Delegation, Monitoring, and Supplier Engagement: A Field Experiment in Supply Chain Governance*

Hao Chen† Guojun He‡ Qinrui Xiahou§

May 2026

Abstract

We conduct a 2×2 factorial field experiment with a leading Chinese apparel brand and 155 of its suppliers to study how governance structure shapes supplier compliance with multi-dimensional reporting mandates. We independently vary decision authority (mandate versus delegation of KPI selection) and monitoring scope (comprehensive versus targeted carbon audit). Contrary to classical organizational theory, simpler governance structures generate substantially higher engagement among capacity-constrained small and medium-sized enterprises. Mandating a fixed KPI set reduces the missing-data rate by 10.4 percentage points relative to the control mean, while delegating KPI choice yields a negligible effect. Targeted carbon audits raise documentation coverage by 0.857 additional categories, roughly four times the effect of comprehensive audits. The two interventions are mutually reinforcing, with gains concentrated among smaller suppliers and in environmental reporting, where measurement standards are clearer and coordination costs are lower. We develop a model of bandwidth-constrained compliance that rationalizes these patterns and draw implications for the private regulation of externalities in global value chains.

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

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

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

1 Introduction

Multinational corporations have fundamentally reorganized modern production, coordinating complex supplier networks across borders to capture substantial efficiency gains (Antràs and Chor, 2013). However, the geographical dispersion and organizational fragmentation that generate these gains also create severe governance frictions (Antràs and Helpman, 2004). Lead firms must secure cooperation from independent suppliers on multi-dimensional and often non-contractible objectives, even though those suppliers possess private information about local constraints and may not share the principal's priorities (Baker et al., 2002; Gereffi et al., 2005; Gibbons, 2005; Boudreau et al., 2023). In such settings, a central organizational question is how the principal should allocate decision-making authority and structure performance monitoring so as to obtain compliance from agents outside the firm's formal boundary.

Organizational economics offers well-developed benchmark predictions on both dimensions. On the vertical dimension, delegation theory establishes that granting agents the authority to tailor tasks allows them to exploit private information about local costs, improving effort allocation beyond what a principal could achieve through mandate (Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008). On the horizontal dimension, the multitasking framework demonstrates that broader performance evaluation can prevent inefficient substitution of effort toward easily verifiable metrics (Holmström and Milgrom, 1991). Both frameworks find robust empirical support within firm boundaries, where delegation improves firm productivity (Acemoglu et al., 2007; Bloom et al., 2012; Bandiera et al., 2021; Kala, 2024) and workers respond to performance measurement and evaluation design (Hong et al., 2018; Manthei and Sliwka, 2019; Gosnell et al., 2020; Rasul and Rogger, 2018).

These benchmark predictions, however, rely on an implicit assumption that agents

possess sufficient organizational bandwidth to process complex instructions and execute multi-dimensional tasks. This assumption frequently fails in supply chains dominated by small and medium-sized enterprises (SMEs) in developing economies. Managers in such firms typically operate with thin administrative structures and face internal organizational frictions when asked to implement new practices (Bloom et al., 2013; Atkin et al., 2017; McKenzie and Woodruff, 2017). When a principal imposes complex or open-ended requirements on bandwidth-constrained agents, the main failure may not be strategic effort misallocation but a breakdown in basic engagement (Ocasio, 1997; Mani et al., 2013; Bhargava and Manoli, 2015; Kaur et al., 2025; Finkelstein and Notowidigdo, 2019). If capacity constraints bind, simpler and more directive governance structures may outperform more flexible or more comprehensive arrangements, reversing the standard predictions of delegation and multitasking.

We test this hypothesis through a randomized controlled trial (RCT) conducted in partnership with a leading Chinese apparel brand and 155 of its suppliers. We leverage the brand's recent introduction of environmental, social, and governance (ESG) reporting requirements as a natural, multi-dimensional compliance environment. Within this setting, we independently vary two dimensions of governance structure. Vertically, suppliers receive either a mandate of eight fixed key performance indicators (KPIs) or the delegated authority to select eight KPIs from a broader menu. Horizontally, suppliers face either comprehensive third-party monitoring of all submitted data or targeted monitoring restricted to carbon-related metrics. Crossing these dimensions yields a 2×2 factorial design, augmented by a control group. Because the experiment runs entirely within the lead firm's digital supply chain platform, suppliers interact with the system in the ordinary course of business, thus eliminating experimenter demand effects.

We evaluate supplier response along two complementary margins. *Basic engage-*

ment captures whether suppliers clear the initial reporting hurdle, measured by the completeness of submitted ESG fields. *Advanced engagement* captures the depth of compliance effort, measured by the breadth of supporting documentation uploaded. We find that simplified governance structures substantially outperform their more complex counterparts among capacity-constrained SMEs. On the vertical dimension, mandating a fixed KPI set reduces the missing-data rate by 10.4 percentage points relative to a 45.2 percent control mean, while delegating KPI choice yields no statistically or economically meaningful reduction. On the horizontal dimension, targeted carbon audits raise documentation coverage by 0.857 attachment categories, roughly four times the effect of comprehensive audits. The two simplifications are mutually reinforcing: the bundle combining mandated KPIs with targeted monitoring produces the strongest overall engagement, reducing the missing-data rate by 11.5 percentage points and increasing documentation coverage by 1.338 categories.

Heterogeneity analysis confirms that organizational capacity is central to these patterns. Among the less-capable majority of suppliers, neither delegation nor comprehensive monitoring generates significant engagement gains. Among the most capable firms, however, delegation improves advanced engagement and comprehensive monitoring begins to produce meaningful documentation effects. This pattern indicates that the standard predictions hold conditional on sufficient organizational slack. Additionally, treatment effects are concentrated in environmental metrics rather than social or governance dimensions, consistent with the idea that environmental metrics are more standardized and less coordination-intensive for constrained suppliers.

This paper makes three main contributions. First, we provide direct causal evidence on the boundary conditions of classical organizational economics in an inter-firm setting. The theories of delegation (Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008; Mookherjee, 2006) and multitasking (Holmström and

Milgrom, 1991) are foundational to the economics of firm organization, yet direct empirical tests have largely been confined to lab or within-firm settings where agents typically have sufficient capacity to respond as theory predicts (Fehr et al., 2013; Hong et al., 2018; Manthei and Sliwka, 2019; Bandiera et al., 2021; Kala, 2024). We show that when agents face binding capacity constraints across organizational boundaries, standard predictions reverse. This highlights organizational bandwidth as a critical constraint that determines which governance regime succeeds.

Second, the paper contributes to the field experimental literature on management and firm capabilities in developing economies. Prior work establishes that managerial and organizational capabilities are important determinants of firm performance (Bloom and Van Reenen, 2007; Bloom et al., 2013, 2020; Atkin et al., 2017; Verhoogen, 2023; Adhvaryu et al., 2022) and that multinationals can enforce inter-firm standards (Distelhorst et al., 2017; Boudreau, 2024; Alfaro-Ureña et al., 2022; Amengual and Distelhorst, 2025). We advance this literature by demonstrating that the specific structure of governance, rather than merely its presence, determines compliance outcomes, and we identify how this structure interacts with agent capacity.

Third, our findings carry direct implications for the design of private governance in global value chains (Gereffi et al., 2005; Locke, 2013). As lead firms increasingly function as *de facto* regulators in environments where public enforcement is limited, the effectiveness of compliance programs depends critically on matching governance complexity to agent capacity (Kitzmüeller and Shimshack, 2012; Dimson et al., 2015). For the SME-dominated supply chains that characterize most emerging-economy manufacturing, simple and directive governance structures are not a practical compromise but often the only design likely to generate meaningful compliance.

The paper proceeds as follows. Section 2 develops the theoretical framework and derives testable predictions. Section 3 describes the institutional context. Section 4

presents the experimental design and empirical strategy. Section 5 reports the main results. Section 6 discusses theoretical and policy implications. Section 7 concludes.

2 Theoretical Framework

We develop a model of governance design under asymmetric information and limited organizational bandwidth. The goal is to embed the classical frameworks of authority allocation (Aghion and Tirole, 1997) and multitask monitoring (Holmström and Milgrom, 1991) within a single principal-agent problem, recover their standard predictions as limiting cases under explicitly stated conditions, and then demonstrate that both predictions implicitly rely on non-binding managerial bandwidth. Once this premise is relaxed to reflect the capacity constraints typical of SME suppliers, the governance structures that maximize supplier engagement can diverge sharply from the classical rankings. Section 2.1 lays out the environment, governance structure, and the behavioral assumptions that underpin both the classical and bandwidth-constrained analyses. Section 2.2 derives the classical benchmarks as limiting cases of the model. Section 2.3 introduces bandwidth frictions and characterizes the conditions under which both classical predictions fail. Section 2.4 states five formal propositions, and Section 2.5 maps them to testable hypotheses.

2.1 Model Setup

Agents, tasks, and bandwidth. A lead firm (principal) oversees a continuum of suppliers (agents), each of whom allocates managerial effort across J reporting dimensions indexed by $j \in \mathcal{J} = \{\text{Env, Soc, Gov}\}$ (environmental, social, and governance). Each dimension is characterized by a tractability parameter $\theta_j \in (0, 1]$, measuring how efficiently a unit of managerial attention converts into verifiable documentation on

that dimension. Environmental metrics such as energy consumption can typically be retrieved from a single utility invoice, whereas social and governance disclosures require qualitative documentation coordinated across HR, legal, and compliance functions. We order dimensions by decreasing tractability, $\theta_{\text{Env}} > \theta_{\text{Soc}} > \theta_{\text{Gov}} > 0$, and normalize $\theta_{\text{Env}} = 1$ without loss of generality.

Each supplier is endowed with organizational bandwidth $B > 0$, representing the stock of managerial attention available for all non-production activities during the reporting period. Bandwidth is heterogeneous across the supplier population and distributed according to a continuously differentiable distribution F with support $[0, \bar{B}]$, $\bar{B} > 0$.

Assumption 1 (Bandwidth heterogeneity)**.** F is continuously differentiable with strictly positive density $f(b) = F'(b) > 0$ for all $b \in (0, \bar{B})$.

The distribution F represents the population of suppliers in the network. Under stratified random assignment, treatment groups are draws from the same F, so expected engagement differences across arms identify the causal effects estimated in Section 4.

Governance structure. The principal designs a governance structure $G = (a, m)$, where $a \in \{M, D\}$ specifies the authority allocation over KPI selection (Mandate or Delegate) and $m \in \{C, T\}$ specifies the scope of third-party monitoring (Comprehensive or Targeted). The authority dimension determines *who* chooses what to report; the monitoring dimension determines *which* reports will be externally verified.

Authority allocation. The reporting program requires each supplier to file data on at least $N = 8$ key performance indicators drawn from the platform's full menu of metrics; suppliers may voluntarily report on additional indicators beyond

this minimum. Under mandate ($a = M$), the principal pre-specifies the required KPIs. Under delegation ($a = D$), the supplier self-selects at least eight indicators that best match its local operational capabilities.

Because the supplier possesses private information about the tractability of its own operations, delegation enables a selection gain. The expected tractability of self-selected indicators, $\bar{\theta} \equiv \mathbb{E}[\theta_j \mid \text{optimal local selection}]$, weakly exceeds the expected tractability of the mandated set, $\tilde{\theta} \equiv \mathbb{E}[\theta_j \mid \text{principal's selection}]$, with $\bar{\theta} > \tilde{\theta}$ strictly whenever local information is non-trivial. This information advantage is the classical rationale for delegation (Aghion and Tirole, 1997; Dessein, 2002).

The selection process, however, is not costless. Evaluating a menu of indicators, verifying data availability across organizational units, and choosing the optimal subset requires dedicated managerial attention. We model this as a bandwidth cost $\phi(D) = \phi > 0$ that the agent must expend before any reporting activity can begin. Under mandate no selection decision is required: $\phi(M) = 0$.

For the selection gain to benefit the principal, the supplier's preferred choices must genuinely serve the principal's objectives rather than merely minimize reporting burden. We therefore impose:

Assumption 2 (Preference congruence)**.** The agent's preferred KPI selection weakly maximizes the principal's payoff conditional on the agent's private information, so that the selection gain $\bar{\theta} - \tilde{\theta} \geq 0$ represents a genuine improvement in reporting quality rather than strategic manipulation.

In our experimental setting, Assumption 2 is grounded in design: prior to the experiment, the lead firm held a formal supply chain conference communicating the strategic rationale for ESG reporting to all suppliers. This shared understanding ensures that supplier KPI choices are informative to the principal (Aghion and Tirole,

1997; Dessein, 2002). The assumption concerns KPI selection alone and places no restriction on suppliers' capacity to act on that selection; it therefore operates on a dimension orthogonal to the bandwidth heterogeneity that drives the main results.

Monitoring scope and the initiation threshold. We capture the cognitive burden of the monitoring scope through an *initiation threshold* $\kappa(m) > 0$: the minimum bandwidth that must be committed before the supplier can begin filing any reports. This threshold reflects the pre-task coordination load of understanding evidentiary requirements, organizing cross-departmental data collection, and mapping the verification scope onto available internal resources.

Under comprehensive monitoring ($m = C$), the agent anticipates that all three ESG dimensions will be verified. Preparing for this broad scope requires coordinating across functionally disconnected units before any individual field can be completed, raising the initiation threshold to $\kappa(C)$. Under targeted monitoring ($m = T$), verification is restricted to a single focal dimension: in our experimental setting, the carbon audit. The clarity of this anchor reduces the pre-task coordination burden and lowers the initiation threshold to $\kappa(T) < \kappa(C)$ (Ocasio, 1997; Dessein et al., 2016).

For comprehensive monitoring to motivate broader reporting effort rather than mere compliance on the verified dimensions, the principal must value engagement across all ESG domains. If the principal's payoff depended on a single dimension, targeted monitoring would be weakly optimal regardless of the agent's bandwidth, since concentrating verification on that dimension would preserve incentives without sacrificing any payoff-relevant effort. We therefore impose:

Assumption 3 (Multi-dimensional valuation)**.** The principal's payoff $\Pi(a, m)$ is strictly increasing in engagement on each dimension $j \in \mathcal{J}$, and no single dimension constitutes a bottleneck that renders improvements on others valueless.

In our setting, Assumption 3 is grounded in how third-party ESG rating agencies evaluate lead firms. Major rating providers aggregate performance across environmental, social, and governance pillars into a composite score, and a persistent shortfall on any single pillar reduces the overall rating even when other dimensions perform well. The lead firm therefore has genuine incentives to promote balanced supplier engagement across dimensions.

Assumption 4 (Threshold and cost ordering)**.** $\phi(M) = 0$, $\phi(D) = \phi > 0$, $\kappa(C) > \kappa(T) > 0$, and $\phi + \kappa(C) \leq \bar{B}$.

The final condition ensures that the most demanding governance structure remains feasible for at least some agents in the population.

Engagement outcomes. After incurring governance overhead, the supplier's *effective execution bandwidth* is:

$$B^e(G,\, B) \;=\; \max\big\{B - \phi(a) - \kappa(m),\; 0\big\}. \tag{1}$$

Basic engagement captures the breadth of platform participation. In the formal model, we represent it as a threshold indicator,

$$\mathcal{E}^B(G,\, B) \;=\; \mathbf{1}\big[B \;\geq\; \phi(a) + \kappa(m)\big], \tag{2}$$

which is equal to one when the supplier clears the combined overhead. This binary representation preserves analytical tractability; the empirical counterpart is the continuous fraction of the 108 platform indicators left unfilled, which captures both required-field completion and voluntary reporting on non-required fields.

Advanced engagement measures the depth of discretionary documentation: sup-

porting materials uploaded to corroborate reported figures, an activity that requires substantially more cross-departmental coordination than data entry alone. For suppliers who clear the threshold:

$$\mathcal{E}^A(G,\, B) \;=\; h\big(B^e(G,\, B)\big) \cdot \theta(a), \tag{3}$$

where $h : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies $h(0) = 0$, $h' > 0$, $h'' \leq 0$, and $\theta(M) = \tilde{\theta}$, $\theta(D) = \bar{\theta} \geq \tilde{\theta}$. The function $h(B^e)$ captures the quantity of documentation produced from available bandwidth, with diminishing marginal returns. The multiplier $\theta(a)$ captures the informational quality of the documented indicators: under delegation, self-selected KPIs are more closely matched to local operational realities (Assumption 2), yielding more informative corroborating materials per unit of effort. Empirically, advanced engagement counts the number of ESG categories for which at least one supporting document is uploaded.

The two forces in equation (3) push in opposite directions under delegation. The bandwidth component $h(B^e)$ is reduced by the decision cost $\phi(D)$, since $B^e(D, m, B) < B^e(M, m, B)$ for any supplier who engages under both structures. The quality component $\theta(a)$ is raised by the selection gain $\bar{\theta} \geq \tilde{\theta}$. Whether mandate or delegation produces higher expected advanced engagement depends on which force dominates, and this in turn depends on the supplier's organizational bandwidth. This tension generates the capability-heterogeneity results in Proposition 4.

Dimension-specific engagement. Conditional on clearing the aggregate initiation threshold, the supplier begins reporting on individual dimensions. We model dimension-specific participation using a tractability-scaled threshold: $\kappa_j(m) = \kappa(m)/\theta_j$. Since $\theta_{\text{Env}} = 1$ by normalization, $\kappa_{\text{Env}}(m) = \kappa(m)$ is the lowest dimension-specific threshold;

social and governance dimensions impose higher thresholds, $\kappa_{\text{Soc}}(m) > \kappa(m)$ and $\kappa_{\text{Gov}}(m) > \kappa(m)$. Dimension-specific basic engagement is:

$$\mathcal{E}_j^B(G,\, B) \;=\; \mathbf{1}\big[B \;\geq\; \phi(a) + \kappa_j(m)\big]. \tag{4}$$

A supplier who clears $\phi(a) + \kappa(m)$ can immediately engage with environmental reporting; social and governance dimensions require additional bandwidth. On the advanced engagement margin, tractability exerts an additional force: producing verifiable supporting documentation requires effort that converts into output at rate θ_j, so less tractable dimensions impose both a higher threshold to clear and a lower documentation yield per unit of effort conditional on clearing it. Effort concentration on environmental metrics is therefore the rational response of bandwidth-constrained agents across both margins, reflecting effort-substitution in the spirit of Holmström and Milgrom (1991).

2.2 Classical Benchmarks

This section derives the classical benchmarks as formal limiting cases of the model. In the frictionless limit where bandwidth is non-binding, the model recovers two standard predictions under explicitly stated parametric conditions, providing the benchmark against which the bandwidth-constrained departures in Section 2.3 are assessed.

The classical case for delegation. Consider the limit $\phi \to 0$, so that KPI selection becomes costless and effective bandwidth converges to $B - \kappa(m)$ under both authority structures. Equation (3) then gives $\mathcal{E}^A = h(B - \kappa(m)) \cdot \theta(a)$. Since

$\theta(D) = \bar{\theta} \geq \tilde{\theta} = \theta(M)$, advanced engagement is weakly higher under delegation:

$$\lim_{\phi \to 0} \mathbb{E}\big[\mathcal{E}^A \mid D,\, m\big] \;\; \geq \;\; \mathbb{E}\big[\mathcal{E}^A \mid M,\, m\big], \tag{5}$$

with strict inequality whenever $\bar{\theta} > \tilde{\theta}$. This is the classical result of Aghion and Tirole (1997) and Dessein (2002): granting real authority to the better-informed agent raises performance quality when exercising that authority is costless and preferences are sufficiently aligned. The prediction holds for any level of supplier capability B in this limit, because the threshold advantage of mandate vanishes when $\phi = 0$.

The classical case for comprehensive monitoring. Now consider the limit $\kappa(C) \to \kappa(T)$, so that both monitoring structures impose the same initiation threshold and therefore the same effective execution bandwidth. Since $\theta(a)$ does not depend on m, equation (3) implies $\mathbb{E}[\mathcal{E}^A \mid a, C] = \mathbb{E}[\mathcal{E}^A \mid a, T]$. In this frictionless limit, the distinction between comprehensive and targeted monitoring does not operate through the reduced-form engagement equation, but through the principal's payoff $\Pi(a, m)$ over the allocation of effort across dimensions. Under Assumption 3, the principal values engagement on all dimensions, so comprehensive monitoring is weakly preferred because it prevents distortion toward the single verified dimension:

$$\kappa(C) = \kappa(T) \;\; \Longrightarrow \;\; \Pi(a,\, C) \;\geq\; \Pi(a,\, T). \tag{6}$$

This recovers the central result of Holmström and Milgrom (1991): when the principal values performance across multiple substitutable tasks, broader monitoring weakly dominates targeted monitoring because it mitigates effort distortion toward the most salient or easily measured activity.

Equations (5) and (6) constitute the classical benchmark: delegation weakly

dominates mandate and comprehensive monitoring weakly dominates targeted. Both results rely on $\phi = 0$ and $\kappa(C) = \kappa(T)$, which together encode the premise that governance overhead is negligible relative to agents' available bandwidth. Section 2.3 shows how relaxing this premise can overturn both benchmark rankings.

2.3 Bandwidth Constraints and Benchmark Departures

When decision-making requires effort and task breadth raises the cognitive cost of initiating, the classical engagement rankings need not hold. The two mechanisms operate independently and reinforce each other when both frictions are present.

Why delegation can depress engagement. When $\phi > 0$, a supplier with $B \in [\kappa(m),\, \phi + \kappa(m))$ clears the initiation threshold under mandate but falls short under delegation. For this segment, delegation converts a potential compliance into a non-response. The aggregate comparison therefore depends on the mass of suppliers in this interval: for SMEs concentrated in the lower tail of F, this interval is densely populated and the population-level effect of delegation is negative. The selection gain $\bar{\theta} - \tilde{\theta}$ accrues only to suppliers who would have engaged under both structures; it does not compensate for the loss of basic engagement among the constrained majority.

Why comprehensive monitoring can depress engagement. When $\kappa(C) > \kappa(T)$, a supplier with $B \in [\phi(a) + \kappa(T),\, \phi(a) + \kappa(C))$ can initiate reporting under targeted monitoring but not under comprehensive monitoring. Comprehensive monitoring therefore excludes a segment of capacity-constrained suppliers by raising the initiation threshold. Targeted monitoring, by anchoring attention on a single tractable dimension, lowers the effective entry barrier and raises expected engagement across both margins when integrated over the supplier distribution.

The governance bundle. The two frictions are additive in their effect on the initiation threshold. Moving from (D, C), with threshold $\phi + \kappa(C)$, to (M, T), with threshold $\kappa(T)$, reduces overhead by $\phi + \kappa(C) - \kappa(T)$. Both savings apply simultaneously under (M, T), making it the governance structure that minimizes the initiation threshold and maximizes expected basic engagement across the supplier distribution.

2.4 Formal Results

The five propositions below formalize the arguments above. Each maps to one testable hypothesis in Section 2.5.

Proposition 1 (Guidance Effect)**.** *Under Assumptions 1, 2, and 4, for any monitoring scope m, mandate strictly dominates delegation on basic engagement:*

$$\Pr\big[\mathcal{E}^B = 1 \mid M,\, m\big] \;>\; \Pr\big[\mathcal{E}^B = 1 \mid D,\, m\big].$$

For advanced engagement, mandate also dominates delegation provided that the extensive-margin gain from suppliers who engage only under mandate outweighs the intensive-margin quality advantage of delegation among suppliers who would engage under either structure. A sufficient condition is

$$\tilde{\theta} \int_{\kappa(m)}^{\phi+\kappa(m)} h(b-\kappa(m))\, f(b)\, db \;>\; \int_{\phi+\kappa(m)}^{\bar{B}} \big[\bar{\theta}\, h(b-\phi-\kappa(m)) - \tilde{\theta}\, h(b-\kappa(m))\big]^{+} f(b)\, db, \tag{7}$$

under which

$$\mathbb{E}\big[\mathcal{E}^A \mid M,\, m\big] \;>\; \mathbb{E}\big[\mathcal{E}^A \mid D,\, m\big].$$

Condition (7) is satisfied when F places sufficient mass near the threshold interval $[\kappa(m),\, \phi + \kappa(m))$, as in supplier populations facing tight organizational bandwidth

constraints. As $\phi \to 0$, basic-engagement probabilities equalize and the advanced-engagement comparison reverts to the classical benchmark $\mathbb{E}[\mathcal{E}^A \mid D, m] \geq \mathbb{E}[\mathcal{E}^A \mid M, m]$ from equation (5).

Proof. For basic engagement, a supplier with $B \in [\kappa(m), \phi+\kappa(m))$ clears the initiation threshold under mandate but not under delegation. Both thresholds $\kappa(m)$ and $\phi + \kappa(m)$ lie in $(0, \bar{B})$ by Assumption 4, where $f > 0$ by Assumption 1; hence $F(\phi + \kappa(m)) > F(\kappa(m))$, so $1 - F(\kappa(m)) > 1 - F(\phi + \kappa(m))$, establishing the basic engagement ranking.

For advanced engagement, the mandate advantage $\mathbb{E}[\mathcal{E}^A \mid M, m] - \mathbb{E}[\mathcal{E}^A \mid D, m]$ decomposes into two terms. First, suppliers with $B \in [\kappa(m),\ \phi + \kappa(m))$ engage under mandate but not under delegation, contributing $\tilde{\theta} \int_{\kappa(m)}^{\phi+\kappa(m)} h(b - \kappa(m)) f(b)\, db > 0$. Second, for $B \geq \phi + \kappa(m)$, both structures produce basic engagement and the comparison is $\tilde{\theta}\, h(B-\kappa(m))$ versus $\bar{\theta}\, h(B-\phi-\kappa(m))$, whose sign depends on parameter magnitudes; $[\,\cdot\,]^+$ denotes the positive part. Condition (7) is a sufficient condition that the first term dominates. In our empirical setting, condition (7) is plausible because the supplier population is concentrated among SMEs whose bandwidth distribution F places substantial mass near the lower end of the support. This concentration ensures that the interval $[\kappa(m),\ \phi + \kappa(m))$ is densely populated, satisfying the condition for populations with a right-skewed capability distribution.

As $\phi \to 0$, the first term vanishes, B^e equalizes across structures, and the comparison reduces to $\bar{\theta} \geq \tilde{\theta}$, yielding (5). □

Proposition 2 (Salience Effect)**.** *Under Assumptions 1, 3, and 4, for any authority allocation a:*

$$\Pr\big[\mathcal{E}^B = 1 \mid a,\, T\big] \;>\; \Pr\big[\mathcal{E}^B = 1 \mid a,\, C\big] \quad \textit{and} \quad \mathbb{E}\big[\mathcal{E}^A \mid a,\, T\big] \;>\; \mathbb{E}\big[\mathcal{E}^A \mid a,\, C\big].$$

As $\kappa(C) \to \kappa(T)$, both strict inequalities collapse to equalities in the reduced-form model; the weak dominance of comprehensive monitoring in the classical benchmark (6) reflects the effort-allocation channel captured by $\Pi(a, m)$ rather than by equation (3).

Proof. For basic engagement, a supplier with $B \in [\phi(a)+\kappa(T), \phi(a)+\kappa(C))$ clears the initiation threshold under targeted monitoring but not under comprehensive monitoring. Both thresholds lie in $(0, \bar{B})$ by Assumption 4, where $f > 0$ by Assumption 1; hence $1 - F(\phi(a) + \kappa(T)) > 1 - F(\phi(a) + \kappa(C))$, establishing the basic engagement ranking.

For advanced engagement, $\theta(a)$ is identical under both monitoring structures for fixed a, so the comparison reduces to $h(B - \phi(a) - \kappa(T))$ versus $h(B - \phi(a) - \kappa(C))$ for $B \geq \phi(a) + \kappa(C)$. Since $\kappa(T) < \kappa(C)$, targeted monitoring yields strictly higher effective bandwidth, so $h(B - \phi(a) - \kappa(T)) > h(B - \phi(a) - \kappa(C))$ for all such B. For $B \in [\phi(a) + \kappa(T), \phi(a) + \kappa(C))$, targeted monitoring contributes strictly positive advanced engagement while comprehensive monitoring contributes zero. As $\kappa(C) \to \kappa(T)$, both inequalities collapse to equalities within the reduced-form model; the classical benchmark (6) then holds for the principal's full payoff Π, as established in the multitasking literature.

Note that the unconditional ranking in Proposition 2 arises because monitoring scope does not affect $\theta(a)$: unlike the authority dimension, where delegation involves a quality–quantity trade-off ($\bar{\theta} > \tilde{\theta}$ versus reduced bandwidth), the monitoring dimension operates purely through the initiation threshold. There is no countervailing quality channel, so the lower-threshold structure dominates unambiguously. □

Remark 1. Propositions 1 and 2 show that the classical predictions are limiting cases of the general model when $\phi \to 0$ and $\kappa(C) \to \kappa(T)$, respectively. The bandwidth-constrained departures documented in Section 5 are therefore not contradictions of the classical frameworks but extensions that apply in the region of high ϕ, high

$\kappa(C) - \kappa(T)$, and low B that characterizes SME-dominated supply networks.

Proposition 3 (Bandwidth Minimization Bundle)**.** *Under Assumptions 1 and 4, the bundle* (M, T) *uniquely minimizes the initiation threshold among all four governance structures and therefore uniquely maximizes expected basic engagement:*

$$\Pr\big[\mathcal{E}^B = 1 \mid M, T\big] \;>\; \Pr\big[\mathcal{E}^B = 1 \mid G'\big] \quad \textit{for all } G' \neq (M, T).$$

Under condition (7) evaluated at $m = T$*,* (M, T) *also uniquely maximizes expected advanced engagement:*

$$\mathbb{E}[\mathcal{E}^A \mid M, T] \;>\; \mathbb{E}[\mathcal{E}^A \mid G'] \quad \textit{for all } G' \neq (M, T).$$

Proof. For basic engagement, the initiation thresholds are $\kappa(T)$ under (M,T); $\kappa(C)$ under (M,C); $\phi + \kappa(T)$ under (D,T); and $\phi + \kappa(C)$ under (D,C). By Assumption 4, $\kappa(T)$ is strictly the smallest, so (M,T) generates the highest basic engagement probability by strict monotonicity of F.

For advanced engagement within the mandate arm, $\theta(a) = \tilde{\theta}$ is held constant and $B^e(M,T,B) = B - \kappa(T) > B - \kappa(C) = B^e(M,C,B)$; since h is strictly increasing, (M,T) strictly dominates (M,C) on both margins. The same argument gives (D,T) strictly dominating (D,C). For the cross-arm comparison, mandate has higher B^e but lower $\theta(a)$, which is the same trade-off as in Proposition 1; under condition (7) at $m = T$, the bandwidth advantage dominates. Transitivity then yields the full ranking. □

Proposition 4 (Capability Threshold)**.** *The guidance and salience effects have the following capability-dependent properties.*

For basic engagement, the advantage of mandate over delegation is concentrated

in the bandwidth interval $[\kappa(m),\, \phi + \kappa(m))$*, while the advantage of targeted over comprehensive monitoring is concentrated in* $[\phi(a) + \kappa(T),\, \phi(a) + \kappa(C))$*. In each case, the lower-overhead structure converts non-participation into participation only for suppliers whose bandwidth lies in the corresponding threshold interval.*

For advanced engagement, define $\Delta_A(B) \equiv \mathcal{E}^A(M, m, B) - \mathcal{E}^A(D, m, B)$ *in the authority dimension. For* $B \in [\kappa(m),\, \phi + \kappa(m))$*, one has* $\Delta_A(B) = \tilde{\theta}\, h(B - \kappa(m))$*, so the mandate advantage rises strictly with* B*. For* $B \geq \phi + \kappa(m)$*, one has* $\Delta'_A(B) < 0$*, so the mandate advantage contracts as bandwidth rises further.*

Similarly, define $\Delta_M(B) \equiv \mathcal{E}^A(a, T, B) - \mathcal{E}^A(a, C, B)$ *in the monitoring dimension. For* $B \in [\phi(a) + \kappa(T),\, \phi(a) + \kappa(C))$*, one has* $\Delta_M(B) = \theta(a)\, h(B - \phi(a) - \kappa(T))$*, so the targeted-monitoring advantage rises strictly with* B*. For* $B \geq \phi(a) + \kappa(C)$*, one has* $\Delta'_M(B) \leq 0$*, so the targeted-monitoring advantage weakens as bandwidth rises further.*

Hence, on the advanced engagement margin, the advantage of the lower-overhead structure is concentrated around the relevant capability threshold: it increases over the intermediate range where only that structure is feasible, and decreases once both structures become feasible.

Proof. For basic engagement, equation (2) gives $\mathcal{E}^B(M, m, B) = \mathbf{1}[B \geq \kappa(m)]$ and $\mathcal{E}^B(D, m, B) = \mathbf{1}[B \geq \phi + \kappa(m)]$, so the mandate advantage is exactly $\mathbf{1}[B \in [\kappa(m),\, \phi + \kappa(m))]$. The monitoring analogue is $\mathcal{E}^B(a, T, B) = \mathbf{1}[B \geq \phi(a) + \kappa(T)]$ and $\mathcal{E}^B(a, C, B) = \mathbf{1}[B \geq \phi(a) + \kappa(C)]$, so the targeted-monitoring advantage is exactly $\mathbf{1}[B \in [\phi(a) + \kappa(T),\, \phi(a) + \kappa(C))]$.

For advanced engagement over these intermediate-capability regions, equation (3) implies $\mathcal{E}^A(G, B) = h(B^e(G, B))\theta(a)$ with $h(0) = 0$ and $h' > 0$. Hence, for $B \in [\kappa(m),\, \phi + \kappa(m))$, one has $B^e(M, m, B) = B - \kappa(m) > 0$ and $B^e(D, m, B) = 0$, so $\mathcal{E}^A(M, m, B) > 0 = \mathcal{E}^A(D, m, B)$; moreover, $\mathcal{E}^A(M, m, B)$ rises strictly with B on

this interval. Likewise, for $B \in [\phi(a) + \kappa(T),\, \phi(a) + \kappa(C))$, one has $B^e(a, T, B) = B - \phi(a) - \kappa(T) > 0$ and $B^e(a, C, B) = 0$, so $\mathcal{E}^A(a, T, B) > 0 = \mathcal{E}^A(a, C, B)$, again with strict monotonicity in B.

Now consider the region where both structures are feasible. In the authority dimension, for $B \geq \phi + \kappa(m)$, we have $\Delta_A(B) = \tilde{\theta}\, h(B - \kappa(m)) - \bar{\theta}\, h(B - \phi - \kappa(m))$. Differentiating and using $h'' \leq 0$ yields $\Delta'_A(B) = \tilde{\theta}\, h'(B - \kappa(m)) - \bar{\theta}\, h'(B - \phi - \kappa(m)) \leq (\tilde{\theta} - \bar{\theta})\, h'(B - \phi - \kappa(m)) < 0$, because $\bar{\theta} > \tilde{\theta}$ and $h' > 0$.

In the monitoring dimension, for $B \geq \phi(a) + \kappa(C)$, we have $\Delta_M(B) = \theta(a)\big[h(B - \phi(a) - \kappa(T)) - h(B - \phi(a) - \kappa(C))\big]$. Differentiating gives $\Delta'_M(B) = \theta(a)\big[h'(B - \phi(a) - \kappa(T)) - h'(B - \phi(a) - \kappa(C))\big] \leq 0$, because $\kappa(C) > \kappa(T)$ implies $B - \phi(a) - \kappa(T) > B - \phi(a) - \kappa(C)$ and h' is non-increasing. Therefore, the lower-overhead advantage on advanced engagement contracts with capability in both dimensions.

The two dimensions are nevertheless not fully identical in structure. In the authority dimension, the comparison additionally involves the selection gain $\bar{\theta} > \tilde{\theta}$, so the contraction in $\Delta_A(B)$ may eventually lead to a reversal in favor of delegation at sufficiently high B. In the monitoring dimension, by contrast, the reduced-form engagement equation contains no analogous quality multiplier, so the proposition yields attenuation of the targeted-monitoring advantage without requiring reversal. □

Proposition 5 (Task Tractability)**.** *Let $\kappa_j(m) = \kappa(m)/\theta_j$ for $j \in \mathcal{J}$, with $\theta_{\text{Env}} = 1$ by normalization. Assume*

$$\phi + \kappa_{\text{Gov}}(C) < \bar{B}. \tag{8}$$

Then the dimension specific basic engagement probability $\Pr[\mathcal{E}_j^B = 1 \mid a, m]$ is strictly decreasing in $1/\theta_j$, so environmental engagement rates strictly exceed social and governance engagement rates. For fixed m, the guidance effect $\Pr[\mathcal{E}_j^B = 1 \mid M, m] - \Pr[\mathcal{E}_j^B = 1 \mid D, m]$ is largest at the environmental dimension under a unimodal

density f whose mode lies below $\kappa_{\text{Env}}(m)$. For fixed a, the salience effect $\Pr[\mathcal{E}_j^B = 1 \mid a, T] - \Pr[\mathcal{E}_j^B = 1 \mid a, C]$ is largest at the environmental dimension whenever $f(\zeta_j)/\theta_j$ is strictly decreasing in $1/\theta_j$, where $\zeta_j \in (\phi(a) + \kappa_j(T),\, \phi(a) + \kappa_j(C))$.

Proof. Condition (8) implies $\phi(a) + \kappa_j(m) < \bar{B}$ for all $j \in \mathcal{J}$ and (a, m), so all relevant thresholds lie in the support of F. Since $\kappa_j(m) = \kappa(m)/\theta_j$ is strictly increasing in $1/\theta_j$ and $\Pr[\mathcal{E}_j^B = 1 \mid a, m] = 1 - F(\phi(a) + \kappa_j(m))$, the dimension specific basic engagement probability is strictly decreasing in $1/\theta_j$, which gives the ordering across environmental, social, and governance dimensions.

For fixed m, the guidance effect is $F(\kappa_j(m) + \phi) - F(\kappa_j(m))$. By the mean value theorem, this equals $f(\xi_j)\phi$ for some $\xi_j \in (\kappa_j(m),\, \kappa_j(m) + \phi)$. If f is unimodal with mode below $\kappa_{\text{Env}}(m)$, then f is strictly decreasing over the relevant range. Since $\kappa_{\text{Env}}(m) < \kappa_{\text{Soc}}(m) < \kappa_{\text{Gov}}(m)$, it follows that $f(\xi_{\text{Env}}) > f(\xi_{\text{Soc}}) > f(\xi_{\text{Gov}})$, so the guidance effect is largest at the environmental dimension.

For fixed a, the salience effect is $F(\phi(a) + \kappa_j(C)) - F(\phi(a) + \kappa_j(T))$. By the mean value theorem, this equals $f(\zeta_j)\big[\kappa(C) - \kappa(T)\big]/\theta_j$ for some $\zeta_j \in (\phi(a) + \kappa_j(T),\, \phi(a) + \kappa_j(C))$. Unlike the fixed-$m$ comparison, tractability now affects both the location of the interval, through ζ_j, and its width, through the factor $1/\theta_j$. Lower-tractability dimensions generate a larger threshold wedge, but that wedge is evaluated farther into the right tail of the bandwidth distribution, where the density is lower. The stated sufficient condition that $f(\zeta_j)/\theta_j$ is strictly decreasing in $1/\theta_j$ ensures that the density decline dominates the wedge expansion. Using $1/\theta_{\text{Env}} < 1/\theta_{\text{Soc}} < 1/\theta_{\text{Gov}}$, the salience effect is therefore also largest at the environmental dimension. □

Remark 2. While the strict ordering in Proposition 5 follows mathematically from the assumptions, the magnitude of basic-engagement differences across dimensions depends on the threshold wedges $\kappa_j(m) - \kappa_{j'}(m)$, which may be quantitatively modest

relative to the spread of the bandwidth distribution. Tractability differences are therefore expected to be most consequential on the advanced engagement margin, where documentation effort must be translated into verifiable supporting materials and where the returns to producing those materials differ more sharply across dimensions.

2.5 Testable Hypotheses

The five propositions generate the following testable predictions, which map directly to the treatment contrasts estimated in Section 5.

Hypothesis 1 (Guidance Effect)**.** Suppliers assigned mandatory KPIs exhibit higher basic and advanced engagement, on average, than suppliers who self-select their reporting metrics.

Hypothesis 2 (Salience Effect)**.** Suppliers subject to targeted carbon monitoring exhibit higher basic and advanced engagement, on average, than suppliers subject to comprehensive audits.

Hypothesis 3 (Bandwidth Minimization Bundle)**.** The combination of mandatory KPIs and targeted monitoring produces the highest basic engagement among the four treatment cells. On the advanced engagement margin, it outperforms other governance structures in bandwidth-constrained supplier populations where the threshold-saving benefit dominates the quality advantage of delegation.

Hypothesis 4 (Capability Threshold)**.** The engagement effects of governance structure depend on suppliers' organizational capability. Among lower-capability suppliers, lower-overhead governance structures are more effective because governance-induced bandwidth costs are more likely to bind. As capability rises, the advantage of mandate over delegation is expected to narrow, with delegation becoming relatively more

effective on the advanced engagement margin. For monitoring scope, the advantage of targeted over comprehensive monitoring should likewise weaken as capability rises, with comprehensive monitoring becoming viable only for suppliers with sufficient organizational slack.

Hypothesis 5 (Task Tractability)**.** Treatment effects on engagement are strongest for environmental metrics, reflecting the lower per-unit bandwidth cost of environmental reporting relative to social and governance dimensions. Effects on less tractable dimensions are expected to be weaker, especially on the advanced engagement margin where documentation effort must be translated into verifiable supporting materials.

3 Institutional Context

3.1 The Lead Firm

The lead firm in this study is an industry-leading consumer goods and apparel brand based in China, with a large outsourced supplier network serving domestic and international markets. It holds top-tier ratings from major international ESG evaluation agencies, reflecting recognized progress on its own operations. However, as with most global consumer brands, the vast majority 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 untenable as investor, regulatory, and consumer pressure for supply chain transparency

intensified. The firm's strategic objective was accordingly to move beyond surface-level compliance toward genuine accountability. This required not only establishing ESG standards but designing governance structures capable of motivating substantive behavioral change among independent suppliers operating under significant resource constraints.

3.2 The Supplier Network

The firm's active supplier network comprises over 300 manufacturing partners across China, falling into four main categories: Cut-Make-Trim (CMT) manufacturers, who assemble garments from pre-cut fabric; Original Design Manufacturers (ODM), who handle product design, material sourcing, and manufacturing; fabric and material suppliers (FAB), producing raw textiles and functional synthetic materials; and ancillary material suppliers (ACC), providing zippers, labels, and other accessories.

For this experiment, the firm randomly selected 155 suppliers from its active supplier base. As Figure 1 illustrates, these suppliers are predominantly micro and small firms, typically operating on thin margins and handling labor-intensive garment assembly or design. Most of these firms operate as family-owned businesses with limited formal management hierarchies, and very few possess dedicated ESG departments or specialized compliance personnel. This profile is highly representative of manufacturing supply chains in emerging economies, where local suppliers play a critical role in global production but lack the administrative bandwidth of larger corporate entities (Gereffi et al., 2005).

3.3 The Digital Platform

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

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

Once suppliers click "Start Filing" on the welcome page, they are directed to the main reporting dashboard shown in Figure 3. The interface organizes all ESG requirements into distinct tabs corresponding to Environment, Social, Governance, Financial, and Certificate metrics. For every active indicator, suppliers can input data, add qualitative remarks, and upload supporting documentation. The interface also displays historical data from previous reporting periods to facilitate quarter-on-quarter comparison.

The platform tracks a comprehensive library of 108 supply chain indicators listed in Appendix A summarized in Table 1. Among these, 30 indicators covering basic financial information and standard factory certifications are mandatory for all suppliers,

constituting the common baseline requirement. The remaining ESG-specific indicators are either mandatory or optional depending on the supplier's assigned treatment group detailed in Section 4.

After the baseline data collection, the lead firm's supply chain management team conducted a structured telephone briefing with a designated manager at each supplier. The briefing outlined reporting requirements and performance evaluation standards without disclosing the randomized nature of the governance structure. Each supplier therefore understood its own obligations clearly while remaining unaware of the broader experimental design. This arrangement helped preserve both the integrity of the experiment and the realism of the governance environment.

3.4 Third-Party Monitoring and Verification

Third-party verification is widely used in global supply chain governance and has been shown to affect firms' compliance and reporting behavior (Duflo et al., 2013). During the experimental period, the lead firm implemented a formal verification process to strengthen data integrity and enforce accountability. Independent professional auditors were engaged to review the quality of submitted data for suppliers assigned to the relevant treatment conditions.

The scope of verification varied by experimental assignment. Under comprehensive monitoring, auditors verified all ESG indicators across the environmental, social, and governance domains through document reviews, facility inspections, and employee interviews. Under targeted monitoring, auditors verified only carbon emissions data, with all remaining ESG indicators left self-reported. This variation allows us to identify the effect of broader versus narrower oversight, holding fixed the presence of third-party verification. It therefore enables us to assess whether broader oversight

improves overall data quality or instead imposes administrative costs that reduce supplier engagement with the reporting process.

4 Experimental Design

4.1 The 2×2 Factorial Design

We implement a 2×2 factorial randomized controlled trial that independently varies two governance dimensions (Figure 4). The first dimension, vertical authority allocation, takes one of two values: (i) *Mandate*, in which the lead firm specifies a fixed set of eight ESG KPIs to be reported by the supplier; and (ii) *Delegation*, in which the supplier is permitted to self-select eight indicators from a broader menu of 25 environmental, 48 social, and 5 governance indicators. The second dimension, horizontal monitoring scope, also takes two values: (i) *Comprehensive Audit*, under which all submitted ESG metrics are subject to third-party verification; and (ii) *Targeted (Carbon) Audit*, under which only carbon-related metrics are externally verified.

Crossing these two dimensions yields four treatment cells. A separate control group, where suppliers report only basic financial and certification information with no additional ESG requirements, provides the comparison benchmark. The five experimental arms are thus:

(1) *Control*: baseline financial and certification reporting only.

(2) *Mandate + Comprehensive Audit*: baseline reporting + eight mandated KPIs with full third-party verification.

(3) *Mandate + Carbon Audit*: baseline reporting + eight mandated KPIs with carbon-only verification.

(4) *Delegation + Comprehensive Audit*: baseline reporting + eight self-selected KPIs with full third-party verification.

(5) *Delegation + Carbon Audit*: baseline reporting + eight self-selected KPIs with carbon-only verification.

Operationally, these five experimental arms correspond to five distinct platform interface designs (Figure 5). The control group is presented with a standard Baseline Interface. The four treatment groups are presented with interfaces matched to their specific treatment cell. The treatment variation is first introduced on the platform's welcome page, which informs each supplier of their reporting obligations and assigned verification scope. This logic carries through to the main filing interface, where the platform dynamically adjusts the required data entry fields based on whether the supplier's KPIs were mandated or delegated. Screenshots and English translations of the group-specific welcome screens are provided in Appendix B.

4.2 Randomization and Sample Balance

We randomly assigned the 155 participating suppliers across the five experimental conditions using a stratified procedure based on supplier type (CMT, ODM, FAB, ACC) and firm size (proxied by historical production value). Stratification improves precision and ensures that treatment effects are not confounded by systematic pre-existing differences in supplier characteristics. The control group received 46 suppliers; the four treatment cells received approximately 27 suppliers each ($N_{treat} = 109$).

Table 2 reports pre-treatment balance checks. Treatment groups are statistically indistinguishable from the control group on all available baseline characteristics: production volume, total employment, firm age, registered capital, number of ESG

certifications, and supplier type composition. No difference is statistically significant at conventional levels, validating the randomization procedure.

4.3 Timeline and Implementation

The experiment proceeded in three phases. In Phase 1 (July–August 2025), all suppliers logged into the platform under their randomly assigned governance conditions to submit retrospective ESG data for fiscal year 2024. Phase 2 (September 2025–March 2026) constituted the main intervention period. In September 2025, the lead firm's supply chain management team conducted structured telephone briefings with designated managers at each supplier, reinforcing the reporting requirements and audit scope specific to their assigned conditions without disclosing the broader experimental design. Quarterly data collection proceeded through three rounds: Round 1 in October 2025, Round 2 in January 2026, and Round 3 in April 2026. Independent third-party audits for treatment groups were conducted in November 2025. Phase 3 (April–June 2026) consolidates the full longitudinal dataset and administers a post-intervention survey to elicit qualitative evidence on perceived cognitive load and coordination costs.

4.4 Outcome Measures and Estimation Sample

We quantify supplier engagement along two complementary margins. The first outcome, *Basic Engagement*, is the missing-data rate: the fraction of required data fields that a supplier leaves empty. A lower missing-data rate indicates higher basic engagement. The second outcome, *Advanced Engagement*, is documentation coverage: the number of distinct ESG categories for which a supplier uploads at least one supporting document. We further decompose this coverage into hard information (e.g., quantitative utility records) and soft information (e.g., qualitative policy documents), with detailed

variable construction outlined in Appendix C. We evaluate these measures both in aggregate and across specific ESG sub-categories to test whether treatment effects vary by coordination complexity.

The estimation sample comprises 591 supplier-quarter observations, drawn from 150 suppliers across four reporting rounds. Of the 155 suppliers initially enrolled, five did not engage with the platform following assignment and are excluded. Among the remaining 150 suppliers, seven did not complete the baseline filing and two suspended reporting in subsequent rounds due to an absence of active orders; these nine observations are omitted from the corresponding rounds. Results are robust to including these suppliers with imputed outcomes or to excluding them altogether.

4.5 Empirical Specification

Given that experimental interfaces are deployed from the first reporting round, all observed platform behavior occurs under the assigned governance conditions. Our main estimating equation is therefore an OLS regression with stratum and round fixed effects:

$$Y_{ist} = \alpha + \beta_1 \cdot \text{Mandate}_i + \beta_2 \cdot \text{Delegation}_i + \delta_s + \lambda_t + \varepsilon_{ist}, \tag{9a}$$

or

$$Y_{ist} = \alpha + \beta_1 \cdot \text{Comprehensive}_i + \beta_2 \cdot \text{Carbon}_i + \delta_s + \lambda_t + \varepsilon_{ist}, \tag{9b}$$

where Y_{ist} is the engagement outcome for supplier i in stratum s during reporting round t. The indicators Mandate_i and Delegation_i identify suppliers assigned to the mandate and delegation arms, respectively, pooled across monitoring scope conditions; analogously, Carbon_i and Comprehensive_i identify the monitoring arms. The control

group (baseline reporting with no additional KPI requirements or monitoring) serves as the omitted category. Stratum fixed effects δ_s absorb cross-sectional variation in supplier type and size; round fixed effects λ_t control for aggregate time trends. Standard errors are clustered at the stratum level.

To test Hypothesis 3 on interaction effects, we augment equation (9) with full treatment-cell indicators:

$$Y_{ist} \;=\; \alpha + \sum_{k \in \{1,2,3,4\}} \gamma_k \cdot \mathbf{1}[\text{cell}_i = k] + \delta_s + \lambda_t + \varepsilon_{ist}, \tag{10}$$

where the four cells are indexed as: $k = 1$ (*Mandate + Comprehensive*), $k = 2$ (*Mandate + Carbon*), $k = 3$ (*Delegation + Comprehensive*), and $k = 4$ (*Delegation + Carbon*). A coefficient γ_2 that is significantly larger than γ_k for $k \in \{1, 3, 4\}$ would confirm the bandwidth minimization prediction of Proposition 3.

To test Hypothesis 4 on capability heterogeneity, we augment the main specification with interactions between treatment indicators and a binary measure of baseline supplier characteristics. The moderating variables include the lead firm's supplier quality ranking (STI, A, or B tiers), production volume, and employment:

$$\begin{aligned} Y_{ist} = \alpha &+ \mu_1 \cdot \text{Mandate}_i + \mu_2 \cdot \text{Delegation}_i \\ &+ \mu_3 \cdot (\text{Mandate}_i \times \text{High}_i) + \mu_4 \cdot (\text{Delegation}_i \times \text{High}_i) \\ &+ \delta_s + \lambda_t + \varepsilon_{ist}, \end{aligned} \tag{11a}$$

or

$$\begin{aligned} Y_{ist} = \alpha &+ \mu_1 \cdot \text{Comprehensive}_i + \mu_2 \cdot \text{Carbon}_i \\ &+ \mu_3 \cdot (\text{Comprehensive}_i \times \text{High}_i) + \mu_4 \cdot (\text{Carbon}_i \times \text{High}_i) \\ &+ \delta_s + \lambda_t + \varepsilon_{ist}, \end{aligned} \tag{11b}$$

where High_i is an indicator equal to one if the supplier falls in the upper tail of the

moderating characteristic, and zero otherwise. The coefficients μ_1 and μ_2 identify the treatment effects for suppliers in the lower tail of the moderating characteristic ($\text{High}_i = 0$), while μ_3 and μ_4 capture the incremental treatment effects for higher-capability suppliers relative to this baseline. The total treatment effects for the latter group are therefore $\mu_1 + \mu_3$ and $\mu_2 + \mu_4$, respectively.

5 Results

5.1 Main Effects: Authority Allocation

We begin by testing the prediction that mandating a fixed KPI set outperforms delegation of KPI choice (Hypothesis 1). The theoretical logic is straightforward: when organizational bandwidth is scarce, the decision cost ϕ of evaluating and selecting among indicators consumes resources that would otherwise be devoted to reporting itself. Mandates eliminate this overhead entirely, lowering the effective initiation threshold from $\phi + \kappa(m)$ to $\kappa(m)$ and thereby converting potential non-participants into active reporters.

Figure 6 presents the main results, with full regression estimates reported in Table D1. On the basic engagement margin, the mandate treatment significantly reduces the missing-data rate by 10.4 percentage points, a 23.0 percent improvement relative to the control group mean of 45.2 percent. Delegation, by contrast, produces a reduction of only 4.3 percentage points that is not distinguishable from zero. A parallel pattern holds when the outcome is measured in raw indicator counts (Column 1 of Table D1).

The pattern is even more pronounced on the advanced engagement margin. Mandated suppliers upload documentation spanning approximately one additional ESG

category beyond the control average of roughly two categories, an increase of nearly 50 percent. This gain is concentrated in hard-information documentation such as utility invoices and quantitative records, where the mandate coefficient is statistically significant, rather than in soft-information materials such as qualitative policy documents, where the point estimate is positive but imprecise. The delegation treatment, by contrast, produces negligible effects on documentation coverage. These findings indicate that the bandwidth savings from eliminating the selection decision translate directly into deeper compliance effort, not merely broader field completion.

The sharp contrast between the mandate and delegation arms is consistent with the extensive-margin mechanism formalized in Proposition 1: for the majority of capacity-constrained suppliers in this network, the cognitive cost of navigating a menu of ESG indicators exceeds the informational benefit of local tailoring. This interpretation aligns with the choice-overload literature in behavioral economics (Iyengar and Lepper, 2000) and the growing evidence that decision simplification improves take-up of complex programs among resource-constrained agents (Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019).

5.2 Main Effects: Monitoring Scope

Hypothesis 2 predicts that targeted carbon monitoring generates higher engagement than comprehensive auditing. The underlying mechanism operates through the initiation threshold $\kappa(m)$: when suppliers anticipate that all ESG dimensions will be verified, the coordination cost of preparing across functionally disconnected departments raises the entry barrier above what many constrained firms can clear. A single focal audit lowers this threshold to $\kappa(T) < \kappa(C)$, providing a clear entry point that channels limited attention efficiently.

Figure 7 displays the treatment effects, with detailed estimates in Table D2. Targeted carbon monitoring reduces the missing-data rate by 9.9 percentage points, or approximately 22 percent relative to the control mean. Comprehensive monitoring yields a smaller reduction of 4.7 percentage points that is statistically insignificant. For documentation coverage, the carbon audit generates an increase of 0.857 additional categories over a control baseline of approximately two categories, four times the point estimate for comprehensive monitoring. The carbon audit effect is driven primarily by hard-information uploads, consistent with the idea that a focused verification signal directs compliance effort toward concrete, quantifiable evidence rather than diffuse documentation.

These results confirm the salience effect derived in Proposition 2. A narrowly targeted audit does not merely verify less; it restructures how suppliers allocate scarce administrative bandwidth by signaling which dimension matters most. This finding resonates with the attention-based view of the firm (Ocasio, 1997) and complements the audit-design evidence of Duflo et al. (2013, 2018) by demonstrating that verification scope shapes not only the accuracy of reported data but also the basic willingness of regulated firms to engage with the reporting process in the first place.

5.3 Interaction Effects: The Bandwidth-Minimization Bundle

The model predicts that simplifying the vertical decision structure and narrowing the horizontal task scope are mutually reinforcing, because both interventions independently reduce the total bandwidth cost of compliance (Hypothesis 3). The bundle (M, T) uniquely minimizes the initiation threshold among all four governance structures, and Proposition 3 establishes that it therefore uniquely maximizes expected engagement under the conditions characterizing SME-dominated supply networks.

Figure 8 plots the engagement outcomes for all four treatment cells alongside the control group, with the underlying estimates reported in Table D3. The results are striking. The only cell that generates statistically significant improvements on both engagement margins is the combination of mandatory KPIs and targeted carbon monitoring. This bundle reduces the missing-data rate by 11.5 percentage points, representing a 25.4 percent improvement over the control mean of 45.2 percent. On the advanced engagement margin, it increases documentation coverage by 1.338 categories, a 67.4 percent gain relative to the control average of roughly two categories. The hard-information component of this gain is particularly large, nearly doubling the control-group level of approximately one hard-information category.

The remaining three cells tell an equally informative story. Mandates paired with comprehensive monitoring produce a sizable but imprecisely estimated reduction in the missing rate and a moderate increase in documentation. Delegation paired with comprehensive monitoring generates essentially zero effect on either margin. Delegation paired with carbon monitoring falls between these extremes. The monotonic ordering across cells tracks the theoretical ordering of initiation thresholds: the cell with the lowest overhead produces the strongest effects, and engagement declines as either vertical complexity (delegation) or horizontal breadth (comprehensive monitoring) is reintroduced.

These patterns confirm the complementarity prediction of Hypothesis 3. Neither simplification alone is sufficient to maximize engagement; the compounding benefit arises because each independently reduces the bandwidth tax on compliance, and their combination pushes the effective cost below the engagement threshold for a larger fraction of the supplier distribution. For practitioners, this result implies that authority allocation and monitoring scope should be designed jointly rather than optimized in isolation.

5.4 Heterogeneity by Supplier Capability

A central prediction of the bandwidth-constraint model is that the advantage of simplified governance is not uniform across the supplier distribution (Hypothesis 4). For suppliers with ample organizational slack, the decision cost ϕ and the coordination threshold $\kappa(m)$ are small relative to available bandwidth, and the classical predictions should begin to reassert themselves: delegation recovers its information advantage, and comprehensive monitoring recovers its effort-allocation benefit. The governance reversal documented above should therefore be concentrated among lower-capability firms and should attenuate or reverse among the most capable ones.

Authority allocation. Figure 9 displays the heterogeneous effects of mandate and delegation across capability levels, with the full set of interaction estimates reported in Table D4. For basic engagement, the mandate generates strong and broad-based reductions in the missing-data rate, with negligible capability interactions across all three moderators (quality rank, production, and employment). Structured guidance thus functions as a universal compliance floor: it lowers the barrier to entry regardless of where a supplier sits on the capability distribution.

Advanced engagement reveals a strikingly different pattern. Among lower-capability suppliers, the mandate produces positive but moderate documentation gains, while delegation generates negative or near-zero effects. As capability rises, however, the delegation treatment exhibits substantially stronger marginal returns than the mandate treatment. In Table D4, the interaction between delegation and high quality rank on documentation categories is 1.894, implying that the total delegation effect for top-ranked suppliers is positive and economically meaningful. A parallel pattern emerges for high-production and high-employment firms: the delegation interaction

terms on attachment categories are 1.491 and 1.429, respectively, each large enough to push the total delegation effect into positive territory for the most capable subset. The mandate interactions, by contrast, are uniformly smaller, indicating that the mandate's documentation advantage narrows as organizational slack increases.

This asymmetry maps directly onto the trade-off formalized in Proposition 4. For constrained suppliers, the bandwidth savings from mandate dominate the selection gain from delegation, producing a net advantage for simplified governance. For capable suppliers, the decision cost ϕ is small relative to available bandwidth, and the quality multiplier $\bar{\theta} > \tilde{\theta}$ begins to dominate, restoring the classical prediction that local information improves reporting quality. Autonomy is therefore not inherently ineffective; its value is contingent on whether the agent possesses the organizational slack to productively leverage local knowledge, consistent with the management-quality evidence of Bloom et al. (2013).

Monitoring scope. Figure 10 presents the analogous heterogeneity analysis for monitoring scope, with supporting estimates in Table D5. For basic engagement, carbon monitoring consistently attracts supplier attention more effectively than comprehensive monitoring, indicating that the salience effect is broadly applicable and not restricted to the least capable firms.

On the advanced engagement margin, a pronounced threshold effect emerges. Among lower-capability suppliers, both monitoring structures generate small or even negative documentation effects. As capability increases, the carbon audit interaction terms are large and positive across all three moderators: the interaction with high quality rank is 1.753 on documentation categories, and the interactions with high production and high employment are 2.155 and 1.719, respectively. Comprehensive monitoring also shows positive capability interactions, but these are smaller in magni-

tude and are economically significant only for the highest-quality-ranked suppliers. The implication is that targeted monitoring works across a broad range of capability, while comprehensive monitoring becomes viable only at the very top of the capability distribution. This pattern suggests a natural lifecycle approach to governance design: targeted, directive structures are most appropriate for constrained firms in the early stages of an ESG program, while comprehensive oversight becomes feasible as supplier capability develops (Verhoogen, 2023; Locke, 2013).

5.5 Heterogeneity Across ESG Dimensions

The model's final prediction concerns the dimension-specific pattern of engagement. Because environmental indicators have higher tractability ($\theta_{\text{Env}} = 1 > \theta_{\text{Soc}} > \theta_{\text{Gov}}$), their effective threshold is lowest, and bandwidth-constrained suppliers rationally concentrate compliance effort on the environmental dimension first (Proposition 5, Hypothesis 5).

Figure 11 decomposes the treatment effects by reporting domain, with the full estimates in Table D6. The evidence strongly supports this prediction. For authority allocation, the mandate reduces the environmental missing-data rate by 14.5 percentage points, a substantial improvement relative to the domain-specific control mean of 74.1 percent. The corresponding effects on social and governance dimensions are smaller and statistically imprecise. On the documentation margin, the mandate increases the probability of uploading at least one environmental attachment by 33.0 percentage points against a control mean of 31.8 percent, effectively doubling the environmental documentation rate. Neither social nor governance attachment rates respond significantly to the mandate.

The monitoring scope results tell a parallel story. Targeted carbon monitoring

reduces the environmental missing-data rate by 15.5 percentage points and raises the environmental attachment rate by 27.9 percentage points, both statistically significant. The corresponding effects on social and governance dimensions are again negligible.

This concentration of treatment effects in the environmental domain has a natural interpretation within the bandwidth-constraint framework. Factory managers can obtain energy consumption data from utility invoices, chemical inventories from procurement records, and renewable energy shares from supplier contracts. These data typically reside within a single organizational function and require minimal cross-departmental coordination. Social and governance metrics, by contrast, demand coordination across HR, legal, and compliance departments and often require qualitative documentation that must be drafted rather than retrieved. Under strict bandwidth constraints, suppliers rationally prioritize the most cognitively tractable dimensions, producing the domain-specific pattern we observe. This finding is consistent with the multi-dimensional effort allocation mechanism of Holmström and Milgrom (1991) and suggests that environmental reporting is the natural entry point for capacity-constrained suppliers entering the ESG compliance landscape.

5.6 Robustness

Tables D7 and D8 report a comprehensive set of robustness checks for both the authority allocation and monitoring scope results. The main findings are stable across four alternative specifications: wild bootstrap clustered standard errors, standard errors clustered at the firm level rather than the stratum level, a collapsed cross-sectional sample averaging outcomes across rounds, and the baseline specification. In every case, the mandate and carbon monitoring coefficients remain significant and comparable in magnitude. The delegation and comprehensive monitoring coefficients

consistently remain small and statistically insignificant. These patterns rule out that the main results are artifacts of the specific clustering assumption or are driven by within-firm serial correlation across reporting rounds.

6 Discussion

6.1 Implications for Organizational Theory

Our results generate three insights for organizational economics. First, they demonstrate that the scope conditions of delegation theory are empirically binding in inter-firm settings. Classical models predict that granting agents the authority to tailor tasks exploits private information about local costs and improves outcomes (Aghion and Tirole, 1997; Dessein, 2002). These benefits are real: they materialize among the most capable suppliers in our sample. But for the median SME factory manager in a developing-economy supply chain, the cognitive cost of evaluating a menu of unfamiliar indicators exceeds the informational gain from local tailoring. The model in Section 2 captures this through the decision cost $\phi(a)$: delegation improves welfare only when ϕ is small relative to the agent's bandwidth B. Our finding that the delegation advantage emerges only for top-tier suppliers maps directly onto this condition and contributes to a growing literature on how heterogeneous agent characteristics determine the returns to managerial practices (Bloom and Van Reenen, 2007; Gibbons, 2005; Bandiera et al., 2020). More broadly, the result highlights that organizational design recommendations drawn from within-firm studies of relatively well-resourced agents (Fehr et al., 2013; Hong et al., 2018) may not transfer to the capacity-constrained inter-firm settings that characterize most global supply chains.

Second, our results extend the multitasking framework of Holmström and Milgrom

(1991) to inter-firm governance with heterogeneous task complexity. Standard theory focuses on how incentive distortions arise when tasks differ in measurability. We document an additional channel: task breadth interacts with bandwidth constraints to determine whether comprehensive oversight motivates or overwhelms the agent. Targeted monitoring works across the capability distribution, whereas comprehensive monitoring generates positive advanced engagement only among the highest-ranked suppliers. This finding complements the audit-design evidence of Duflo et al. (2013) and the inspection-targeting results of Duflo et al. (2018), who show that the structure and scope of third-party verification shapes both reporting accuracy and behavioral responses. Our contribution highlights an administrative-burden channel that operates alongside the strategic-misreporting motives emphasized in the literature: even absent any incentive to misreport, audit breadth alone can depress engagement by raising the initiation threshold beyond what constrained agents can clear.

Third, our interaction results shed light on how governance instruments combine. The model predicts that the (M, T) bundle uniquely minimizes the initiation threshold because the two simplifications reduce bandwidth costs along independent dimensions (Proposition 3). Consistent with this prediction, (M, T) is the only treatment cell that produces statistically significant improvements on both engagement margins, and its point estimates exceed those of every other cell. The underlying mechanism is additive threshold reduction: mandate eliminates the decision cost ϕ, and targeted monitoring lowers the initiation threshold from $\kappa(C)$ to $\kappa(T)$, so the two savings apply simultaneously and bring the effective cost of compliance below the engagement threshold for a larger fraction of the supplier distribution.

6.2 Implications for Supply Chain Governance Practice

For practitioners, the central message is that governance complexity is itself a cost, and one that is particularly consequential for capacity-constrained suppliers. The temptation to maximize information collection by offering broad autonomy and comprehensive oversight is natural, but our evidence indicates that it is counterproductive at the initial stages of a compliance program. Several design principles follow from the findings.

First, governance should be matched to supplier capability. Lead firms should segment their supplier base and adopt directive, narrow governance for SME suppliers at the outset of a compliance program. As suppliers develop internal capacity and demonstrated familiarity with reporting, progressive delegation and expanded monitoring scope become appropriate. This lifecycle approach avoids the binary choice between full standardization and full customization that characterizes much current practice (Locke, 2013).

Second, clear communication of expectations is essential. On the vertical dimension, a clear directive about which indicators to report eliminates the cognitive overhead of menu evaluation and ensures that constrained managers can proceed directly to execution. On the horizontal dimension, a focused verification scope signals which compliance domain matters most, channeling limited attention toward a tractable entry point rather than dispersing it across all reporting categories. In both cases, the mechanism is the same: clarity reduces the bandwidth cost of initiating compliance, consistent with the attention-allocation literature (Ocasio, 1997; Gabaix, 2014).

Third, governance tools should be designed jointly. Choosing authority allocation and monitoring scope independently risks forgoing the compounding engagement gains that arise from simultaneous simplification. In our setting, the bundle that pairs

mandated KPIs with targeted monitoring dominates all other configurations precisely because both sources of bandwidth cost are reduced at once.

Fourth, lead firms should account for the differential tractability of various compliance dimensions. Metrics with standardized measurement protocols are the natural entry point for capacity-constrained suppliers, given their clearer evaluation criteria and lower coordination requirements. Programs that immediately impose broad and qualitative requirements risk generating superficial compliance. A phased approach that begins with quantifiable reporting and progressively expands to other dimensions as supplier capacity develops is more likely to generate genuine accountability (Christensen et al., 2021).

Finally, these findings carry implications for the design of supply chain due diligence regulation. Policymakers who impose broad and uniform disclosure obligations on lead firms, and by extension on their supplier networks, risk triggering the same attention-dilution effects documented here (Boudreau, 2024; Chen et al., 2018). Regulatory frameworks that allow lead firms flexibility in sequencing and prioritizing supplier requirements are more likely to produce effective behavioral change in SME-dominated production networks.

6.3 Limitations and Extensions

Several limitations merit acknowledgment. First, the experiment was conducted in a single industry (apparel) and country (China). The specific magnitudes of the treatment effects may differ in other supply chain contexts, particularly in industries with higher baseline environmental or social risk where suppliers may already possess more developed compliance infrastructure. Whether the directional findings of the effects generalizes more broadly is an empirical question for future work.

Second, our outcomes measure engagement with the reporting platform rather than underlying economic performance. The link between reporting engagement and genuine operational improvement, such as actual emissions reductions or labor standard gains, is an important step that requires a longer time horizon. Preliminary evidence from the platform data suggests that delegation and carbon monitoring generate positive early-stage effects on certificate acquisition and energy metrics, but the sample size and follow-up period are not yet sufficient for definitive conclusions.

Third, the current analysis does not directly measure the bandwidth constraint at the individual manager level. Post-experiment survey data on perceived cognitive load and coordination costs will allow more direct validation of the mechanism. Similarly, testing whether the engagement effects documented here generate positive or negative spillovers on production performance, a central concern of the multitasking framework (Holmström and Milgrom, 1991), requires more granular production data than are currently available.

7 Conclusion

Managing supplier compliance with multi-dimensional reporting requirements is among the most consequential organizational challenges facing multinational firms under intensifying regulatory and investor pressure. Yet the practical question of how to design governance structures that alter behavior in networks dominated by small, under-resourced manufacturers has received little rigorous causal scrutiny. This paper provides experimental evidence that governance design choices matter enormously, and that for bandwidth-constrained agents, the most effective structures are often far simpler than standard organizational economics would predict.

For a network of capacity-constrained SME suppliers facing novel multi-dimensional

ESG requirements, the most effective governance action is to make expectations explicit, focus verification on the dimensions that matter most, and resist the impulse to maximize reporting comprehensiveness at the expense of operational tractability. When governance is simple and directive, suppliers can devote their limited bandwidth to execution. When it is complex and open-ended, the cognitive demands of navigating the governance structure itself become a barrier that many suppliers cannot clear.

These findings rest on an important qualification of classical principal-agent models. The predictions of delegation theory and the multitasking framework describe optimal governance under the assumption that agents possess sufficient bandwidth to process the demands placed upon them. When this assumption is relaxed to accommodate the organizational reality of multi-tasked factory managers with limited administrative capacity, the predictions reverse in systematic and interpretable ways. Crucially, the reversal is not a refutation of the classical frameworks but an extension that applies in the region of high decision cost, high coordination load, and low organizational slack that characterizes SME-dominated supply networks. As supplier capability rises, the classical predictions reassert themselves, and the advantages of delegation and comprehensive oversight re-emerge.

The lessons generalize beyond the specific production network examined here. The core dynamics, capacity-constrained SMEs grappling with multi-dimensional compliance requirements and lead firms attempting to fill institutional voids left by limited public regulatory capacity, are defining features of modern supply chains across industries and geographies (Gereffi et al., 2005; Boudreau et al., 2023). As global regulatory requirements expand, governance structures that account for agent constraints will be critical in transforming external compliance pressure into genuine and sustained supply chain 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.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham, "Management and Shocks to Worker Productivity," *Journal of Political Economy*, 2022, *130* (1), 1–47.

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

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P. Vasquez, "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages," *Quarterly Journal of Economics*, 2022, *137* (3), 1495–1552.

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

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

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

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

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.

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.

_ , Aprajit Mahajan, David McKenzie, and John Roberts, "Do Management Interventions Last? Evidence from India," *American Economic Journal: Applied Economics*, 2020, *12* (2), 198–219.

_ , 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.

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.

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.

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.

Dimson, Elroy, Oguzhan Karakas, and Xi Li, "Active Ownership," *Review of Financial Studies*, 2015, *28* (12), 3225–3268.

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.

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

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

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

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

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

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

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

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

Gosnell, Greer K., John A. List, and Robert D. Metcalfe, "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains," *Journal of Political Economy*, 2020, *128* (4), 1195–1233.

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.

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.

Iyengar, Sheena S. and Mark R. Lepper, "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?," *Journal of Personality and Social Psychology*, 2000, *79* (6), 995–1006.

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

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.

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

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

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

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

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

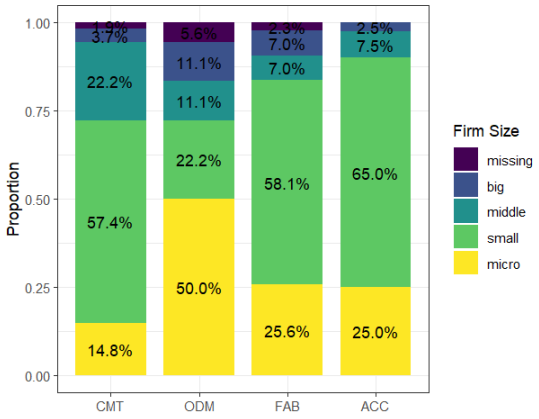
Mookherjee, Dilip, "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 2006, *44* (2), 367–390.

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

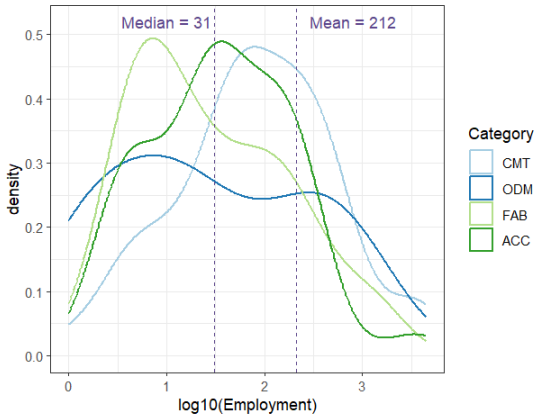
Rasul, Imran and Daniel Rogger, "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal*, 2018, *128* (608), 413–446.

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

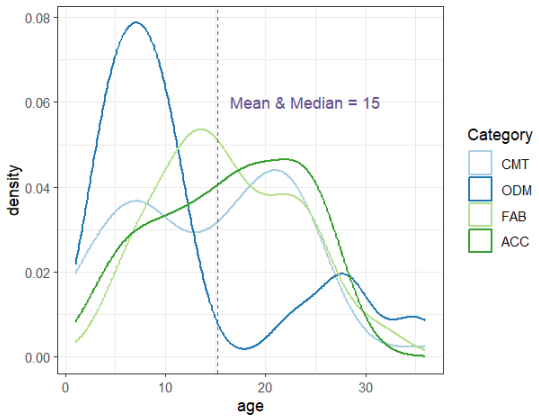
Figures and Tables



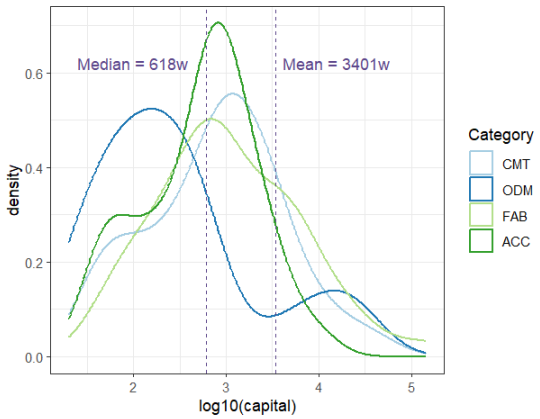
Panel A: Supplier Size



Panel B: Supplier Employment



Panel C: Supplier Age



Panel D: Supplier Registered Capital

Figure 1: Characteristics of the Supplier Network

Notes: This figure presents the distribution of key characteristics for the 155 suppliers in the experimental sample. Panel A displays firm size by supplier type: Cut-Make-Trim (CMT), Original Design Manufacturer (ODM), fabric/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 '0000. The distributions indicate a supplier network concentrated in small and micro-sized enterprises, consistent with the capacity constraints that motivate the empirical design.



Figure 2: Digital Platform Interface: Welcome Page

Notes: This figure displays the platform welcome page, with English translation, presented to suppliers upon login at the start of each reporting cycle. The text introduces the ESG reporting requirements and the scope of third-party auditing assigned to the supplier. While the page layout is identical across users, the substantive content varies with the supplier's randomly assigned treatment condition, making this screen the primary channel through which the experimental variation is delivered.

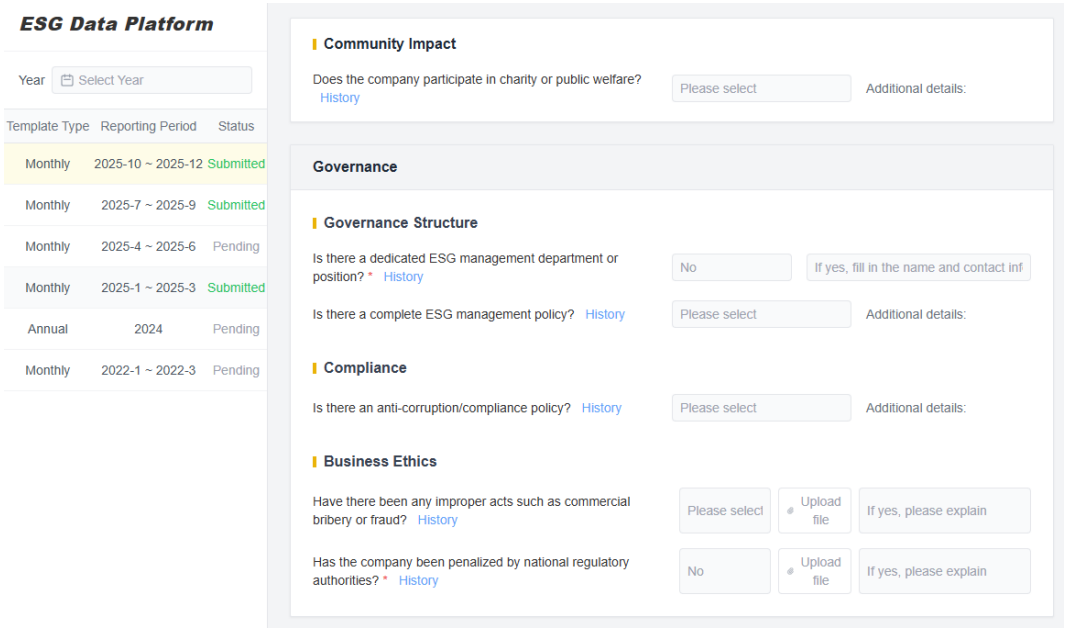


Figure 3: Digital Platform Interface: Main Reporting Dashboard

Notes: This figure shows the main reporting dashboard, with English translation, where suppliers enter ESG data. The interface organizes reporting requirements into Environmental, Social, Governance, Financial, and Certification tabs. For each active indicator, suppliers may enter data, append qualitative remarks, and upload supporting documentation. Historical submissions are displayed alongside current-period fields, and the set of mandatory items changes with the supplier's assigned treatment condition.

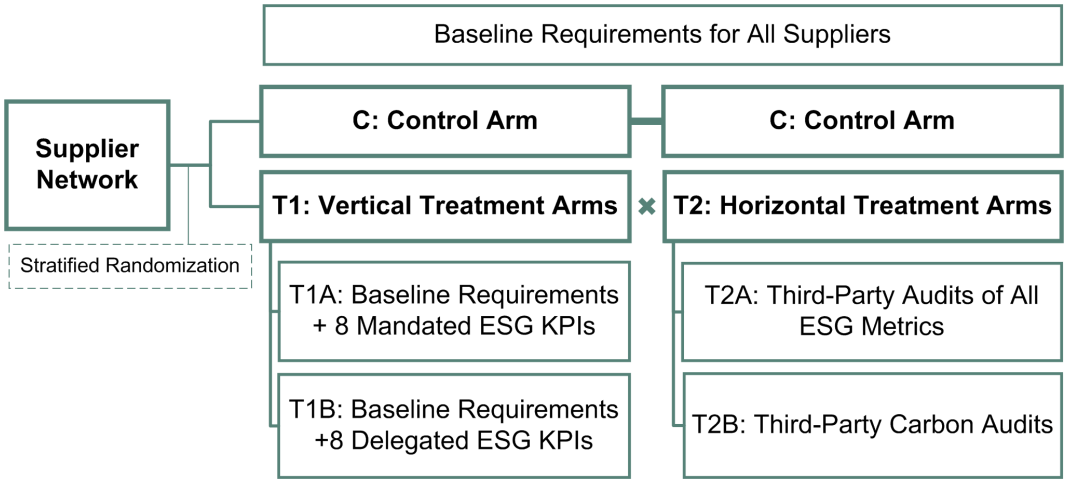


Figure 4: The 2×2 Experimental Design

Notes: This figure illustrates the structure of the experimental assignment. All suppliers face common baseline reporting requirements covering financial and certification indicators. Treatment suppliers are cross-randomized along two dimensions: vertical authority allocation over KPI selection (Mandated KPIs versus Delegated KPIs) and horizontal monitoring scope (Comprehensive Audit versus Carbon Audit). Crossing these two dimensions yields four treatment cells, alongside a pure control group assigned to baseline reporting only.

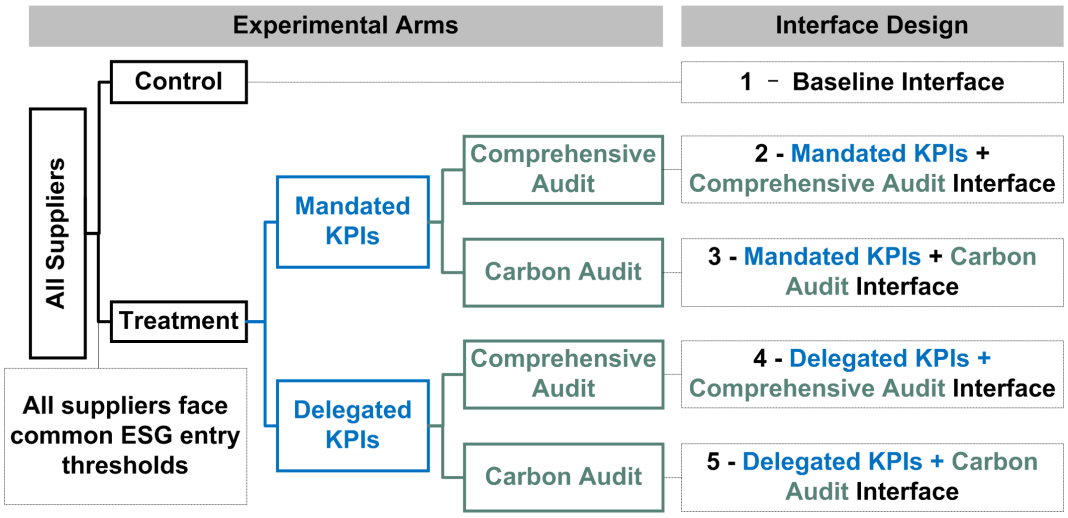


Figure 5: Interface Design by Experimental Arm

Notes: This figure maps the five experimental conditions to their corresponding platform implementations. Interface 1 is assigned to the control group. Interfaces 2 through 5 correspond to the four treatment cells formed by crossing authority allocation and monitoring scope. The assigned interface determines both the welcome-page instructions and the data-entry fields marked as mandatory in the subsequent reporting dashboard.

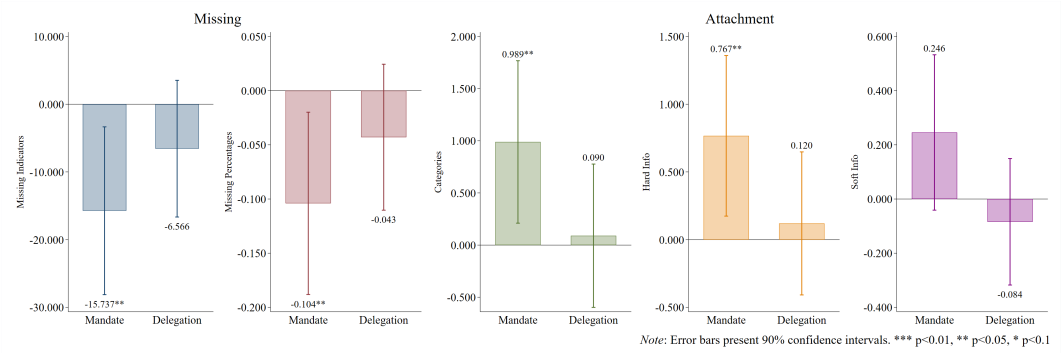


Figure 6: The Guidance Effect on Supplier Engagement

Notes: This figure illustrates the treatment effects of vertical authority allocation (Mandate versus Delegation) on basic engagement, measured by the missing data rate, and advanced engagement, measured by documentation coverage. Clear mandates consistently generate stronger participation than delegated autonomy, indicating that explicit guidance lowers the cognitive barrier to entry for resource-constrained suppliers. Bars show coefficient estimates with 90% confidence intervals. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. The corresponding regression table can be found in Table D1.

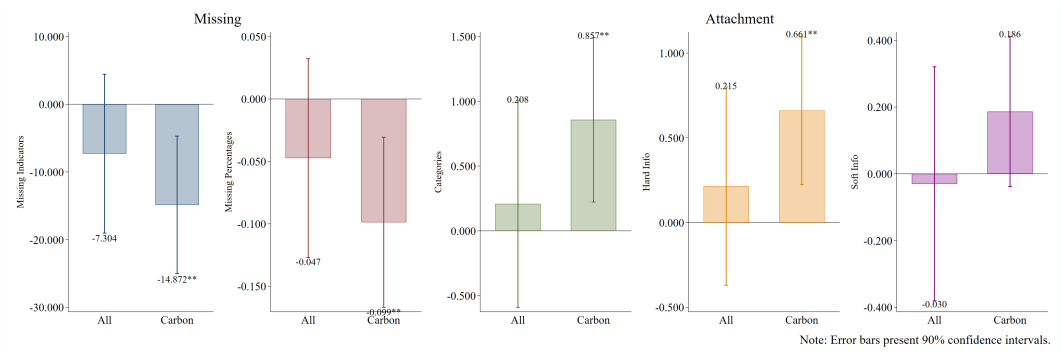


Figure 7: The Salience Effect of Targeted Monitoring

Notes: This figure compares the impact of comprehensive ESG auditing and targeted carbon auditing on supplier engagement. Targeted monitoring outperforms comprehensive monitoring on both information completeness and documentation uploads, indicating that a focused compliance signal helps suppliers prioritize limited administrative resources. Bars show coefficient estimates with 90% confidence intervals. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. The corresponding regression table can be found in Table D2.

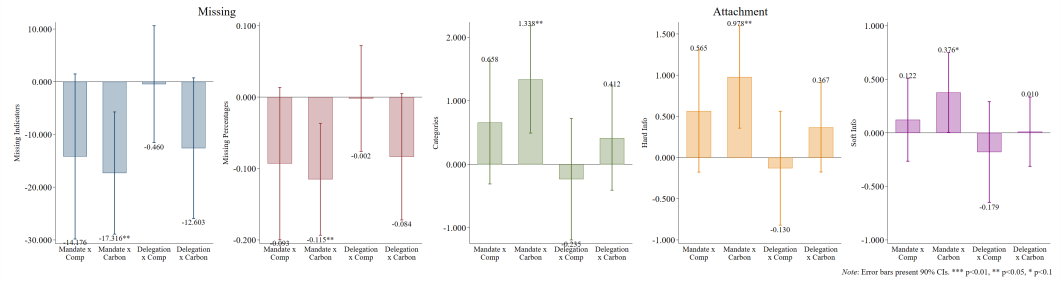


Figure 8: Interaction Effects of Simplified Governance

Notes: This figure plots engagement outcomes across all four interacted treatment cells and the control group. The combination of Mandated KPIs and Targeted Carbon Monitoring produces the strongest improvement in both basic and advanced engagement, confirming that removing decision complexity and narrowing task breadth jointly reduces the organizational friction required to initiate reporting. Bars show coefficient estimates with 90% confidence intervals. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. The corresponding regression table can be found in Table D3.

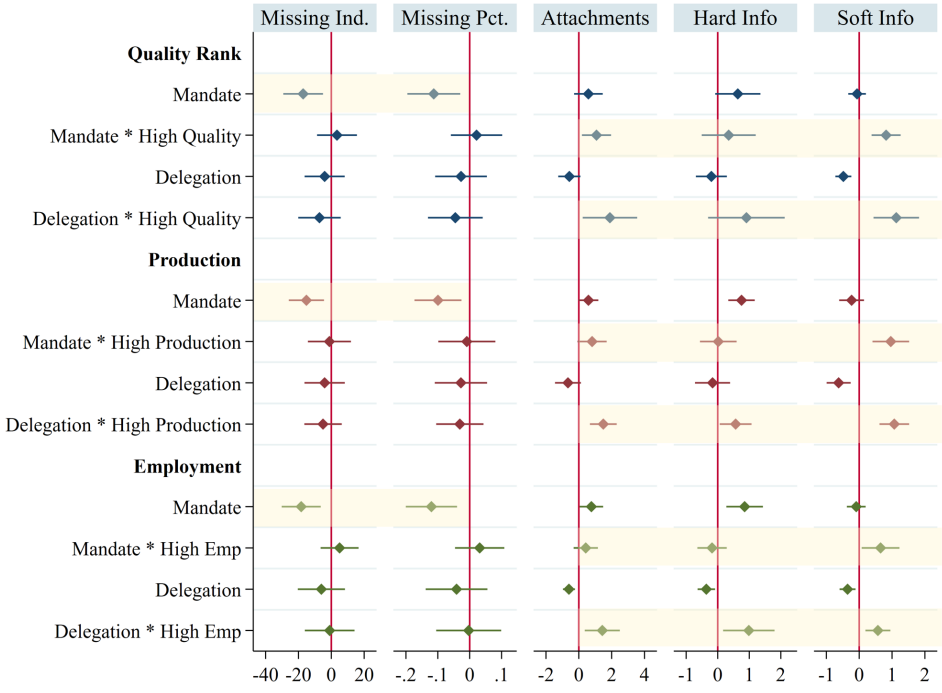


Figure 9: Capability Threshold Heterogeneity in Authority Allocation

Notes: This figure displays the estimated treatment effects of vertical authority allocation, Mandate versus Delegation, by supplier capability group. Mandates improve compliance broadly on the basic engagement margin. On the advanced engagement margin, delegation yields strong positive returns only among the higher-capability suppliers, indicating that autonomy requires organizational slack to be effective. The main treatment coefficients correspond to suppliers in the lower-capability group; the total treatment effects for higher-capability suppliers are the sum of the main coefficient and the interaction term. Points show coefficient estimates with error bars indicating 90% confidence intervals. The corresponding regression table can be found in Table D4.

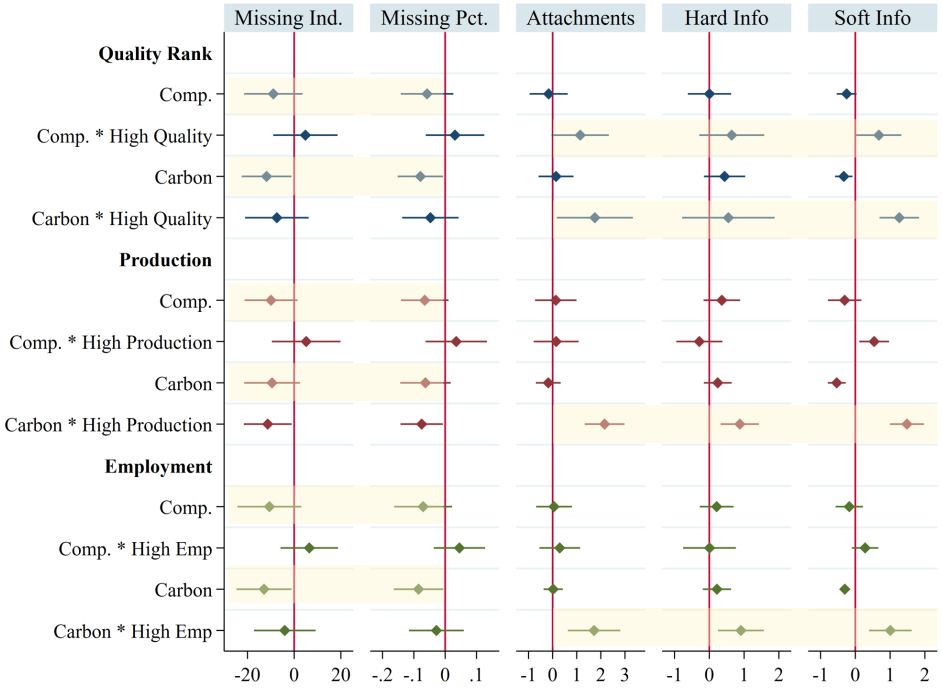


Figure 10: Capability Threshold Heterogeneity in Monitoring Scope

Notes: This figure displays the estimated treatment effects of horizontal monitoring scope, Comprehensive Audit versus Carbon Audit, by supplier capability group. Carbon audits generate consistent basic engagement gains across capability groups; comprehensive audits show weaker and less precise effects on this margin. On the advanced engagement margin, carbon monitoring exhibits a robust positive interaction with supplier capability across all three moderating characteristics, whereas the incremental gains from comprehensive monitoring are detectable only among the highest-ranked suppliers. This pattern is consistent with broader audit scope imposing bandwidth demands that only exceptionally capable suppliers can satisfy. The main treatment coefficients correspond to suppliers in the lower-capability group; the total treatment effects for higher-capability suppliers are the sum of the main coefficient and the interaction term. Points show coefficient estimates with error bars indicating 90% confidence intervals. The corresponding regression table can be found in Table D5.

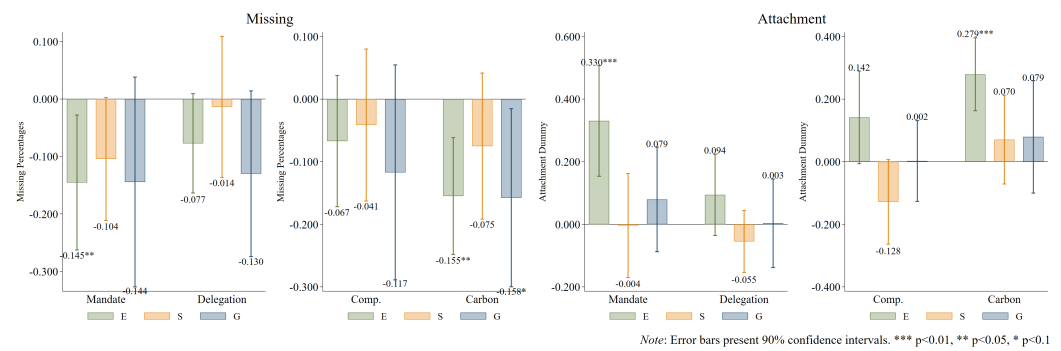


Figure 11: Engagement Responses Across ESG Dimensions

Notes: This figure breaks down treatment effects by reporting domain: Environmental, Social, and Governance. Suppliers concentrate their compliance effort primarily on environmental metrics. This pattern is consistent with the idea that, under organizational bandwidth constraints, firms prioritize tasks with clearer measurement protocols and lower internal coordination costs. Bars show coefficient estimates with 90% confidence intervals. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. The corresponding regression table can be found in Table D6.

Table 1: ESG Indicators on the Digital Platform

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

Notes: The platform tracks 108 supply chain indicators organized by reporting domain. Financial and certification metrics are mandatory for all suppliers. Requirements for environmental, social, and governance indicators vary by experimental treatment arm. Suppliers in the Mandate arm are required to report a fixed set of eight ESG KPIs spanning these three domains, whereas suppliers in the Delegation arm self-select eight KPIs from the whole menu.

Table 2: Balance Table

	Control (C)	Mandate (T1A)	Delegation (T1B)	Comp. (T2A)	Carbon (T2B)
Firms	46	54	55	54	55
Production (CNY '0000)	31,232	67,776 [0.36]	31,537 [0.98]	31,215 [1.00]	68,098 [0.35]
Employment	111	203 [0.25]	304 [0.11]	198 [0.25]	310 [0.11]
Firm age (years)	15.52	14.43 [0.49]	15.82 [0.85]	15.33 [0.90]	14.93 [0.72]
Registered capital (CNY '0000)	4,417	3,316 [0.74]	2,670 [0.61]	2,624 [0.60]	3,349 [0.75]
ESG certificates (count)	1.00	0.83 [0.60]	0.85 [0.66]	0.72 [0.38]	0.96 [0.91]
CMT	0.33	0.37 [0.65]	0.35 [0.84]	0.33 [0.94]	0.38 [0.56]
ODM	0.11	0.11 [0.97]	0.13 [0.78]	0.11 [0.97]	0.13 [0.78]
FAB	0.33	0.28 [0.61]	0.24 [0.33]	0.26 [0.47]	0.25 [0.44]
ACC	0.24	0.24 [0.99]	0.29 [0.56]	0.30 [0.52]	0.24 [0.97]
STI ranking	0.07	0.11 [0.42]	0.09 [0.63]	0.06 [0.84]	0.15 [0.19]
A ranking	0.29	0.27 [0.82]	0.29 [0.99]	0.29 [0.99]	0.27 [0.85]
B ranking	0.63	0.60 [0.77]	0.61 [0.80]	0.65 [0.91]	0.57 [0.53]
Large/medium size	0.15	0.17 [0.85]	0.22 [0.40]	0.19 [0.66]	0.20 [0.53]
Small size	0.52	0.57 [0.60]	0.56 [0.68]	0.56 [0.74]	0.58 [0.55]
Micro size	0.28	0.24 [0.64]	0.22 [0.46]	0.24 [0.64]	0.22 [0.46]

Notes: This table reports pre-treatment means for each experimental arm. For treatment columns, p-values from two-sided t-tests against the control mean are reported in brackets beneath each mean where applicable. T1A = Mandate; T1B = Delegation; T2A = Comprehensive Audit; T2B = Carbon Audit. Because treatment assignment is cross-randomized, Control + T1A + T1B and Control + T2A + T2B each span the full sample. No difference is statistically significant at conventional levels, confirming successful randomization. Production values and registered capital are reported in ten thousands of CNY.

Appendix A. Complete List of ESG 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	2002	Recruitment and Termination	2002001	Total terminated employees
3	Social	2002	Recruitment and Termination	2002002	Termination rate (auto-calculated)
3	Social	2002	Recruitment and Termination	2002003	Total voluntary resignations
3	Social	2002	Recruitment and Termination	2002004	Resignation rate (auto-calculated)

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
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
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	2004	Social Insurance	2004001	Employees covered by social insurance**

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
3	Social	2004	Social Insurance	2004002	Social insurance coverage rate (auto-calculated)
3	Social	2004	Social Insurance	2004003	Employees covered by work injury insurance
3	Social	2004	Social Insurance	2004004	Work injury insurance coverage rate (auto-calculated)
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	2001	Working Hours	2001009	Actual factory working days
3	Social	2001	Working Hours	2001010	Statutory working days in the quarter (auto-filled)
3	Social	2006	Employee Representation	2006001	Democratically elected female worker representatives
3	Social	2006	Employee Representation	2006002	Democratically elected male worker representatives

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
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)
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

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
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)*
5	Certification	3002	Social Responsibility	3002003	WCA (Workplace Conditions Assessment)*
5	Certification	3002	Social Responsibility	3002004	SLCP (Social Responsibility Verification)*

Continued on next page

Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
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)*
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)*

Continued on next page

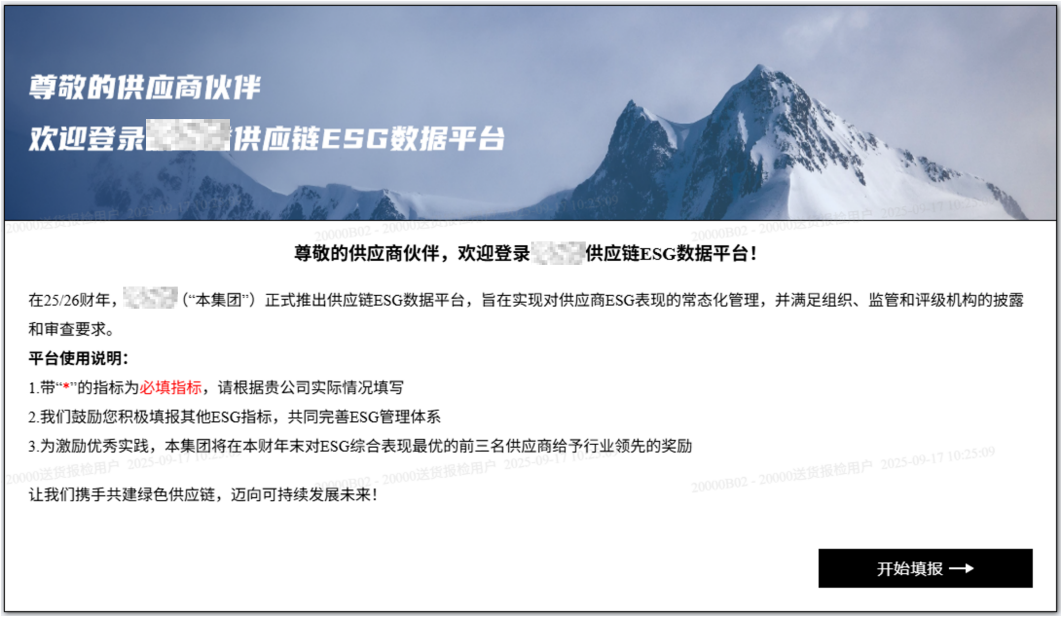
Cat. Code	Category	Sub. Code	Subcategory	Ind. Code	Indicator Name
5	Certification	3003	Material Certification	3003005	OEKO-TEX (Ecological Textile Standard)*
5	Certification	3003	Material Certification	3003006	Bluesign certification*
5	Certification	3003	Material Certification	3003007	FSC (Forest Certification)*

Notes: * Required for all suppliers. ** Required for suppliers in the Mandate treatment arm.

Appendix B. Platform Interface Screenshots

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

Panel A. Interface 1 – Baseline

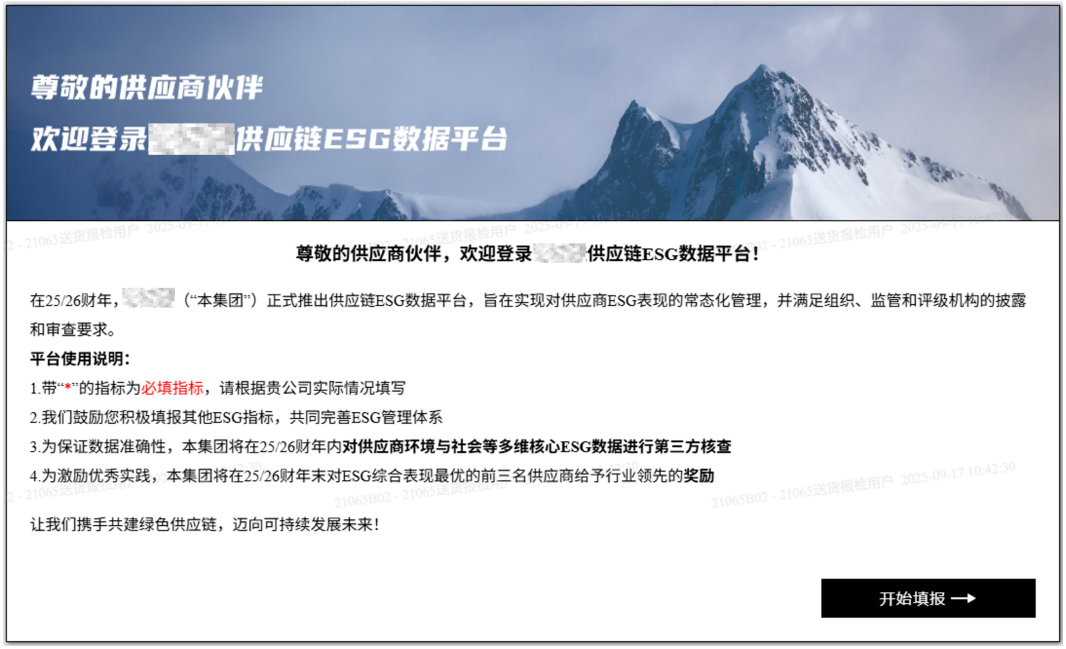


(Original screenshot)



(English translation)

Panel B. Interface 2 – Mandated KPIs, Comprehensive Audit

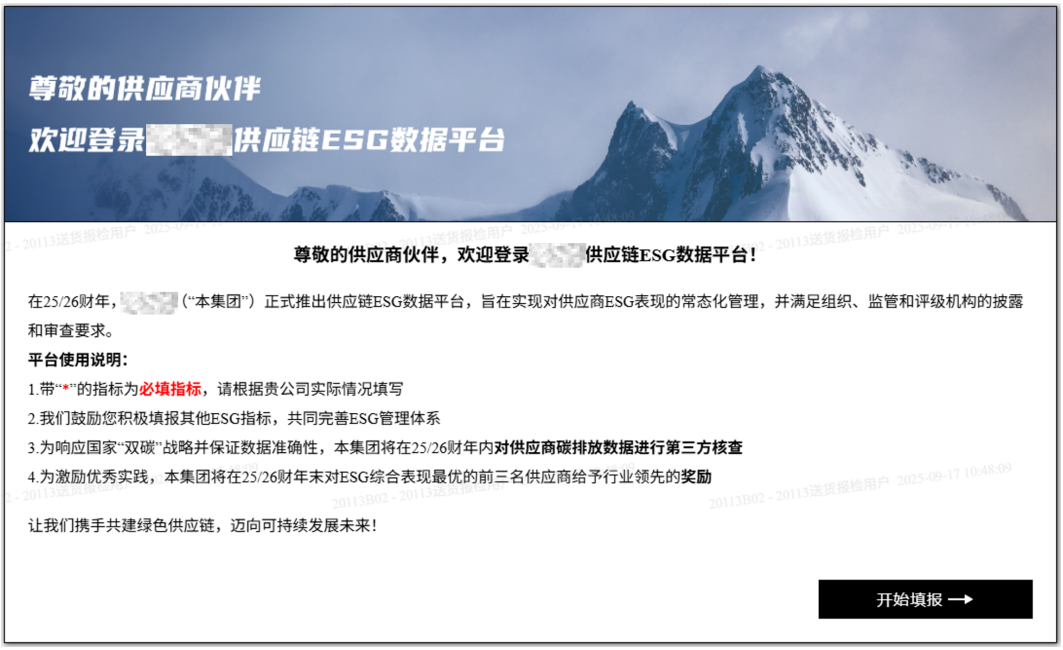


(Original screenshot)



(English translation)

Panel C. Interface 3 – Mandated KPIs, Carbon Audit

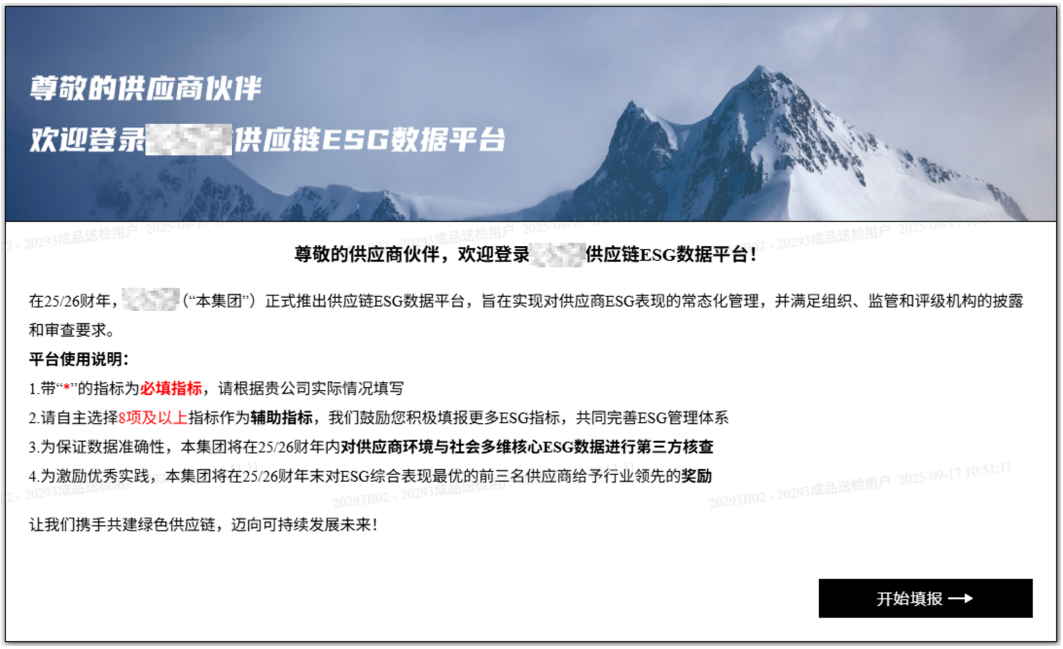


(Original screenshot)

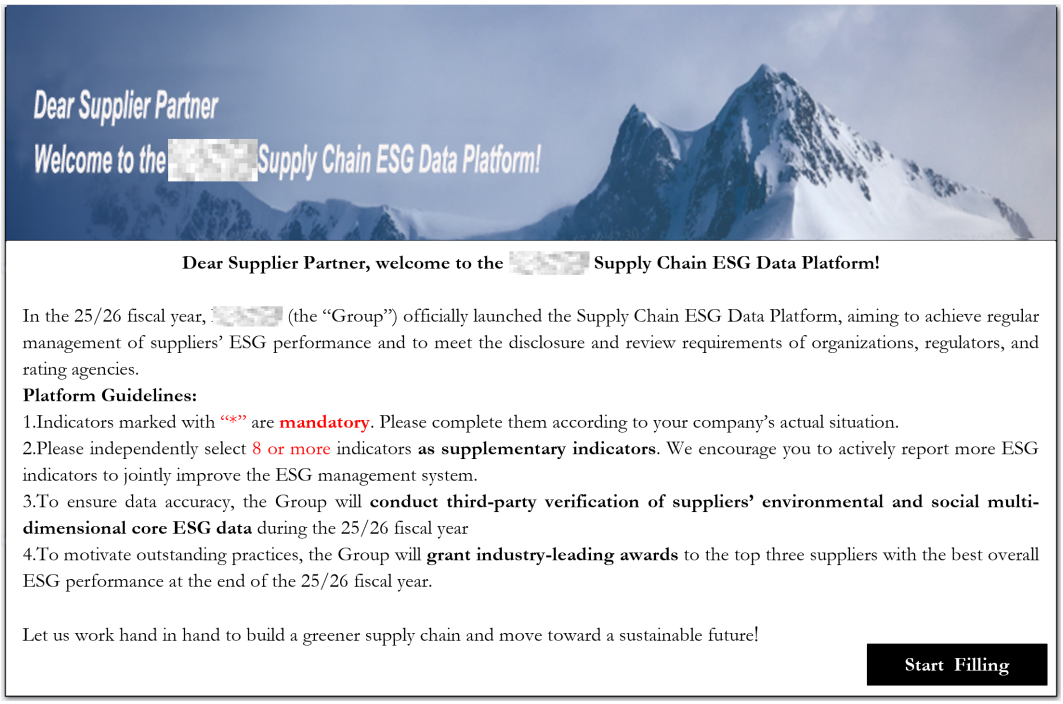


(English translation)

Panel D. Interface 4 – Delegated KPIs, Comprehensive Audit

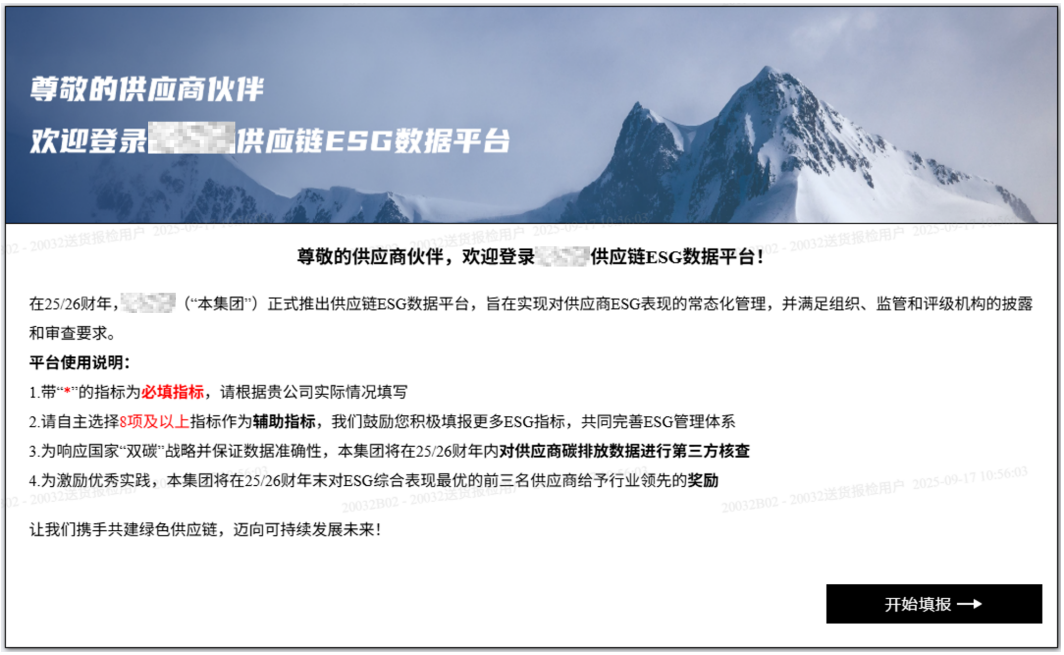


(Original screenshot)



(English translation)

Panel E. Interface 5 – Delegated KPIs, Carbon Audit



(Original screenshot)



(English translation)

Appendix C. Processing ESG Documentation Files

This appendix describes the semi-automated procedure used to convert unstructured supplier-uploaded ESG documentation files into structured variables for empirical analysis. The objective is to transform heterogeneous supporting documents, such as scanned files, spreadsheets, images, and text documents, into three supplier-level outcome measures: *Categories*, *Hard Info*, and *Soft Info*.

The workflow consists of three stages. First, raw files are converted into machine-readable inputs. Second, a large language model (LLM) is used to classify each file into one or more pre-specified ESG categories and to determine whether the evidence is quantitative and verifiable (*hard*) or qualitative and descriptive (*soft*). Third, all machine-generated outputs are reviewed and then aggregated to the supplier level.

C.1. File Parsing and Standardization

Supplier attachments arrive in multiple formats, including PDF files, Word documents, Excel spreadsheets, scanned images, and photographs. To enable unified downstream processing, each file is first 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. For spreadsheets, all worksheets are converted into structured text representations that preserve the row-column relationships needed for interpretation. For images and scanned files, the visual content is passed to a multimodal model through encoded image inputs, allowing direct optical recognition of the underlying evidence. This preprocessing stage standardizes otherwise heterogeneous inputs while preserving the substantive content relevant for classification.

C.2. LLM-Assisted Classification

After parsing, each file is submitted to an LLM-based classifier. The classifier is instructed to return a structured output containing two elements: (i) one or more ESG category labels and (ii) an evidence-type label, either `hard` or `soft`.

The classification taxonomy is closed and fixed ex ante. The admissible categories are: Production, Electricity, Resource Consumption, Solid and Hazardous Waste, Chemicals, Green Initiatives, Water Resources, Employees, Certificates, Community Impact, Corporate Governance, and Other. The prompt also imposes a non-redundancy rule: a file may receive multiple labels only when it contains substantively distinct forms of evidence, but categories are not added on the basis of implied or indirect relationships.

The prompt further distinguishes between two forms of evidence. *Hard* evidence refers to quantitative and directly verifiable materials, such as invoices, metering records, statistical ledgers, and tabular resource-use statements. *Soft* evidence refers to qualitative or descriptive materials, such as policy documents, manuals, reports without quantitative records, and photographs documenting activities or practices. Certificate-related files are uniformly coded as *Certificates*, regardless of whether they pertain to environmental, social-responsibility, or material-certification standards.

C.3. Human Validation and Quality Control

The procedure is intentionally semi-automated rather than fully automated. All machine-generated classifications are reviewed before final aggregation.

This validation step serves two purposes. First, it corrects occasional extraction errors arising from low-quality scans, formatting irregularities, or incomplete optical recognition. Second, it ensures strict adherence to the coding rules, especially in borderline cases involving overlapping categories or mixed qualitative and quantitative content. 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

After validation, document-level classifications are aggregated to the supplier level. Let F_s denote the set of files uploaded by supplier s. For each file $f \in F_s$, let C_{sf} denote the set of assigned ESG categories, and let H_{sf} and S_{sf} denote the subsets of categories supported by hard and soft evidence, respectively.

We then define the three outcome variables as:

$$Categories_s = \left| \bigcup_{f \in F_s} C_{sf} \right|,$$

$$Hard\ Info_s = \left| \bigcup_{f \in F_s} H_{sf} \right|,$$

$$Soft\ Info_s = \left| \bigcup_{f \in F_s} S_{sf} \right|.$$

These variables are counts of unique categories rather than counts of files. As a result, uploading multiple documents within the same category does not mechanically inflate the measured outcome. By contrast, a single file may contribute to multiple categories if it contains substantively distinct information relevant to more than one ESG domain. This aggregation rule ensures that the final measures capture the breadth and type of disclosure rather than the raw volume of uploaded material.

C.5. Illustrative Example

Table C1 provides a de-identified example of the document-level coding and supplier-level aggregation. The example illustrates two important features of the procedure. First, a single spreadsheet may map into multiple categories when it contains several distinct types of

evidence. Second, supplier-level outcomes are based on unique-category aggregation, so the final variables summarize disclosure breadth rather than the number of uploaded files.

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

Index	File	Assigned Category	Evidence Type
1	Chemical usage screenshot.png	Chemicals	Hard
2	2025 SLCP report.pdf	Certificates	Soft
3	Electricity report, Oct. 31, 2025.pdf	Electricity	Hard
4	Tap water data, Oct. 2025.pdf	Water Resources	Hard
5	Steam data, Oct. 2025.pdf	Resource Consumption	Hard
6	Sludge disposal agreement, 2025.pdf	Solid and Hazardous Waste	Soft
7	Code of ethics manual.pdf	Corporate Governance	Soft
8	ESH statistics (including green electricity).xls	Solid and Hazardous Waste	Hard
8	ESH statistics (including green electricity).xls	Resource Consumption	Hard
8	ESH statistics (including green electricity).xls	Water Resources	Hard
8	ESH statistics (including green electricity).xls	Green Initiatives	Hard
8	ESH statistics (including green electricity).xls	Production	Hard
9	Charity fund agreement.pdf	Community Impact	Soft
Supplier-level aggregation		*Categories*	10
		Hard Info	7
		Soft Info	4

Notes. This table provides a de-identified example of the coding procedure. File names are translated and simplified for exposition. The spreadsheet *ESH statistics (including green electricity).xls* receives multiple labels because it contains distinct information on production, resource use, water, green electricity, and waste. Supplier-level variables are constructed from the number of unique categories represented in the validated document-level classifications. Consequently, repeated uploads within the same category do not increase the corresponding outcome variable.

C.6. Interpretation

The resulting variables capture three conceptually distinct margins of supplier disclosure. *Categories* measures the breadth of ESG topics addressed by a supplier's uploaded evidence.

Hard Info measures the breadth of categories supported by quantitative and verifiable documentation. *Soft Info* measures the breadth of categories supported by qualitative or descriptive materials.

The procedure addresses a central measurement problem in our setting: suppliers submit evidence in highly heterogeneous formats and with widely varying informational content. The coding procedure converts these materials into supplier-level measures that are comparable across firms while distinguishing between disclosure scope and evidentiary hardness.

Appendix D. Additional Figures and Tables

Table D1: Main Effects: Vertical Authority Allocation (Mandate vs. Delegation)

	Basic Engagement (Missing)		Advanced Engagement (Attachments)		
	(1) Indicators	(2) Percentages	(3) Categories	(4) Hard Info	(5) Soft Info
Mandate	-15.737**	-0.104**	0.989**	0.767**	0.246
	(6.952)	(0.047)	(0.436)	(0.332)	(0.161)
Delegation	-6.566	-0.043	0.090	0.120	-0.084
	(5.662)	(0.038)	(0.384)	(0.296)	(0.131)
Constant	68.160***	0.452***	1.984***	1.027***	1.098***
	(3.990)	(0.027)	(0.233)	(0.179)	(0.082)
Observations	591	591	591	591	591
R^2	0.430	0.337	0.190	0.183	0.145
p-value	0.101	0.104	0.078	0.094	0.080
Round FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table reports OLS estimates of the effect of vertical authority allocation on supplier engagement outcomes. Missing outcomes are decreasing in engagement, so lower missing rates indicate higher compliance. Attachment outcomes are increasing in engagement, where hard info denotes quantitative supporting documents and soft info denotes qualitative policy documents. The control group comprises suppliers assigned to baseline reporting with no additional ESG requirements. The p-values reported in the row labeled p-value are from two-sided tests of equality between the Mandate and Delegation coefficients within each column. Round and stratum fixed effects are included in all specifications. Standard errors clustered at the stratum level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D2: Main Effects: Horizontal Monitoring Scope (Comprehensive vs. Carbon)

	Basic Engagement (Missing)		Advanced Engagement (Attachments)		
	(1) Indicators	(2) Percentages	(3) Categories	(4) Hard Info	(5) Soft Info
Comprehensive	-7.304	-0.047	0.208	0.215	-0.030
	(6.578)	(0.045)	(0.448)	(0.328)	(0.197)
Carbon	-14.872**	-0.099**	0.857**	0.661**	0.186
	(5.699)	(0.038)	(0.356)	(0.245)	(0.126)
Constant	68.137***	0.452***	1.986***	1.029***	1.099***
	(3.901)	(0.026)	(0.221)	(0.171)	(0.078)
Observations	591	591	591	591	591
R^2	0.427	0.334	0.178	0.172	0.138
p-value	0.148	0.150	0.217	0.168	0.387
Round FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table reports OLS estimates of the effect of horizontal monitoring scope on supplier engagement. Comprehensive monitoring involves third-party verification of all submitted ESG metrics, whereas Carbon monitoring involves verification of carbon-related metrics only. The p-values reported in the row labeled *p*-value are from two-sided tests of equality between the Comprehensive and Carbon coefficients within each column. All other notes are as in Table D1.

Table D3: Interaction Effects: Full 2×2 Treatment Cell Structure

	Basic Engagement (Missing)		Advanced Engagement (Attachments)		
	(1) Indicators	(2) Percentages	(3) Categories	(4) Hard Info	(5) Soft Info
Mandate × Comprehensive	-14.176	-0.093	0.658	0.565	0.122
	(8.784)	(0.060)	(0.544)	(0.416)	(0.217)
Mandate × Carbon	-17.316**	-0.115**	1.338**	0.978**	0.376*
	(6.504)	(0.044)	(0.475)	(0.349)	(0.209)
Delegation × Comprehensive	-0.460	-0.002	-0.235	-0.130	-0.179
	(6.220)	(0.042)	(0.537)	(0.388)	(0.264)
Delegation × Carbon	-12.603	-0.084	0.412	0.367	0.010
	(7.481)	(0.050)	(0.461)	(0.304)	(0.181)
Constant	68.163***	0.452***	1.984***	1.027***	1.098***
	(3.944)	(0.027)	(0.230)	(0.177)	(0.081)
Observations	591	591	591	591	591
R^2	0.441	0.350	0.204	0.194	0.151
Round FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table estimates the engagement effects of each treatment cell relative to the control group using a saturated specification. The bandwidth-minimization bundle, Mandate × Carbon Audit, is the only cell that generates statistically significant improvements on both basic and advanced engagement margins. All other notes are as in Table D1.

Table D4: Capability Heterogeneity: Authority Allocation

	Basic Engagement (Missing)		Advanced Engagement (Attachments)		
	(1)	(2)	(3)	(4)	(5)
	Indicators	Percentages	Categories	Hard Info	Soft Info
Panel A: Quality Rank Heterogeneity					
Mandate	-17.127**	-0.113**	0.583	0.637	-0.065
	(6.783)	(0.047)	(0.489)	(0.402)	(0.150)
Mandate × High Quality	3.476	0.022	1.077*	0.348	0.819***
	(6.813)	(0.045)	(0.498)	(0.481)	(0.248)
Delegation	-4.036	-0.027	-0.578	-0.201	-0.482***
	(6.857)	(0.045)	(0.379)	(0.276)	(0.137)
Delegation × High Quality	-7.199	-0.045	1.894*	0.910	1.129**
	(7.255)	(0.048)	(0.931)	(0.681)	(0.389)
Constant	68.200***	0.452***	1.980***	1.025***	1.097***
	(4.062)	(0.027)	(0.272)	(0.191)	(0.116)
Panel B: Production Heterogeneity					
Mandate	-15.151**	-0.099**	0.586	0.755***	-0.231
	(6.018)	(0.041)	(0.338)	(0.235)	(0.212)
Mandate × High Production	-1.150	-0.009	0.809	0.018	0.963***
	(7.350)	(0.050)	(0.497)	(0.326)	(0.313)
Delegation	-4.053	-0.027	-0.660	-0.163	-0.626**
	(6.894)	(0.046)	(0.442)	(0.312)	(0.206)
Delegation × High Production	-5.024	-0.031	1.491***	0.568*	1.069***
	(6.385)	(0.042)	(0.457)	(0.283)	(0.253)
Constant	68.131***	0.452***	1.994***	1.030***	1.106***
	(3.997)	(0.027)	(0.269)	(0.183)	(0.116)
Panel C: Employment Heterogeneity					
Mandate	-18.320**	-0.120**	0.765*	0.853**	-0.091
	(6.667)	(0.045)	(0.402)	(0.325)	(0.160)
Mandate × High Employment	5.055	0.031	0.422	-0.179	0.651*
	(6.491)	(0.043)	(0.415)	(0.261)	(0.322)
Delegation	-6.043	-0.041	-0.601**	-0.361**	-0.355**
	(8.059)	(0.054)	(0.207)	(0.155)	(0.136)
Delegation × High Employment	-0.985	-0.003	1.429**	0.986*	0.569**
	(8.494)	(0.057)	(0.599)	(0.457)	(0.213)
Constant	68.151***	0.452***	1.985***	1.029***	1.098***
	(3.972)	(0.027)	(0.228)	(0.172)	(0.086)
Observations	591	591	591	591	591
Round FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table reports heterogeneity in the effects of authority allocation by baseline supplier capability. Panels A–C interact Mandate and Delegation with indicators for high quality rank, high production, and high employment, respectively. Columns (1)–(2) report basic engagement outcomes; columns (3)–(5) report advanced engagement outcomes. All other notes are as in Table D1.

Table D5: Capability Heterogeneity: Monitoring Scope

	Basic Engagement (Missing)		Advanced Engagement (Attachments)		
	(1) Indicators	(2) Percentages	(3) Categories	(4) Hard Info	(5) Soft Info
Panel A: Quality Rank Heterogeneity					
Comprehensive	-8.914	-0.058	-0.165	0.004	-0.250
	(7.007)	(0.047)	(0.444)	(0.349)	(0.160)
Comprehensive × High Quality	4.827	0.031	1.145	0.645	0.676*
	(7.734)	(0.052)	(0.666)	(0.524)	(0.363)
Carbon	-11.807*	-0.079*	0.141	0.438	-0.331**
	(5.973)	(0.041)	(0.408)	(0.333)	(0.140)
Carbon × High Quality	-7.403	-0.047	1.753*	0.548	1.265***
	(7.630)	(0.050)	(0.885)	(0.746)	(0.319)
Constant	68.122***	0.452***	1.986***	1.028***	1.099***
	(3.918)	(0.026)	(0.259)	(0.179)	(0.113)
Panel B: Production Heterogeneity					
Comprehensive	-9.919	-0.065	0.132	0.360	-0.307
	(6.358)	(0.043)	(0.482)	(0.295)	(0.270)
Comprehensive × High Production	5.157	0.035	0.146	-0.287	0.541**
	(8.241)	(0.055)	(0.522)	(0.372)	(0.241)
Carbon	-9.481	-0.063	-0.182	0.242	-0.534***
	(6.692)	(0.045)	(0.289)	(0.224)	(0.144)
Carbon × High Production	-11.324*	-0.075*	2.155***	0.878**	1.484***
	(5.751)	(0.038)	(0.462)	(0.309)	(0.274)
Constant	68.165***	0.452***	1.989***	1.028***	1.104***
	(3.901)	(0.026)	(0.253)	(0.176)	(0.112)
Panel C: Employment Heterogeneity					
Comprehensive	-10.595	-0.070	0.057	0.211	-0.174
	(7.672)	(0.052)	(0.416)	(0.273)	(0.221)
Comprehensive × High Employment	6.492	0.045	0.295	0.006	0.282
	(6.913)	(0.046)	(0.472)	(0.425)	(0.214)
Carbon	-12.903*	-0.085*	0.022	0.220	-0.302***
	(6.609)	(0.044)	(0.223)	(0.228)	(0.083)
Carbon × High Employment	-4.029	-0.028	1.719**	0.908**	1.006**
	(7.401)	(0.049)	(0.610)	(0.370)	(0.344)
Constant	68.143***	0.452***	1.985***	1.028***	1.098***
	(3.897)	(0.026)	(0.221)	(0.169)	(0.087)
Observations	591	591	591	591	591
Round FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table reports heterogeneity in the effects of monitoring scope by baseline supplier capability. Panels A–C interact Comprehensive and Carbon monitoring with indicators for high quality rank, high production, and high employment, respectively. Columns (1)–(2) report basic engagement outcomes; columns (3)–(5) report advanced engagement outcomes. All other notes are as in Table D1.

Table D6: Heterogeneity Across ESG Dimensions

	Missing Percentages			Attachment Dummy		
	(1)	(2)	(3)	(4)	(5)	(6)
	E	S	G	E	S	G
Panel A: Authority Allocation						
Mandate	-0.145**	-0.104	-0.144	0.330***	-0.004	0.079
	(0.066)	(0.060)	(0.102)	(0.099)	(0.093)	(0.094)
Delegation	-0.077	-0.014	-0.130	0.094	-0.055	0.003
	(0.049)	(0.069)	(0.081)	(0.073)	(0.055)	(0.079)
Constant	0.741***	0.361***	0.406***	0.318***	0.198***	0.151***
	(0.035)	(0.043)	(0.059)	(0.045)	(0.050)	(0.054)
Panel B: Monitoring Scope						
Comprehensive	-0.067	-0.041	-0.117	0.142	-0.128	0.002
	(0.059)	(0.068)	(0.096)	(0.083)	(0.076)	(0.072)
Carbon	-0.155**	-0.075	-0.158	0.279***	0.070	0.079
	(0.052)	(0.065)	(0.080)	(0.065)	(0.080)	(0.101)
Constant	0.741***	0.361***	0.406***	0.319***	0.198***	0.151***
	(0.034)	(0.042)	(0.058)	(0.042)	(0.048)	(0.054)
Observations	591	591	591	591	591	591
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum

Notes: This table decomposes engagement outcomes by ESG reporting domain. Columns (1)–(3) and (4)–(6) report effects on the environmental (E), social (S), and governance (G) dimensions, respectively. The attachment dummy in columns (4)–(6) is an indicator for whether a supplier uploaded at least one supporting document for the given domain. Treatment effects are concentrated in the environmental dimension. All other notes are as in Table D1.

Table D7: Robustness of Main Results: Vertical Authority Allocation

	Missing Percentages				Attachment Categories			
	(1) Main	(2) Wild Bootstrap	(3) Firm Clusters	(4) Collapsed Sample	(5) Main	(6) Wild Bootstrap	(7) Firm Clusters	(8) Collapsed Sample
Mandate	-0.104**	-0.104**	-0.104***	-0.108**	0.989**	0.989**	0.989**	1.018**
	(0.047)	(0.048)	(0.036)	(0.047)	(0.436)	(0.441)	(0.412)	(0.442)
Delegation	-0.043	-0.043	-0.043	-0.043	0.090	0.090	0.090	0.114
	(0.038)	(0.038)	(0.036)	(0.040)	(0.384)	(0.389)	(0.409)	(0.395)
Constant	0.452***	0.290***	0.452***	0.454***	1.984***	2.533***	1.984***	1.970***
	(0.027)	(0.029)	(0.028)	(0.028)	(0.233)	(0.296)	(0.306)	(0.238)
Observations	591	591	591	150	591	591	591	150
R^2	0.337	0.337	0.337	0.216	0.190	0.190	0.190	0.195
Round FE	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Firm	Stratum	Stratum	Stratum	Firm	Stratum

Notes: This table reports robustness checks for the main effects of vertical authority allocation on supplier engagement. Columns (1) and (5) reproduce the baseline specification from Table D1. Columns (2) and (6) report wild-bootstrap inference. Columns (3) and (7) report standard errors clustered at the firm level rather than the stratum level. Columns (4) and (8) collapse the panel to the firm level by averaging outcomes across rounds and therefore exclude round fixed effects. In the wild-bootstrap columns, estimation uses `areg` rather than `reghdfe`, because `reghdfe` does not support the `vce(bootstrap)` option. This implementation absorbs stratum fixed effects while including round indicators as explicit regressors, which mechanically changes the constant term but leaves the treatment estimates substantively unchanged. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D8: Robustness of Main Results: Horizontal Monitoring Scope

	Missing Percentages				Attachment Categories			
	(1) Main	(2) Wild Bootstrap	(3) Firm Clusters	(4) Collapsed Sample	(5) Main	(6) Wild Bootstrap	(7) Firm Clusters	(8) Collapsed Sample
Comprehensive	-0.047	-0.047	-0.047	-0.052	0.208	0.208	0.208	0.219
	(0.045)	(0.045)	(0.035)	(0.046)	(0.448)	(0.453)	(0.413)	(0.446)
Carbon	-0.099**	-0.099**	-0.099***	-0.098**	0.857**	0.857**	0.857**	0.888**
	(0.038)	(0.039)	(0.037)	(0.038)	(0.356)	(0.360)	(0.409)	(0.367)
Constant	0.452***	0.289***	0.452***	0.454***	1.986***	2.551***	1.986***	1.973***
	(0.026)	(0.028)	(0.027)	(0.027)	(0.221)	(0.281)	(0.304)	(0.224)
Observations	591	591	591	150	591	591	591	150
R^2	0.334	0.334	0.334	0.206	0.178	0.178	0.178	0.182
Round FE	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	Stratum	Stratum	Firm	Stratum	Stratum	Stratum	Firm	Stratum

Notes: This table reports robustness checks for the main effects of horizontal monitoring scope on supplier engagement. Columns (1) and (5) reproduce the baseline specification from Table D2. Columns (2) and (6) report wild-bootstrap inference. Columns (3) and (7) report standard errors clustered at the firm level rather than the stratum level. Columns (4) and (8) collapse the panel to the firm level by averaging outcomes across rounds and therefore exclude round fixed effects. In the wild-bootstrap columns, estimation uses `areg` rather than `reghdfe`, because `reghdfe` does not support the `vce(bootstrap)` option. This implementation absorbs stratum fixed effects while including round indicators as explicit regressors, which mechanically changes the constant term but leaves the treatment estimates substantively unchanged. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.