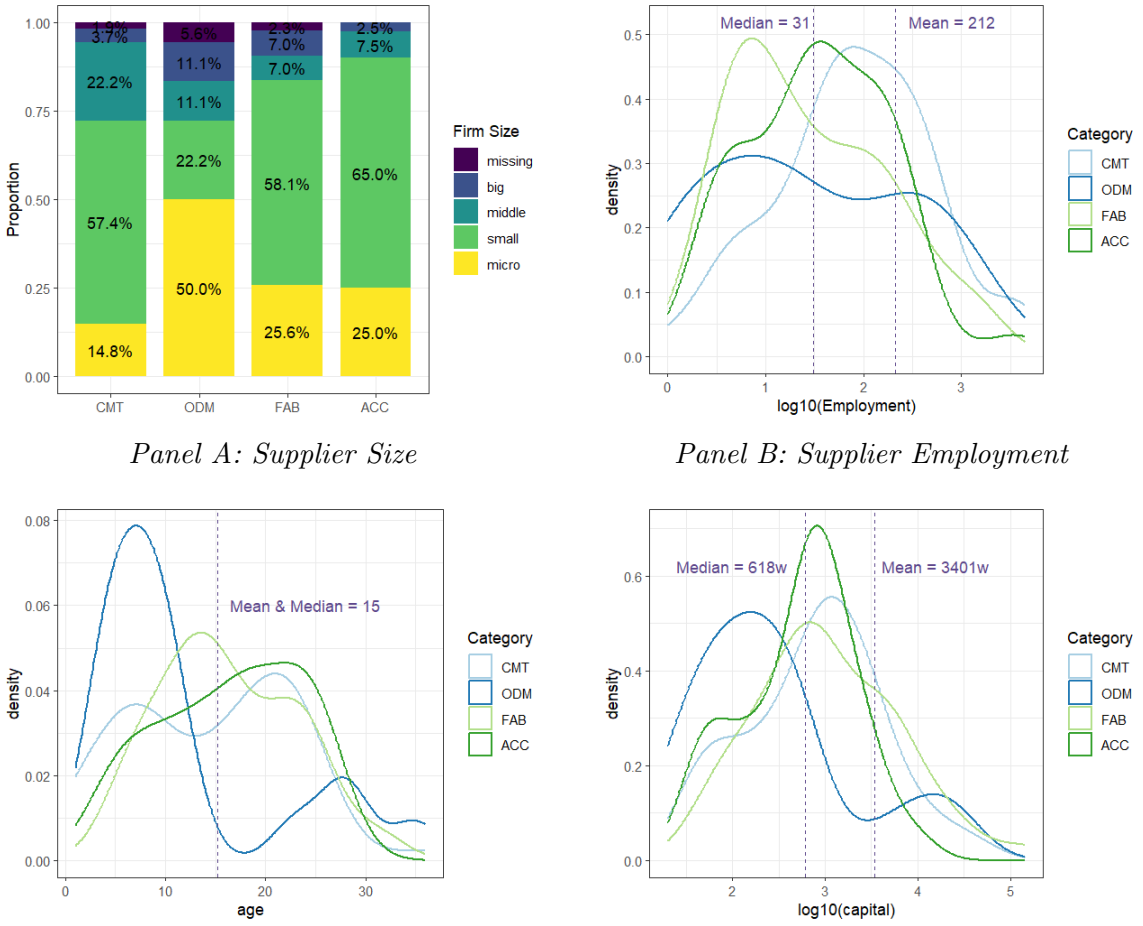
Implication (iii) concerns the capacity that reaches the task inside the firm. The baseline treats the supplier's capacity as a primitive, but the competing load of the person handling the filing reduces what is available. Let K denote the organization's capacity endowment and $0 \leq \ell \leq K$ the 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.

Corollary D.4 (Reporting roles)**.** *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 $[\, r(M, h),\ r(D, h)\,]$. When capacity lies between these entry points, increasing the filer's competing load reduces the contrast to zero as k falls to $r(M, h)$. The horizontal contrast satisfies the same properties.*

Proof of Corollary D.4. Continuity follows from Lemma D.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 $k \downarrow r(M, h)$. On this interval, increasing the filer's competing load lowers $k = K - \ell$ toward the entry point of the simpler structure, and the contrast converges to zero continuously. The horizontal contrast replaces δ with μ throughout, so the same argument applies with entry points $r(V, T) < r(V, C)$. □

The effect of governance on engagement depends on the capacity available to the reporting task. Dedicated staff face fewer competing demands, while owners and general managers filing personally have less capacity available. As competing demands reduce task capacity toward the lower entry threshold, governance contrasts shrink toward zero.

Appendix E. Additional Figures and Tables



Panel A: Supplier Size

Panel B: Supplier Employment

Panel C: Supplier Age

Panel D: Supplier Registered Capital

Figure A1: 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.

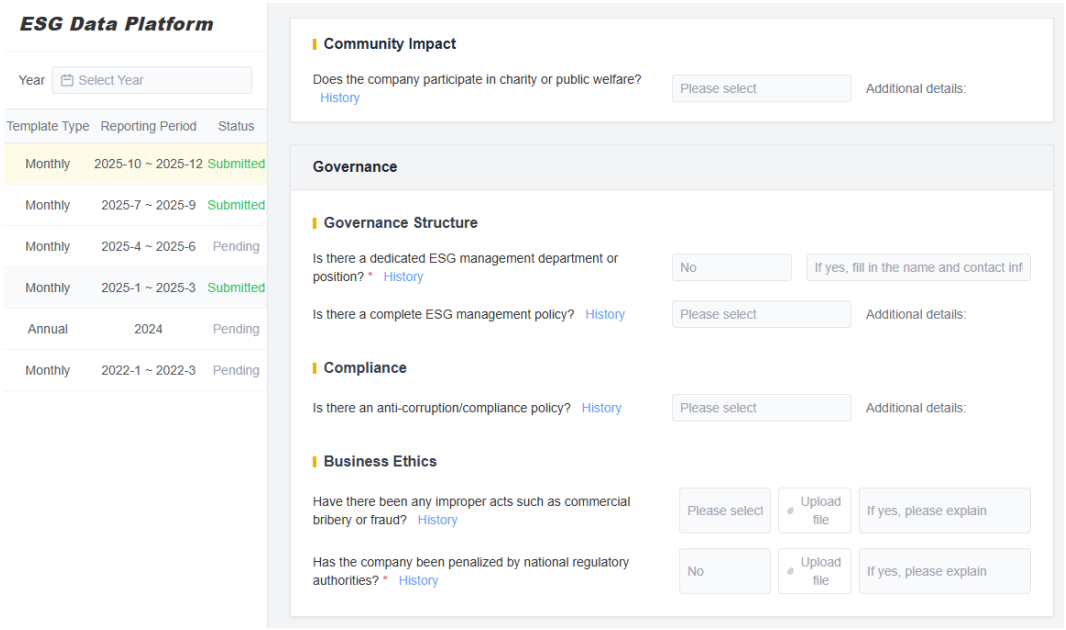


Figure A2: Digital Platform Interface: Main Reporting Dashboard

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.

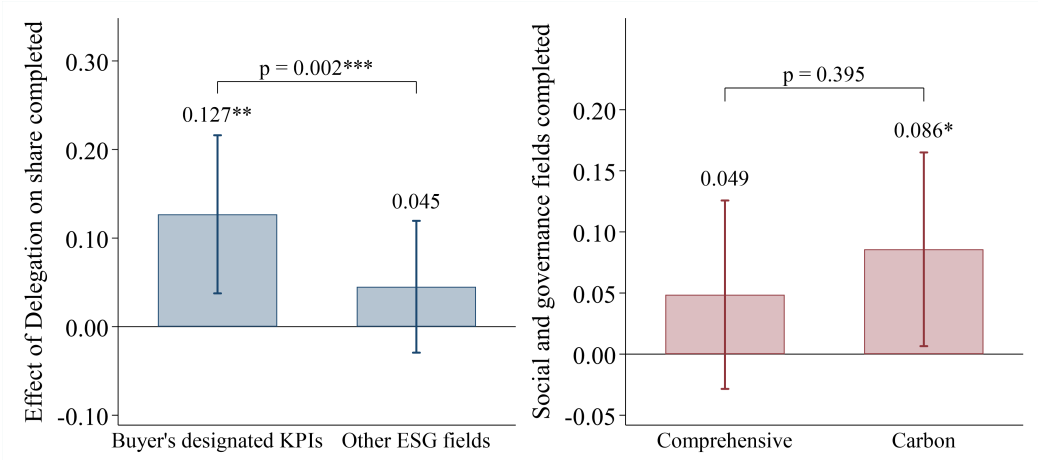


Figure A3: Supplier Choices and the Buyer's Interests

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 buyer's designated KPIs and for the other 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. 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. Brackets report two-sided p-values for equality of the two effects in each panel. Delegation increased completion more on the buyer's designated KPIs than on other ESG fields, and targeted carbon monitoring increased completion of the unaudited social and governance fields. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

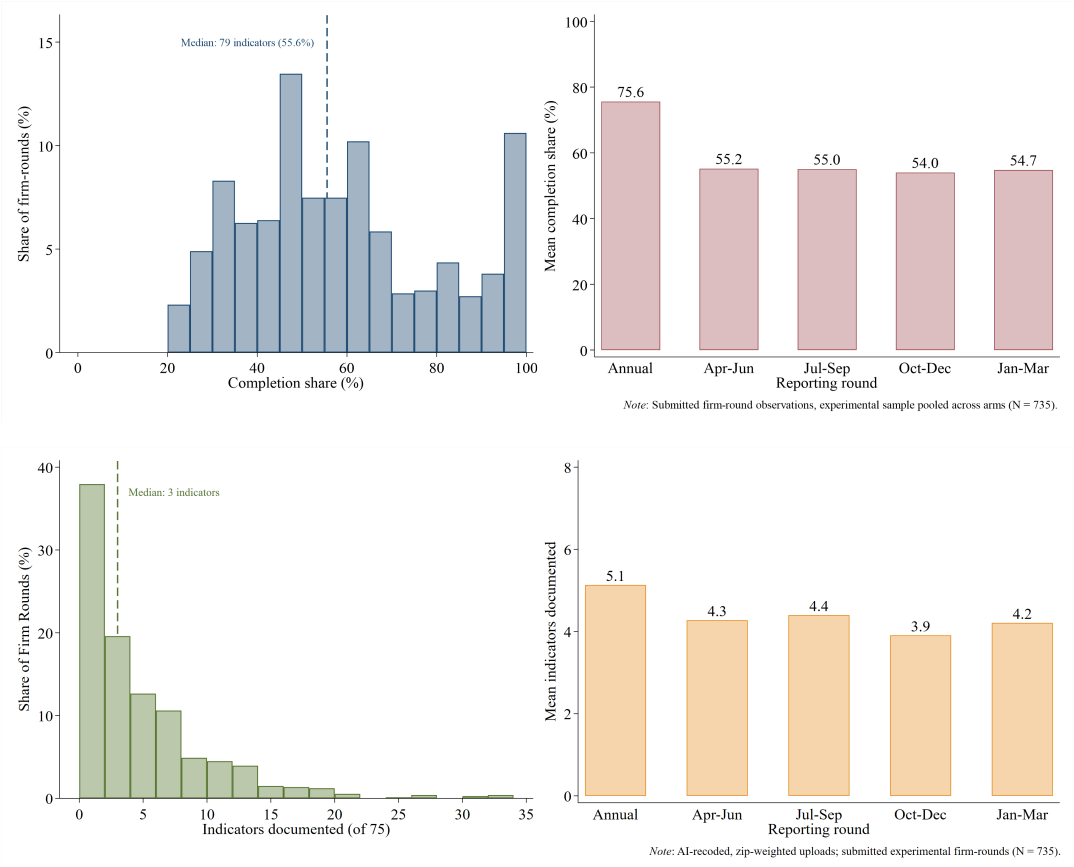
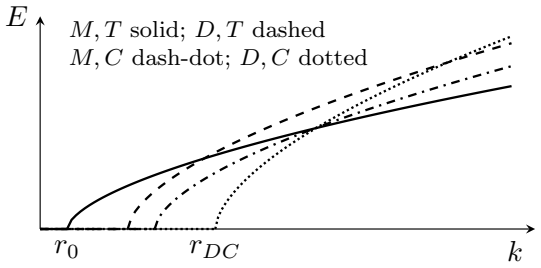
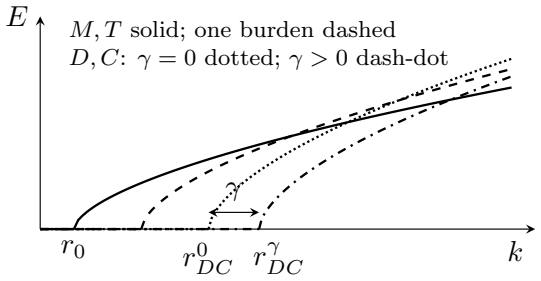


Figure A4: Reporting Completion and Supporting Attachments Across Filing Rounds

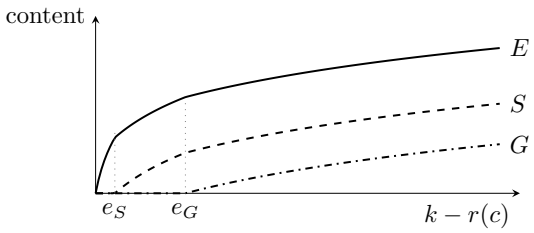
Notes: This figure presents four descriptive panels for the 735 submitted supplier-round observations in the experimental sample. Panel A shows the distribution of the share of indicators completed. The median supplier-round completes 79 indicators, equivalent to 55.7 percent of the indicator set. Panel 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.



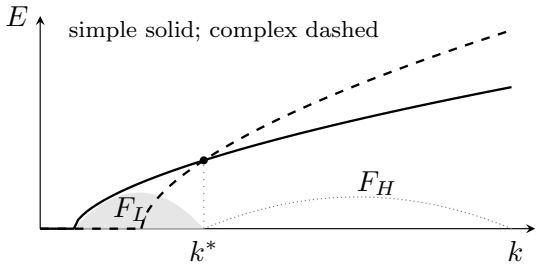
Panel A: Four Governance Structures



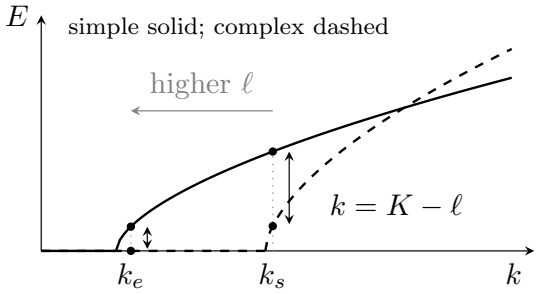
Panel B: Compounding Burdens



Panel C: Task Tractability



Panel D: The Capacity Distribution



Panel E: Reporting Roles

Figure A5: Governance Structures and Extensions of the Capacity Model

Notes: Panel A retains the four governance curves, with $g(e) = e^{0.6}$, $r_0 = 0.8$, $\delta = 1.8$, $\mu = 2.6$, $\pi_D = 0.42$, and $\pi_C = 0.30$. Panel B uses $r_0 = 1$, $\delta = \mu = 2$, and $\pi_D = \pi_C = 0.25$ to compare additive overheads with an additional comprehension cost $\gamma = 1.5$. Concavity can produce $I(k) < 0$ even at $\gamma = 0$ when all structures are feasible. Panel C solves the task-allocation problem with $\psi(z) = \log(1 + z)$, unit content weights, $\rho = 0.1$, and tractabilities $(3, 1, 0.45)$ for environmental, social, and governance tasks. The capacity constraint binds over the range drawn. Social and governance tasks activate at $e_S = 2/3$ and $e_G = 28/9$. Panel D compares simple and complex curves with $(r, \phi) = (1, 1)$ and $(3, 1.55)$. The shaded and dotted densities have support below and above the crossing, respectively, and are scaled for display. Panel E uses the same curves. Greater competing load moves task capacity from the staff example $k_s = 3.1$ toward the executive example $k_e = 1.2$, shrinking the contrast near the simple structure's entry point. These are illustrative parameterizations, not estimates.

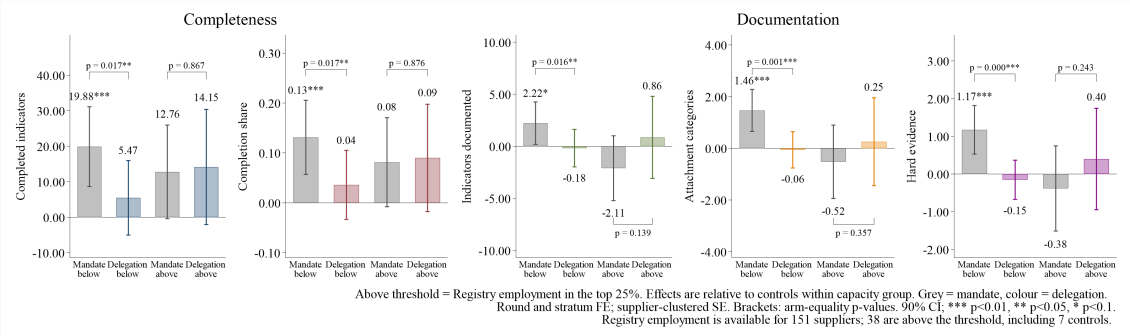


Figure A6: Authority Allocation and Capacity Measured by Registry Employment

Notes: This figure re-estimates Figure 8 using registry employment as the capacity measure. Employment is the maximum of the available registry headcounts, observed for 151 of the 152 suppliers. The threshold is the top quartile, above 168.5 employees, containing 38 suppliers: 7 control, 17 mandate, and 14 delegation. 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 arm-by-capacity-group interactions, a capacity-group indicator, and reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Each coefficient is an arm effect measured against controls in the same capacity group. Bars show coefficients and whiskers show 90 percent confidence intervals. Brackets report the two-sided p-value for equality of the Mandate and Delegation coefficients within each capacity group. The point estimates favor Mandate below the threshold and Delegation above it on all five outcomes. Small treatment and control groups above the threshold leave some estimates imprecise. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

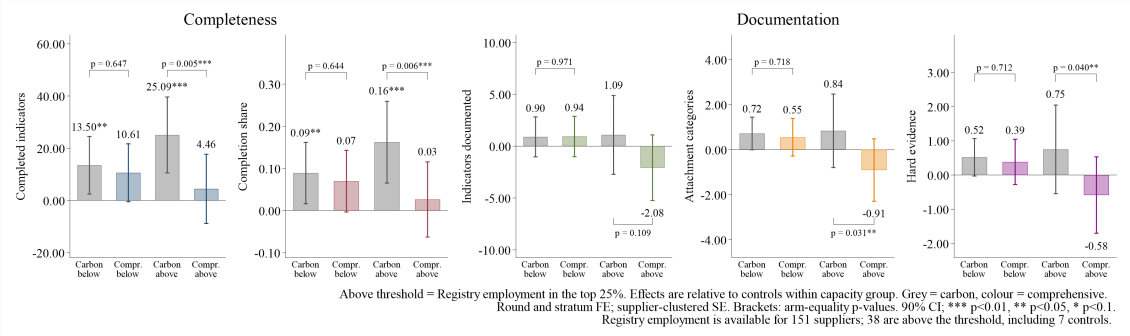


Figure A7: Monitoring Scope and Capacity Measured by Registry Employment

Notes: This figure repeats the design of Figure A6 for Targeted Carbon Monitoring and Comprehensive Monitoring, using the same registry-employment measure, top-quartile threshold, 151-supplier sample, and five outcomes. The group above the threshold contains 38 suppliers: 7 control, 14 targeted, and 17 comprehensive. Each coefficient is a treatment effect relative to controls in the same capacity group. Regressions include arm-by-capacity-group interactions, a capacity-group indicator, and reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Whiskers show 90 percent confidence intervals, and brackets report the two-sided p-value for equality of the two monitoring coefficients within each capacity group. The estimated Targeted Carbon–Comprehensive gap is larger above the threshold on all five outcomes, although documentation estimates are less precise given the small treatment and control groups. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

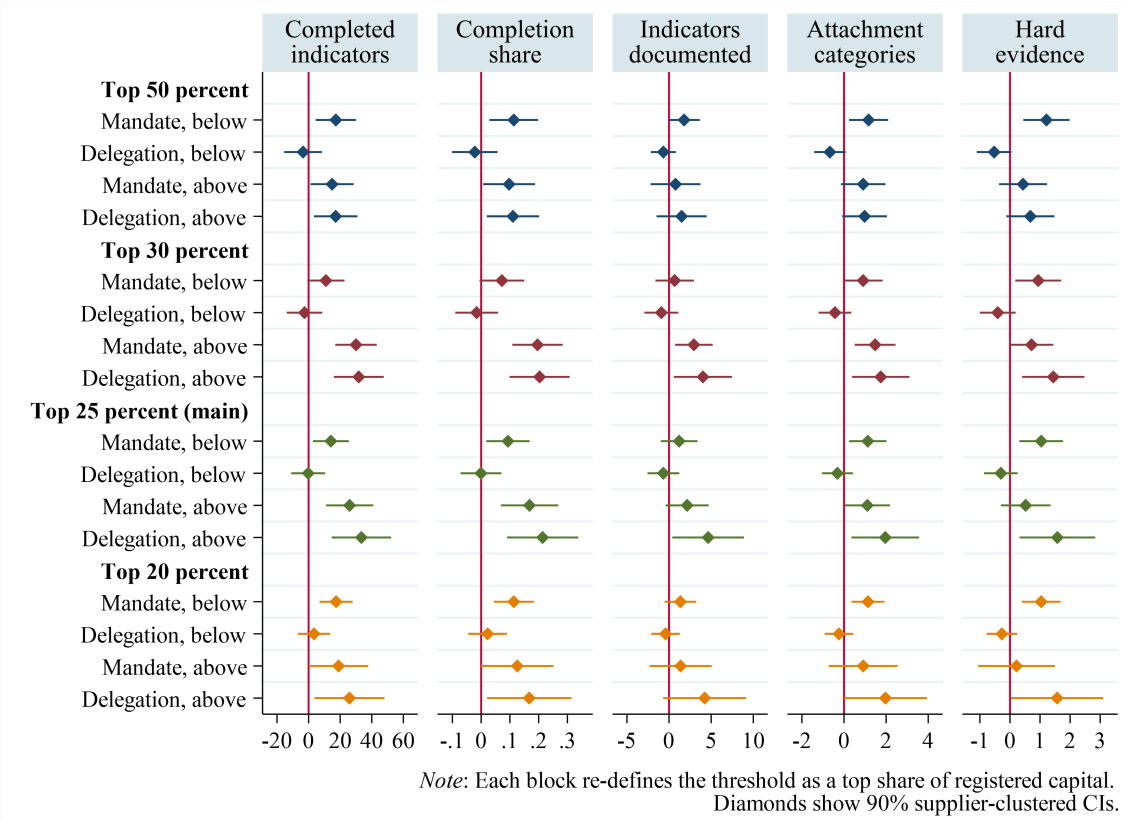


Figure A8: Authority Allocation and the Definition of the Capacity Threshold

Notes: This figure re-estimates Figure 8 using four capacity thresholds: registered capital in the top 50, 30, 25, and 20 percent of the sample. The top 25 percent block is the main specification. Each block reports treatment effects relative to controls in the same capacity group. 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 arm-by-capacity-group interactions, a capacity-group indicator, and reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Diamonds show coefficient estimates, and horizontal lines show 90 percent confidence intervals. At every cutoff, Delegation's point estimates remain near zero below the threshold and positive above it. Mandate's point estimates exceed Delegation's below the threshold, while Delegation's exceed Mandate's above it on all five outcomes.

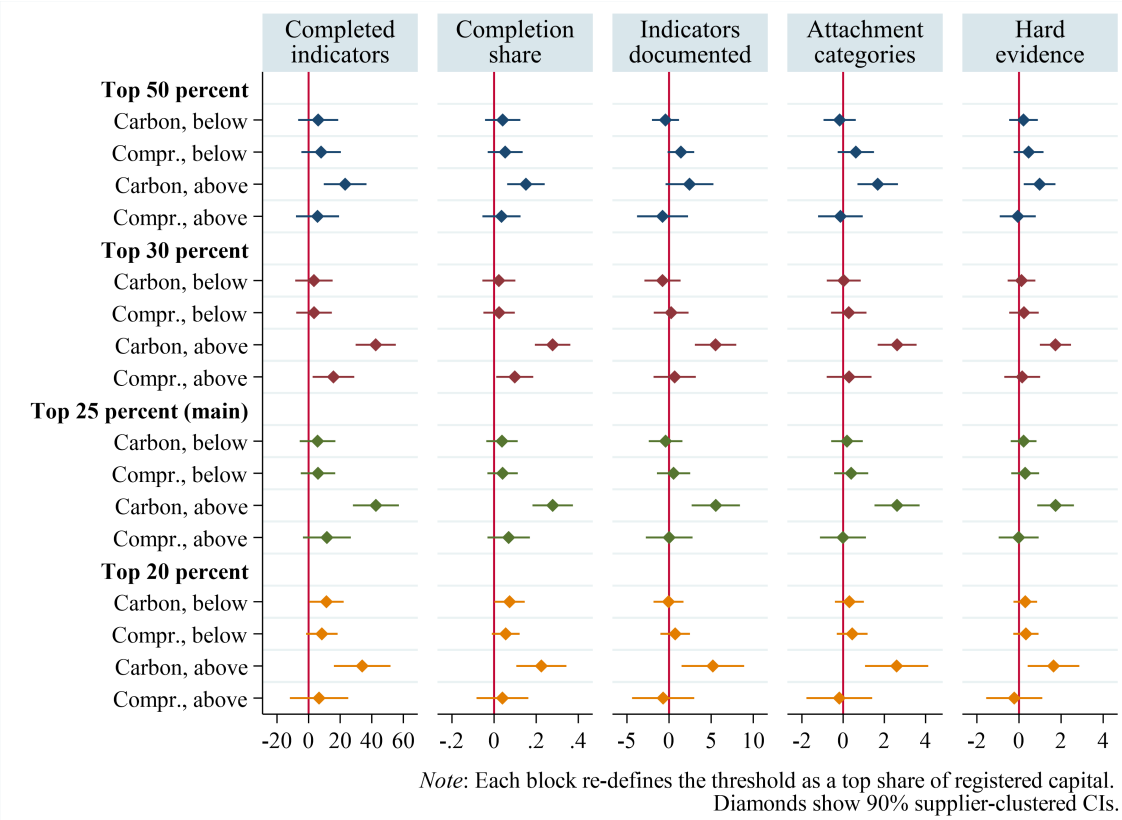


Figure A9: Monitoring Scope and the Definition of the Capacity Threshold

Notes: This figure repeats the design of Figure A8 for Targeted Carbon Monitoring and Comprehensive Monitoring, using the same four registered-capital thresholds and five outcomes. Each coefficient measures a treatment effect relative to controls in the same capacity group. Regressions include arm-by-capacity-group interactions, a capacity-group indicator, and reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Diamonds show coefficient estimates, and horizontal lines show 90 percent confidence intervals. At every cutoff, Targeted Carbon Monitoring's point estimates exceed Comprehensive Monitoring's above the threshold on all five outcomes. Below the threshold, the differences are smaller and their direction varies across outcomes and cutoffs.

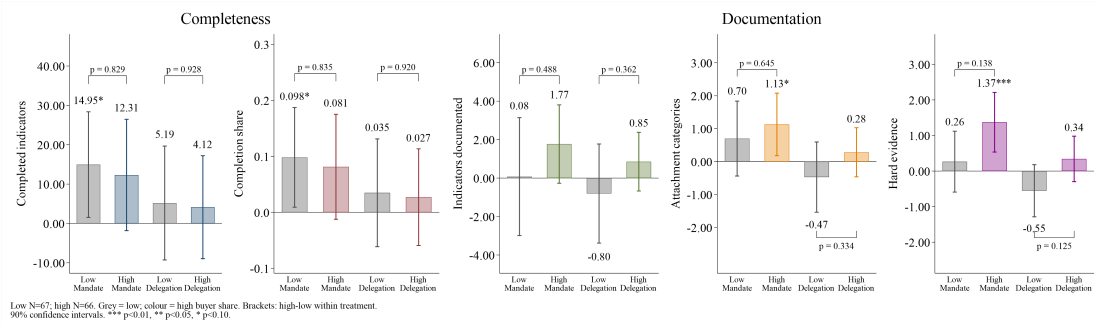


Figure A10: Buyer Dependence and the Returns to Authority Allocation

Notes: This figure reports the effects of Mandate and Delegation separately for suppliers with low and high buyer dependence. Dependence is sales to the lead firm divided by supplier annual revenue, both for FY2024 and measured in 10,000 CNY. We retain nonnegative buyer sales and positive revenue with shares no greater than one; the high group is strictly above the median among the 133 suppliers with valid shares in the main estimation sample. Each arm effect is measured against control suppliers in the same dependence group. 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 interact both treatment arms with dependence groups and include the group main effect, reporting-round and randomization-stratum fixed effects, and supplier-clustered standard errors. Bars show coefficients and whiskers show 90 percent confidence intervals. Brackets compare high and low dependence within the same treatment. These comparisons do not reject equality at the five percent level; their confidence intervals do not establish equivalence. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

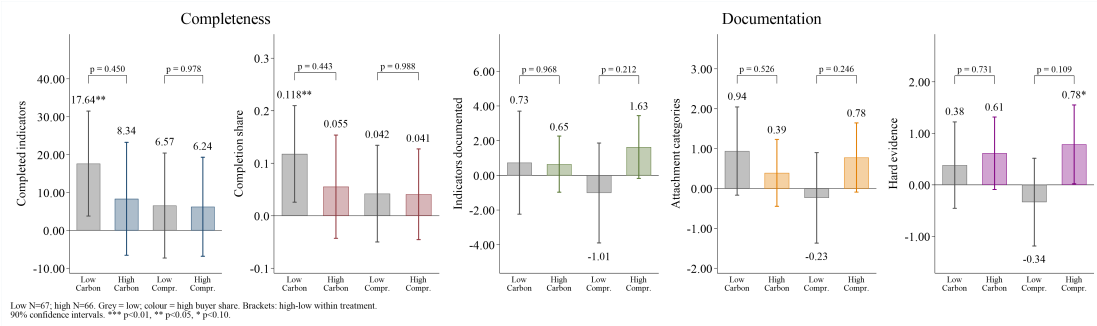


Figure A11: Buyer Dependence and the Returns to Monitoring Scope

Notes: This figure repeats the design of Figure A10 for Targeted Carbon Monitoring and Comprehensive Monitoring, using the same FY2024 buyer-sales share, median split, supplier sample, and five outcomes. Each coefficient is a treatment effect relative to controls within the same dependence group. Brackets report the two-sided p-value for equality of the high- and low-dependence effects within each monitoring arm. Regressions include dependence-group main effects, reporting-round and randomization-stratum fixed effects, with standard errors clustered by supplier. Whiskers show 90 percent confidence intervals. None of the high–low comparisons rejects at the five percent level, but economically meaningful moderation remains possible. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

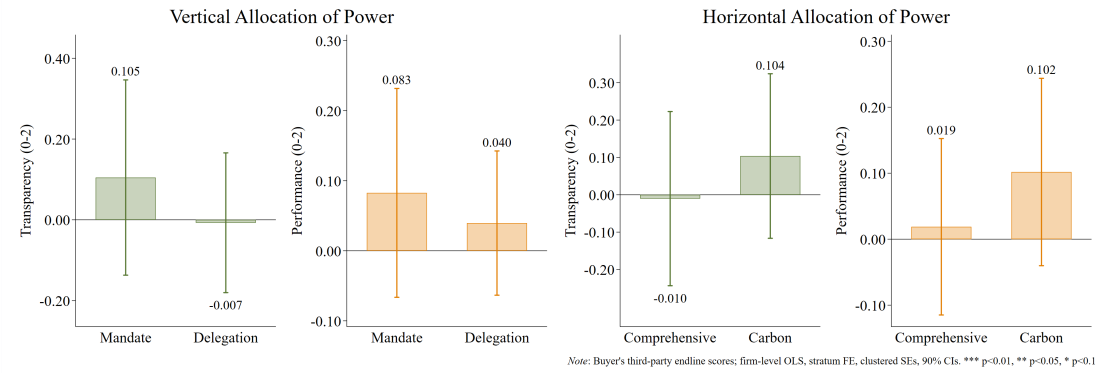
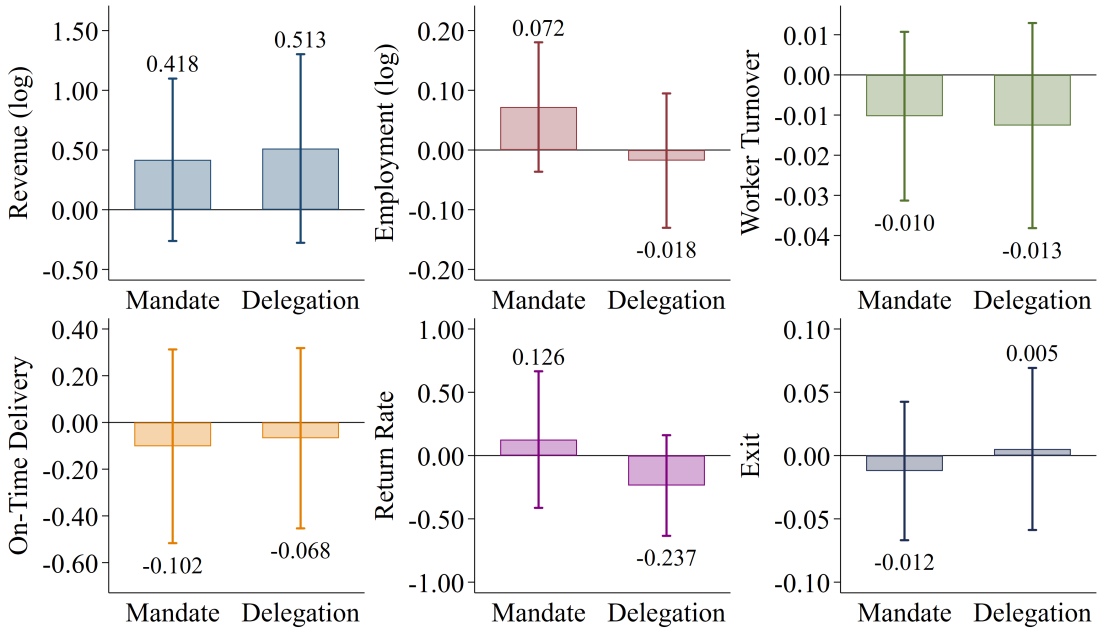


Figure A12: Supplier ESG Performance Rated by the Lead Firm

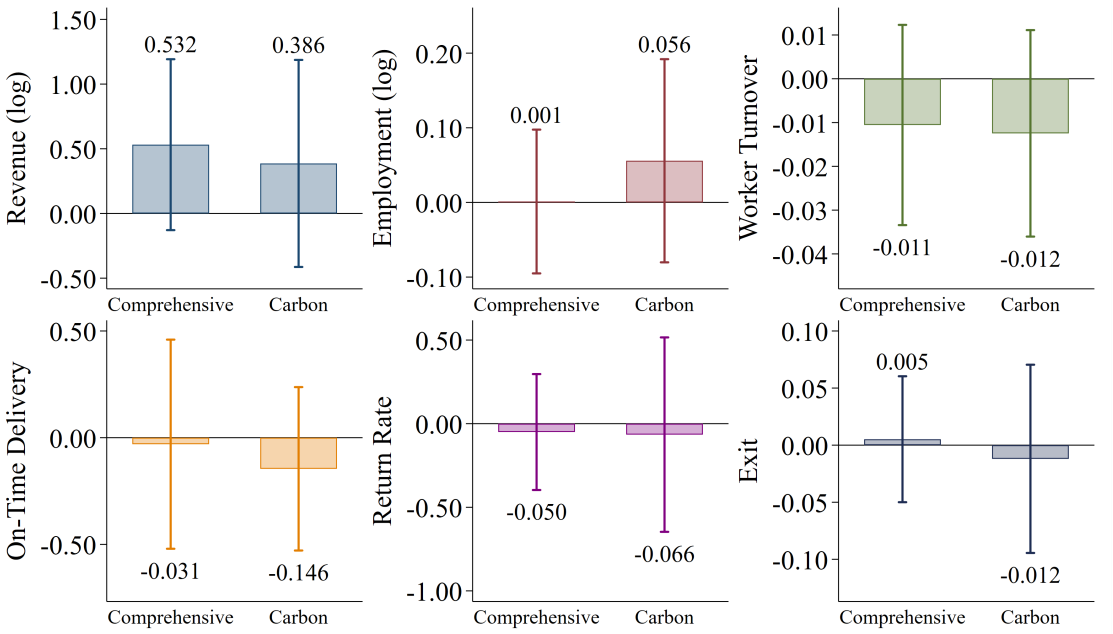
Notes: This figure reports treatment effects on two endline supplier scores assigned by the lead firm: transparency of optional information and a performance bonus. The scoring team did not observe treatment status. Each cross-sectional regression includes randomization-stratum fixed effects. The mandate and targeted-monitoring coefficients are positive for both outcomes, while the confidence intervals include zero. Table A13 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 A13: 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, worker turnover and exit are fractions, and on-time delivery and return rates are in percentage points. 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 A14 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 A14: 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 A13, under the same specifications. No arm has a statistically significant effect on these outcomes, and every confidence interval covers zero.

Table A2: 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: This table presents the 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 B.

Table A3: Authority Allocation and Supplier Engagement

	(1) Completed indicators	(2) Completion share	(3) Indicators documented	(4) Attachment categories	(5) Hard evidence
Mandate	16.037***	0.105***	1.075	0.958**	0.766**
	(5.396)	(0.036)	(1.044)	(0.422)	(0.334)
Delegation	6.894	0.045	0.221	0.081	0.029
	(5.488)	(0.036)	(1.024)	(0.401)	(0.307)
Constant	77.578***	0.536***	3.934***	1.892***	1.065***
	(4.218)	(0.028)	(0.821)	(0.302)	(0.237)
Control mean	77.293	0.535	3.947	1.907	1.053
Observations	735	735	735	735	735
R^2	0.204	0.320	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	0.061	0.062	0.327	0.028	0.018

Notes: This table presents 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 the voluntary ESG reporting program, which is the comparison the buyer faces. The final row reports the p-value for equality of the two arm coefficients. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A4: Monitoring Scope and Supplier Engagement

	(1) Completed indicators	(2) Completion share	(3) Indicators documented	(4) Attachment categories	(5) Hard evidence
Comprehensive monitoring	7.285	0.047	0.219	0.227	0.175
	(5.279)	(0.035)	(1.008)	(0.418)	(0.331)
Targeted carbon monitoring	15.622***	0.103***	1.075	0.806**	0.615**
	(5.579)	(0.037)	(1.057)	(0.405)	(0.311)
Constant	77.596***	0.537***	3.936***	1.894***	1.067***
	(4.201)	(0.028)	(0.820)	(0.301)	(0.236)
Control mean	77.293	0.535	3.947	1.907	1.053
Observations	735	735	735	735	735
R^2	0.201	0.318	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	0.090	0.089	0.321	0.139	0.151

Notes: This table presents OLS estimates of the effects of Comprehensive Monitoring and Targeted Carbon Monitoring relative to the control group. Outcome definitions follow Table A3. 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 the voluntary ESG reporting program, which is the comparison the buyer faces. The final row reports the p-value for equality of the two arm coefficients. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A5: Engagement Across the Four Governance Bundles

	(1) Completed indicators	(2) Completion share	(3) Indicators documented	(4) Attachment categories	(5) Hard evidence
Mandate, comprehensive	14.703**	0.096**	0.661	0.695	0.585
	(6.045)	(0.040)	(1.221)	(0.546)	(0.441)
Mandate, targeted carbon	17.393**	0.114**	1.521	1.240***	0.961***
	(6.741)	(0.045)	(1.172)	(0.445)	(0.347)
Delegation, comprehensive	-0.349	-0.003	-0.230	-0.250	-0.242
	(6.033)	(0.040)	(1.127)	(0.457)	(0.347)
Delegation, targeted carbon	13.982**	0.092**	0.662	0.405	0.294
	(6.527)	(0.043)	(1.252)	(0.500)	(0.378)
Constant	77.573***	0.536***	3.933***	1.892***	1.065***
	(4.210)	(0.028)	(0.822)	(0.302)	(0.237)
Control mean	77.293	0.535	3.947	1.907	1.053
Observations	735	735	735	735	735
R^2	0.223	0.337	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	0.009	0.112	0.002	0.001

Notes: This table presents estimates for each crossed treatment cell relative to the control group. The four cells combine mandated or delegated KPI selection with comprehensive or targeted carbon monitoring. Outcome definitions follow Table A3. 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 the voluntary ESG reporting program. The final row reports, for completeness, the p-value from a test of equality between the fully simplified and the fully complex cell. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A6: 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	0.105***	0.105**	0.105**	0.109**	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	0.536***	0.536***	0.536***	0.532***	1.892***	1.892***	1.892***	1.856***
	(0.028)	(0.026)	(0.026)	(0.027)	(0.302)	(0.218)	(0.218)	(0.220)
Control mean	0.535	0.535	0.535	0.531	1.907	1.907	1.907	1.873
Observations	735	735	735	152	735	735	735	152
R^2	0.320	0.320	0.320	0.215	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 presents robustness checks for the authority-allocation estimates. Main is the repeated supplier-round specification of Table A3, which clusters standard errors on the supplier. Stratum cluster retains that coefficient specification and clusters on the stratum instead, allowing a different within-stratum dependence structure. 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 0.04. Parentheses report standard errors. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A7: 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)	(0.046)	(0.418)	(0.438)	(0.438)	(0.434)
Targeted carbon monitoring	0.103***	0.103**	0.103***	0.103**	0.806**	0.806**	0.806**	0.843**
	(0.037)	(0.037)	(0.037)	(0.037)	(0.405)	(0.315)	(0.315)	(0.318)
Constant	0.537***	0.537***	0.537***	0.532***	1.894***	1.894***	1.894***	1.858***
	(0.028)	(0.026)	(0.026)	(0.027)	(0.301)	(0.207)	(0.207)	(0.209)
Control mean	0.535	0.535	0.535	0.531	1.907	1.907	1.907	1.873
Observations	735	735	735	152	735	735	735	152
R^2	0.318	0.318	0.318	0.209	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 presents robustness checks for 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 0.04. Parentheses report standard errors. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A8: Completion Share: Sample and Measurement Robustness

	Completion share				
	Definition of completeness			Reporting coverage	
	(1) ESG fields only	(2) Monthly entries averaged first	(3) Eight fields deducted	(4) Include unsubmitted drafts	(5) Missing reports at control minimum
Panel A. Authority allocation					
Mandate	0.132***	0.088***	0.085**	0.108***	0.087**
	(0.045)	(0.032)	(0.037)	(0.036)	(0.037)
Delegation	0.056	0.031	0.021	0.047	0.032
	(0.045)	(0.032)	(0.037)	(0.037)	(0.036)
p-value, arm equality	0.061	0.048	0.060	0.062	0.093
Panel B. Monitoring scope					
Comprehensive	0.059	0.039	0.024	0.050	0.043
	(0.044)	(0.031)	(0.036)	(0.035)	(0.035)
Targeted carbon	0.128***	0.079**	0.082**	0.105***	0.075*
	(0.046)	(0.033)	(0.038)	(0.037)	(0.038)
p-value, arm equality	0.090	0.162	0.089	0.089	0.337
Control mean	0.419	0.639	0.535	0.534	0.529
Observations	735	735	735	736	775
Suppliers	152	152	152	152	155
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: This table presents completion-share estimates under alternative outcome definitions and reporting samples. Observations are supplier-rounds. Columns (1)–(3) use 735 submitted reports from 152 suppliers. (1) Excludes the 8 financial and 22 certificate indicators from numerator and denominator in every arm, leaving 78 annual or 128 quarterly ESG entries. (2) Averages the three monthly completion entries within each environmental indicator before averaging the 108 indicator concepts. (3) Deducts eight filled and eight applicable entries from treated reports, leaving controls unchanged. This is a uniform count adjustment, not deletion of the actual selected KPIs. (4) Adds one unsubmitted annual draft with 27 filled entries from an already included supplier. (5) Retains all 155 suppliers over five rounds, including the draft and 39 no-filling records. Six annual slots lack confirmed early assignment. No-filling records receive the same-round submitted-control minimum, 50/108 annually and 36/158 quarterly. Inference conditions on this imputation. Each coefficient compares an arm with control. All models include round and randomization-stratum fixed effects. Supplier-clustered standard errors are in parentheses. Arm-equality p-values compare mandate with delegation and comprehensive with targeted carbon monitoring. Control means are unadjusted means of each dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A9: 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	15.052	14.331	17.743	29.383	−11.640 (9.461)	0.220
Completion share (pp)	9.941	9.487	11.752	19.428	−7.676 (6.256)	0.222
Indicators documented	0.890	0.892	1.750	1.783	−0.032 (1.694)	0.985
Attachment 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 presents linear combinations of the four-cell regression of Table A5 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 A3. 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 A10: 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***	0.107**	0.146*	0.281***	-0.013	0.077
	(0.054)	(0.048)	(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***	0.622***	0.600***	0.317***	0.195***	0.155***
	(0.043)	(0.036)	(0.063)	(0.057)	(0.056)	(0.056)
R^2	0.216	0.179	0.104	0.158	0.095	0.089
Panel B: Monitoring scope						
Comprehensive monitoring	0.066	0.041	0.123	0.099	-0.126*	0.007
	(0.053)	(0.047)	(0.079)	(0.081)	(0.066)	(0.072)
Targeted carbon monitoring	0.159***	0.078	0.156**	0.243***	0.053	0.084
	(0.058)	(0.048)	(0.076)	(0.081)	(0.078)	(0.079)
Constant	0.254***	0.622***	0.600***	0.317***	0.195***	0.155***
	(0.043)	(0.035)	(0.063)	(0.056)	(0.056)	(0.056)
R^2	0.223	0.163	0.105	0.139	0.131	0.092
Control mean	0.251	0.621	0.596	0.316	0.200	0.156
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 presents 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. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A11: Supplier Capacity and the Returns to Governance

	(1) Completed indicators	(2) Completion share	(3) Indicators documented	(4) Attachment categories	(5) Hard evidence
Panel A: Authority allocation					
Mandate, below threshold	14.126**	0.093**	1.198	1.132**	1.039**
	(6.878)	(0.046)	(1.318)	(0.536)	(0.439)
Delegation, below threshold	-0.232	-0.001	-0.679	-0.312	-0.302
	(6.487)	(0.043)	(1.139)	(0.442)	(0.338)
Mandate, above threshold	25.996***	0.168***	2.151	1.111*	0.522
	(9.024)	(0.060)	(1.542)	(0.644)	(0.498)
Delegation, above threshold	33.440***	0.214***	4.643*	1.961**	1.577**
	(11.363)	(0.075)	(2.567)	(0.973)	(0.761)
Above threshold	-17.095*	-0.109*	-1.772	-0.606	-0.211
	(9.086)	(0.060)	(1.677)	(0.651)	(0.521)
Constant	81.216***	0.560***	4.096***	1.936***	1.049***
	(5.410)	(0.036)	(1.037)	(0.371)	(0.300)
R^2	0.230	0.340	0.211	0.215	0.224
p-value, arm equality below	0.009	0.010	0.064	0.003	< 0.001
p-value, arm equality above	0.496	0.528	0.322	0.370	0.147
p-value, change in arm gap	0.081	0.091	0.122	0.039	0.005
Panel B: Monitoring scope					
Comprehensive, below threshold	5.968	0.040	0.537	0.390	0.294
	(6.586)	(0.044)	(1.194)	(0.500)	(0.400)
Targeted carbon, below threshold	5.684	0.038	-0.411	0.188	0.221
	(6.827)	(0.045)	(1.207)	(0.463)	(0.370)
Comprehensive, above threshold	11.601	0.069	0.014	-0.009	-0.014
	(9.161)	(0.061)	(1.677)	(0.674)	(0.573)
Targeted carbon, above threshold	42.635***	0.278***	5.552***	2.615***	1.738***
	(8.790)	(0.058)	(1.732)	(0.661)	(0.527)
Above threshold	-19.358**	-0.123**	-2.198	-0.827	-0.406
	(8.644)	(0.057)	(1.607)	(0.611)	(0.498)
Constant	81.872***	0.564***	4.220***	2.001***	1.106***
	(5.301)	(0.035)	(1.033)	(0.365)	(0.295)
R^2	0.246	0.357	0.239	0.221	0.191
p-value, arm equality below	0.960	0.940	0.288	0.649	0.835
p-value, arm equality above	< 0.001	< 0.001	0.003	< 0.001	0.003
p-value, change in arm gap	0.004	0.003	0.002	0.001	0.008
Control mean, below threshold	79.406	0.548	3.882	1.865	1.000
Control mean, above threshold	68.867	0.484	3.200	1.667	1.000
Observations	725	725	725	725	725
Suppliers	150	150	150	150	150
Round fixed effects	Yes	Yes	Yes	Yes	Yes
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the regressions underlying Figures 8 and 9. The threshold is the top quartile of pre-treatment registered capital. Each arm effect is relative to control suppliers in the same capacity group. The above-threshold coefficient is the level difference between control suppliers above and below the cutoff. Outcome definitions follow Table A3. Both treatment arms are fully interacted with the capacity-group indicators and the group main effect is included. Standard errors clustered by supplier are in parentheses. The change-in-gap test compares the within-group arm differences across the two capacity groups. Control means are unadjusted supplier-round means of the dependent variable among controls in the corresponding group and regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A12: Reporting Roles and Supplier Engagement

	(1) Completed indicators	(2) Completion share	(3) Indicators documented	(4) Attachment categories	(5) Hard evidence
Panel A: Authority allocation					
Mandate, staff respondent	17.440***	0.115***	1.442	1.236**	0.930**
	(6.255)	(0.041)	(1.218)	(0.482)	(0.371)
Mandate, executive respondent	-1.002	-0.012	-0.643	-0.709	-0.114
	(12.875)	(0.086)	(1.457)	(0.980)	(0.969)
Delegation, staff respondent	6.772	0.044	-0.297	0.040	-0.058
	(6.141)	(0.040)	(1.191)	(0.459)	(0.339)
Delegation, executive respondent	-11.167	-0.073	0.594	-0.746	-0.291
	(12.696)	(0.085)	(1.192)	(0.854)	(0.853)
Executive respondent	-0.918	-0.006	-2.077	0.276	-0.147
	(12.279)	(0.082)	(1.345)	(0.833)	(0.843)
Constant	79.691***	0.550***	4.399***	1.934***	1.137***
	(4.745)	(0.031)	(0.968)	(0.348)	(0.260)
R^2	0.240	0.350	0.222	0.213	0.199
Panel B: Monitoring scope					
Comprehensive, staff respondent	6.392	0.041	0.138	0.358	0.237
	(6.027)	(0.040)	(1.186)	(0.491)	(0.376)
Comprehensive, executive respondent	-4.988	-0.035	-0.596	-1.075	-0.592
	(12.491)	(0.083)	(1.296)	(0.862)	(0.861)
Targeted, staff respondent	16.359**	0.108**	0.832	0.787*	0.524
	(6.273)	(0.041)	(1.235)	(0.465)	(0.347)
Targeted, executive respondent	-2.669	-0.023	0.992	0.284	0.952
	(12.487)	(0.083)	(1.378)	(0.967)	(0.964)
Executive respondent	-1.254	-0.008	-2.173	0.204	-0.219
	(12.016)	(0.080)	(1.333)	(0.813)	(0.833)
Constant	79.786***	0.551***	4.413***	1.946***	1.150***
	(4.724)	(0.031)	(0.968)	(0.345)	(0.257)
R^2	0.236	0.347	0.210	0.187	0.177
Control mean, staff respondent	79.283	0.549	4.266	1.918	1.109
Control mean, executive respondent	78.500	0.542	3.067	2.333	1.100
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: This table presents arm effects relative to 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 A3. All specifications include reporting-round and randomization-stratum fixed effects. Robust standard errors clustered by supplier are in parentheses. Control means are unadjusted means of the dependent variable among controls in the corresponding respondent group and regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A13: 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)
Control mean	0.756	0.756	0.459	0.459
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 presents 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. Control means are unadjusted means of the dependent variable among controls in the corresponding regression estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table A14: 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
Control mean	8.206	5.156	0.024	98.471	0.621	0.065
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 presents 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 on-time delivery and return 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. Control means are unadjusted means of each dependent variable on its regression scale among controls in the corresponding estimation sample. Significance stars denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.