

Bridging the Carbon Trust Deficit: Strategic Signaling and Regulatory Retrenchment in Supply Chains

Jiguang Wang

Shanxi University

Kexin Wang

wangkexin6@sxu.edu.cn

Shanxi University

Research Article

Keywords: Low-carbon supply chain, Carbon trust, Evolutionary game theory, Information asymmetry, Strategic regulatory retrenchment

Posted Date: July 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8988938/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Bridging the Carbon Trust Deficit: Strategic Signaling and Regulatory Retrenchment in Supply Chains

Corresponding author: Kexin Wang

Abstract: Achieving credible carbon disclosure in multi-tier supply chains is persistently hindered by information asymmetry and the resulting “carbon trust deficit”. This study investigates the dynamic transition from coercive regulation to an autonomous, market-based governance regime using a four-player evolutionary game theoretical framework involving a government regulator, a focal firm, a third-party certifier, and an upstream supplier. By endogenizing the credibility of carbon information, we model how regulatory penalties, green premiums, and reputational mechanisms jointly shape the evolution of carbon trust under replicator dynamics. The analytical and numerical results reveal that the system can endogenously converge to an “Autonomous Carbon Trust Equilibrium”, where the government retreats to a supportive role. We identify the focal firm’s strategic trust repair as a key catalyst that initiates this transition. Furthermore, we uncover a critical threshold effect in the reputation mechanism, demonstrating that market discipline can effectively substitute for administrative penalties once reputational losses exceed a specific tipping point. These findings provide a dynamic explanation of when and how governance can shift from front-line enforcement to market-disciplined self-enforcement, and suggest that policy should prioritize activating market catalysts and optimizing verification infrastructure to support this transition.

Keywords: Low-carbon supply chain; Carbon trust; Evolutionary game theory; Information asymmetry; Strategic regulatory retrenchment

1 Introduction

Global efforts toward carbon neutrality have intensified over the past decade as climate science has underscored the urgency of deep and rapid emission reductions. In this context, firms are under increasing pressure not only to cut their own direct emissions, but also to manage and report emissions that occur along their supply chains. A growing body of empirical work shows that upstream and downstream Scope 3 emissions often account for the majority of a company’s carbon footprint, making supply chains a central arena for low-carbon transformation (Fan, Wu, & He, 2024; Qian & Zhou, 2025). As carbon pricing, green finance and low-carbon labels diffuse across markets, reliable information on product and supply chain emissions has become a critical input for regulatory compliance, operational decisions and low-carbon business strategy (Ma et al., 2024).

However, generating and governing such information in multi-tier global networks remains challenging. Complex supply chain structures, heterogeneous measurement capabilities and fragmented information systems create persistent carbon information asymmetries among upstream supplier, downstream focal firm, regulator and end-users (Wang et al., 2023). Even when firms disclose Scope 3

emissions, these disclosures often rely on secondary data, industry averages or coarse allocation rules rather than supplier-specific measurements, introducing substantial uncertainty and inconsistency in reported figures (Butt et al., 2025). In parallel, the credibility of environmental and carbon-related claims faces increasing scrutiny. Recent empirical work shows that perceived greenwashing undermines customer satisfaction, trust and brand attachment. Consumers interpret inconsistencies between stated environmental ambitions and actual implementation as signs of corporate hypocrisy, and they adjust their responses to firms accordingly (Ioannou, Kassinis, & Papagiannakis, 2023; Tu et al., 2024). These developments indicate that the pivotal challenge is not merely to disclose carbon data, but to ensure that this information is sufficiently reliable to be trusted and acted upon.

In this study, we formally define the “Carbon Trust Deficit” not merely as a lack of information, but as a systemic governance failure where the perceived credibility of disclosed carbon data falls below the critical threshold required for stakeholder decision-making. Conceptually, this deficit represents a collective action dilemma (Yi & Cui, 2019; Olson, 1965) driven by misaligned incentives. While all supply chain participants would collectively benefit from a high-trust, verified low-carbon ecosystem, individual rationality drives players toward opportunistic behaviors: suppliers have incentives to misreport to reduce abatement costs; certifiers may lower verification standards to cut costs; and focal firms may engage in superficial greenwashing. These strategic interactions create a “Prisoner’s Dilemma” dynamic, causing the system to converge toward a Pareto-inefficient equilibrium of mutual distrust. Therefore, the “Carbon Trust Deficit” is essentially a structural breakdown in the credibility mechanism of the supply chain, creating a barrier that prevents the realization of green premiums. For regulators, limited trust in self-reported and supplier-reported emissions raises monitoring and enforcement costs and may weaken the effectiveness of instruments that rely on robust emissions baselines, while focal firms face reputational and legal risks if “low-carbon” products are later challenged as misleading. Upstream suppliers may find that investments in robust carbon accounting and genuine abatement are not financially rewarded when buyers cannot distinguish credible from non-credible disclosures, dampening incentives for substantive low-carbon efforts, and repeated controversies around environmental labelling and offset-based “neutrality” claims risk eroding consumers’ willingness to pay for genuine low-carbon attributes (Tu et al., 2024; Papagiannakis et al., 2023). These interconnected tensions pose a central question: how can carbon trust be established, repaired and sustained when players face both economic incentives to misreport and intensifying external scrutiny? Addressing this question is crucial not only for the effective implementation of emerging climate and trade policies, but also for understanding how governance arrangements can support credible, long-term low-carbon transformation in supply chains.

Scholars have approached these issues from several directions. Research on low-carbon supply chains frequently employs evolutionary game models to analyse how regulation and contracts shape emission reduction and technology adoption, but largely treats carbon information as exogenous and overlooks the strategic credibility

of disclosure itself (Zhang et al., 2025). Studies on digitalisation highlight how platforms and data technologies improve information integration and monitoring, yet they seldom examine how these tools interact with firms' incentives to manipulate or withhold carbon data, or how they influence the long-term evolution of trust (Fan et al., 2024; Ma et al., 2024). Meanwhile, work on information asymmetry, reputation and blockchain demonstrates how reputational mechanisms and well-designed incentives can foster operational trust in uncertain or technology-enabled settings (Wang et al., 2023; Zhang et al., 2022). What remains less understood is how the credibility of carbon disclosure co-evolves with regulation, green premiums, reputational sanctions and third-party certification in a multi-player supply chain setting where the reliability of carbon information itself is a strategic variable.

Motivated by these challenges, this study aims to bridge these gaps by investigating the dynamics of carbon trust within a supply chain governance system. We pose three pertinent research questions:

RQ1: Under carbon information asymmetry, how do regulatory penalties, green premiums and reputational losses from detected misreporting jointly shape the evolutionary dynamics of carbon disclosure and carbon trust in supply chains?

RQ2: Under what conditions can the four-party governance system shift from coercion-dominated regulation to an autonomous, trust-based regime sustained mainly by market and reputational incentives?

RQ3: How do changes in external parameters affect the stability of equilibrium strategy profiles and the long-run level of carbon trust in low-carbon supply chains?

To address these questions, we develop a four-party evolutionary game model involving a government regulator, a focal firm, a third-party certifier, and a supplier. In the model, regulatory penalties and enforcement costs, green premiums, certification fees, reputational losses from detected misreporting, and carbon price signals jointly determine the players' payoffs. This setup explicitly links governance parameters to the evolutionary dynamics of carbon disclosure and trust. We derive the replicator dynamics and equilibrium properties of this system and conduct numerical experiments under alternative policy and market scenarios to examine how different governance configurations shape the evolution of strategies and the emergence or breakdown of carbon trust in supply chains.

This paper makes three primary contributions. First, at the conceptual level, it introduces the notion of a *carbon trust deficit* in supply chains and frames carbon trust as the outcome of interacting regulatory, price-based and reputational mechanisms, enriching research that typically considers emissions, technology choices or digitalisation in isolation (Fan et al., 2024; Ma et al., 2024). Second, at the methodological level, it extends evolutionary game models in green and low-carbon governance to a four-party setting that endogenises the credibility of carbon information and incorporates both policy parameters and dynamic market conditions. Third, from a managerial and policy perspective, it identifies the conditions under which governments should rely on coercive regulation versus step back and allow

market- and reputation-based mechanisms, anchored in green premiums and third-party certification, to sustain autonomous carbon trust in supply chains.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the model setup, key assumptions, payoff structure and replicator dynamics. Section 4 analyzes the evolutionary dynamics and equilibrium properties. Section 5 reports numerical experiments and sensitivity analyses that illustrate the evolutionary outcomes under different policy and market scenarios. Section 6 concludes by summarizing the main findings and discussing managerial and policy implications, limitations and directions for future research.

2 Literature Review

2.1 Carbon information asymmetry and carbon trust in supply chains

Carbon-related decisions in supply chains are strongly shaped by information asymmetry. Upstream firms often hold private information about emission intensities, abatement costs and green innovation capabilities, while downstream focal firms and regulators observe only imperfect signals. Recent analytical work shows that such asymmetry significantly affects contract design and incentives for low-carbon or green product innovation, and that information rents can lead to distorted effort and profit allocation (Ni et al., 2024). Dynamic trade-credit models under information asymmetry further indicate that opportunistic reporting behaviour emerges when costs are private and that reputation-based compensation may be required to realign incentives (Wang et al., 2023).

On this basis, a growing literature examines carbon disclosure and transparency more directly. Studies on mandatory carbon disclosure and supply chain engagement find that lead firms increasingly require suppliers to report emissions, but disclosure quality remains uneven and depends on measurement capabilities and institutional pressures (Dahlmann & Roehrich, 2019; Deng et al., 2024). Research on carbon transparency highlights that institutional forces and firms' innovative capacities jointly drive supplier disclosure and reduce decision uncertainty (Ma et al., 2024). At the same time, work on greenwashing shows that perceived inconsistencies between stated environmental ambitions and actual practices erode customer satisfaction and trust (Ioannou, Kassinis, & Papagiannakis, 2023; Tu et al., 2024). Overall, these strands suggest that the credibility of carbon information is critical, yet "carbon trust" itself, confidence in the reliability of carbon information in supply chains, has rarely been modeled as an endogenous, evolving construct.

2.2 Governance mechanisms: regulation, third-party certification and digital platforms

Three key research streams address governance mechanisms for low-carbon supply chain transformation. First, environmental regulation studies highlight that calibrated policy mixes can steer firms toward low-emission equilibria by shaping green innovation and compliance strategies, while weak enforcement or excessive burdens risk high-emission lock-in (Zhang et al., 2025; Eghbali, Rasti-Barzoki, & Safarzadeh,

2024; Zhang, Li, & Lu, 2024). Empirical work confirms well-designed regulation, paired with reputation and taxation, promotes corporate green innovation, yet most studies assume regulators have access to reliable environmental information for enforcement.

Second, third-party governance research emphasizes certification as a trust-repair tool, leveraging third parties' neutrality and guarantees to validate green claims and mitigate greenwashing (Bachmann, Gillespie, & Priem, 2015). Game-theoretic and evolutionary analyses show credible (though imperfect) certification reduces misrepresentation incentives, but note certifiers face moral hazard requiring incentive and constraint schemes (Cai, Li & Wang, 2024; Chen, Ping & Liang, 2023). These studies underscore third-party importance but rarely focus on carbon-specific information or long-term trust dynamics.

Third, digital governance research highlights platforms and blockchain as enablers of transparency and trust. Empirical evidence links supply chain digitalization to enhanced information integration, monitoring, and improved green performance (Ma et al., 2024; Hu, Deng, & Li, 2024), while evolutionary models show digital tools lower coordination and monitoring costs to accelerate green transformation (Fan, Wu, & He, 2024). Blockchain frameworks further offer traceability and decentralized verification to bridge trust gaps (Sabeti et al., 2019; Zhang et al., 2022).

2.3 Evolutionary game models in green and low-carbon governance

Evolutionary game theory has been widely used to study behavioural dynamics in environmental and low-carbon governance. Many studies construct two- or three-player evolutionary games to analyse how governments, firms and sometimes consumers adjust strategies over time under bounded rationality and imitation. In the context of green technology innovation and emissions reduction, evolutionary models show how policy support, enforcement strength and cost–benefit structures determine whether enterprises converge to green strategies or remain in non-cooperative equilibria (Zhang et al., 2025; Eghbali, Rasti-Barzoki, & Safarzadeh, 2024). Other work applies evolutionary games to green supply chain transformation, examining how government subsidies, carbon prices and consumer green preferences interact with firms' operational choices and digital cooperation decisions (Fan et al., 2024).

More recent contributions extend evolutionary analysis to multi-player settings involving platforms, third-party environmental service providers and blockchain-based networks. For example, Zhang et al., (2022) develop evolutionary game-based incentive models and smart contracts to sustain trust and participation in a blockchain-based shared manufacturing network, showing how reputation and economic incentives jointly influence strategic behaviour in a technology-enabled environment. Tripartite games involving governments, polluters and environmental service enterprises reveal rich evolutionary dynamics under different reward-penalty, pricing and supervision schemes (Li et al., 2023; Gao et al., 2023). Nevertheless, existing evolutionary game studies in green governance focus mainly on decisions such as whether to innovate, comply with regulation or cooperate in emissions reduction. The

credibility of carbon information itself is rarely treated as an endogenous strategic variable, and four-player interactions among government, focal firm, certifier and supplier in the context of carbon disclosure and trust repair remain largely unexplored.

2.4 Research gap and positioning

Synthesising these streams, several gaps emerge. Research on information asymmetry and carbon disclosure has shown how private information, institutional pressures and disclosure mechanisms affect carbon management, but has not explicitly conceptualised or modeled carbon trust as a dynamic outcome of strategic disclosure, verification and reputational processes. Studies on regulation, third-party certification and digital platforms generally analyse these mechanisms in isolation and treat the reliability of environmental information as given, rather than examining how they jointly shape trust under carbon information asymmetry. Finally, while evolutionary games are widely used in green and low-carbon governance, they mainly address innovation, compliance and cooperation decisions and seldom centre on the credibility of carbon information or on the transition from coercion-dominated regulation to autonomous, trust-based governance in multi-player supply chains.

To address these gaps, this paper develops a four-party evolutionary game framework grounded in Signaling Theory (Connelly et al., 2025; Taj, 2016). We conceptualize the focal firm's trust repair not merely as an operational cost, but as a strategic signal to overcome the "lemons problem" in carbon markets, allowing firms to convert unobservable green efforts into observable market value. Specifically, the model captures the dynamic interactions among a government regulator, a focal firm, a third-party certifier, and an upstream supplier. In this system, carbon disclosure strategies and trust are jointly determined by the interplay of regulatory penalties, green premiums, reputational losses, and certification behaviors. Furthermore, numerical experiments under alternative policy and market scenarios are conducted to examine how different governance configurations influence the emergence, repair, and long-run stability of carbon trust.

3 Model and Assumptions

3.1 Problem description

This study models a low-carbon supply chain system involving four key players: a government regulator, a downstream focal firm, a third-party certifier, and an upstream supplier. Each participant is boundedly rational and aims to maximize their own benefits through strategic interactions. The core challenge is whether credible carbon disclosure and high carbon trust can emerge endogenously as players adjust their strategies based on payoffs.

Government regulator: Responsible for enforcing carbon information disclosure rules, the regulator chooses between strict regulation, with intensive monitoring and penalties for misreporting, and lax regulation, with weaker enforcement and lower

penalties. Strict regulation improves the accuracy of carbon information but incurs higher enforcement costs.

Focal firm: This downstream firm can adopt responsible procurement, which involves strict supplier screening and offering green premiums to low-carbon suppliers, or conventional procurement, where no premiums are paid, and suppliers' carbon performance is not verified. Responsible procurement enhances brand reputation but raises procurement and screening costs.

Third-party certifier: An independent agency that certifies suppliers' carbon information, choosing between strict certification, involving rigorous checks and higher costs, and lax certification, with superficial checks and lower costs, but a higher risk of reputational damage if misreporting is revealed.

Upstream supplier: A supplier, responsible for generating emissions, chooses between truthful disclosure of their carbon data, potentially gaining green premiums, or opportunistic disclosure, where they understate emissions or overstate low-carbon performance to reduce costs and capture premiums. Opportunistic suppliers face regulatory penalties and reputational losses if detected.

These four players interact within a reward-punishment framework, forming a closed-loop system that determines the evolution of carbon trust. The government regulates the focal firm, the third-party agency, and the supplier. The focal firm incentivizes the supplier through green order premiums and relies on the third-party agency for verification services. The third-party certifier's certification directly influences both the supplier's disclosure behavior and the focal firm's information accuracy. Consequently, the supplier's disclosure strategy is simultaneously constrained by government inspections, the focal firm's procurement conditions, and third-party certification costs.

Over time, these players, each acting under bounded rationality, adjust their strategy shares through payoff-based learning, giving rise to the evolutionary dynamics of carbon disclosure and trust in the supply chain. The interaction structure among these participants is visualized in Figure 1, which depicts the flow of regulatory, commercial, and verificational influences. Verification

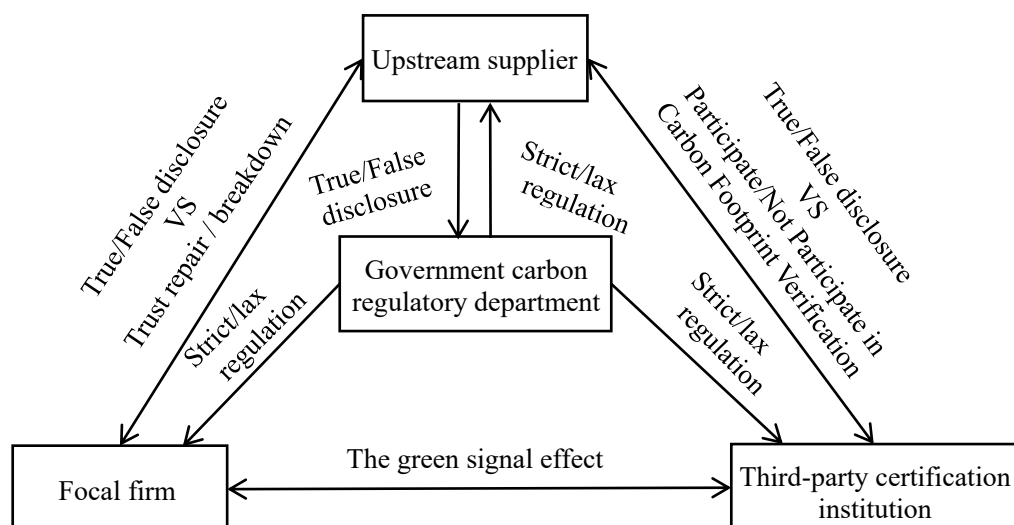


Figure 1: Interaction structure of the four-player evolutionary game

The specific strategic choices available to each player are systematically summarized in Table 1, which outlines the key choices made by the government, the focal firm, the third-party certifier, and the upstream supplier. These strategies, ranging from regulatory decisions to carbon disclosure practices, collectively shape the evolution of carbon trust in the supply chain.

Table 1. Strategic choices of game participants

Participant	Strategy 1	Strategy 2
Government	Strict Regulation	Lax Regulation
Focal Firm	Trust Repair	Trust Breakdown
Third-Party Certifier	Participate	Not Participate
Upstream Supplier	Truthfully Disclose	False Disclosure

3.2 Model assumptions s and parameter settings

Based on the feasibility of the research, we made the following assumptions:

A1. Bounded Rationality and Payoff-Based Learning. All players are boundedly rational. They adjust their strategies over time through a process of imitation and learning, favoring strategies that have yielded higher payoffs. This learning process is formally described by the replicator dynamics.

A2. Information Asymmetry. Asymmetric information exists regarding the upstream supplier's true carbon footprint. The supplier possesses private information, while other players cannot observe this without incurring cost. This asymmetry is the root cause of the "trust crisis" and potential "greenwashing". This is consistent with the principal-agent theory literature in supply chain management (e.g., Aftab et al., 2022).

A3. Government Regulation. The government regulator's strategy set is {Strict, Lax}. Strict regulation incurs a higher cost (C_1) than lax regulation (λC_1), but it improves the accuracy of carbon information and enables the imposition of fines (π_2, π_3, π_4) on non-compliant players. The government also provides rewards (S_2, S_3, S_4) to incentivize positive behaviors.

A4. Focal Firm's Strategic Choice. The focal firm's strategy set is {Trust Repair, Trust Breakdown}. Drawing on Signaling Theory, the Trust Repair strategy requires an upfront screening cost (C_2). This cost serves as a credible signal of low-carbon commitment, enabling the firm to screen out non-compliant suppliers and capture the green premium (r_2) and market share expansion (α). Conversely, Trust Breakdown avoids this cost but relies on information asymmetry to gain short-term risk benefits (η_2) while risking penalties (π_2).

A5. Third-Party Certification. The third-party certifier's strategy set is {Participate, Not Participate}. Participation entails a cost (C_3) but can generate revenue through

collaboration with the focal firm (r_3) and government rewards (s_3). Non-participation risks a fine (π_3), and lax certification carries reputational risks.

A6. Supplier's Disclosure Decision. The supplier's strategy set is {Truthful Disclosure, False Disclosure}. The cost of disclosure ($C_{43}, C_{44}, C_{41}, C_{42}$) varies with others' strategies. Truthful disclosure may bring rewards (s_4), while false disclosure risks penalties (π_4) and market reputation loss (βR_4).

A7. Role of Carbon Price Signals. The green premium (r_2) captured by the focal firm is positively correlated with the market carbon price (P_C), modeled as $r_2 = 20 + 2kP_C$. This assumption incorporates the critical influence of external carbon markets and carbon pricing policies on the economic incentives within the supply chain, which is widely acknowledged in environmental economics (e.g., Qian & Zhou, 2025).

A8. Parameter Stability. All structural parameters (e.g., costs, rewards, penalties) remain constant within each analyzed scenario. This assumption simplifies the analysis and allows us to focus on how strategic interactions evolve under stable conditions.

To clarify the notation used in the model, we define the specific parameter symbols and their meanings in Table 2.

Table 2. Parameter setting and symbol description

Game player	Parameter	Meaning
Government Regulator	R_1	Baseline revenue
	C_1	Cost of strict regulation
	λC_1	Cost of lax regulation, where $\lambda \in [0, 1]$
	r_1	Additional environmental benefit from the supplier's truthful disclosure
	c_1	Additional governance cost from the supplier's false disclosure
	π_2	Fine imposed on the focal firm for trust breakdown
	π_3	Fine imposed on the third-party certifier for non-participation
Focal Firm	π_4	Fine imposed on the upstream supplier for false disclosure
	R_2	Baseline revenue
	C_2	Cost of implementing trust repair
	s_2	Government reward for trust repair
	e	Environmental benefit from the supplier's truthful disclosure
	r_2	Green premium obtained when the third-party certifier and the supplier discloses truthfully
	P_C	Daily average price in the national carbon market
	η_2	Additional risk benefit from choosing trust breakdown
	π_2	Loss (fine) from choosing trust breakdown
Third-Party Certifier	α	Market share amplification factor due to trust repair, $\alpha \in [0, 1]$
	R_3	Baseline revenue
	r_3	Collaboration benefits from the focal firm's trust repair efforts
	C_3	Cost of participating in carbon footprint verification
	s_3	Government reward for participation

	π_3	Fine for non-participation
	R_4	Baseline revenue
	C_{41}	Cost of false disclosure when the focal firm restores trust and the third-party certifier
Upstream	C_{42}	Cost of false disclosure when the third-party participates but the focal firm does not restore trust
Supplier	C_{43}	Cost of truthful disclosure when the focal firm restores trust
	C_{44}	Cost of truthful disclosure when the focal firm does not restore trust
	s_4	Government reward for truthful disclosure
	π_4	Fine for false disclosure
	β	Market reputation loss coefficient for false disclosure, $\beta \in [0,1]$

3.3 Payoff matrix and evolutionary dynamics

Based on the assumptions above, we define the payoff structures for the four players. The complete payoff matrix for all 16 strategy combinations is provided in Table 3. The following outlines the core components of each player's payoff. Each player's payoff is contingent on the strategies chosen by all others. We introduce the replicator dynamics to model strategic evolution over time based on the payoffs. Let x , y , z and w represent the probabilities of the government choosing strict regulation, the focal firm opting for trust repair, the third-party certifier participating, and the upstream supplier disclosing truthfully, respectively. Based on the payoff structure in Table 3, we employ replicator dynamics to describe how the strategy shares evolve over time, i.e. $\dot{F}(x) = x(U_{G1} - U_G)$, and similarly for y , z and w . The explicit forms of the replicator dynamics and their stability properties are derived in Section 4.

Table 3. The payoff matrix for supply chain trust repair under asymmetric carbon footprint information among four players

		Strict government regulation (x)		Lax government regulation (1-x)	
		Real disclosure from upstream supplier (w)	False disclosure from upstream supplier (1-w)	Real disclosure from upstream supplier (w)	False disclosure from upstream supplier (1-w)
Focal firm trust repair (y)	Participation of third-party certification (z)	$a_1=R_1-C_1+r_1-s_2-s_3-s_4$	$a_2=R_1-C_1-c_1+\pi_4-s_2-s_3$	$a_3=R_1-\lambda C_1+r_1$	$a_4=R_1-\lambda C_1-c_1$
		$b_1=(1+\alpha)R_2-C_2+s_2+e+r_2$	$b_2=(1+\alpha)R_2-C_2+s_2+r_2$	$b_3=(1+\alpha)R_2-C_2+e+r_2$	$b_4=(1+\alpha)R_2-C_2+r_2$
		$c_1=R_3+r_3+s_3-C_3$	$c_2=R_3+r_3+s_3-C_3$	$c_3=R_3+r_3-C_3$	$c_4=R_3+r_3-C_3$
		$d_1=R_4-C_{43}+s_4$	$d_2=(1-\beta)R_4-C_{41}-\pi_4$	$d_3=R_4-C_{43}$	$d_4=(1-\beta)R_4-C_{41}$
	Non-participation of third-party certification (1-z)	$a_5=R_1-C_1+r_1+\pi_3-s_2-s_4$	$a_6=R_1-C_1-c_1-s_2+\pi_3+\pi_4$	$a_7=R_1-\lambda C_1+r_1$	$a_8=R_1-\lambda C_1-c_1$
		$b_5=(1+\alpha)R_2-C_2+s_2+e$	$b_6=(1+\alpha)R_2-C_2+s_2$	$b_7=(1+\alpha)R_2-C_2+e$	$b_8=(1+\alpha)R_2-C_2$
		$c_5=R_3-\pi_3$	$c_6=R_3-\pi_3$	$c_7=R_3$	$c_8=R_3$
		$d_5=R_4-C_{43}+s_4$	$d_6=(1-\beta)R_4-\pi_4$	$d_7=R_4-C_{43}$	$d_8=(1-\beta)R_4$
Focal firm trust breakdown (1-y)	Participation of third-party certification (z)	$a_9=R_1-C_1+r_1+\pi_2-s_3-s_4$	$a_{10}=R_1-C_1-c_1-s_3+\pi_2+\pi_4$	$a_{11}=R_1-\lambda C_1+r_1$	$a_{12}=R_1-\lambda C_1-c_1$
		$b_9=R_2+\eta_2-\pi_2$	$b_{10}=R_2+\eta_2-\pi_2$	$b_{11}=R_2+\eta_2$	$b_{12}=R_2+\eta_2$
		$c_9=R_3-C_3+s_3$	$c_{10}=R_3-C_3+s_3$	$c_{11}=R_3-C_3$	$c_{12}=R_3-C_3$
		$d_9=R_4-C_{44}+s_4$	$d_{10}=(1-\beta)R_4-C_{42}-\pi_4$	$d_{11}=R_4-C_{44}$	$d_{12}=(1-\beta)R_4-C_{42}$
	Non-participation of third-party certification (1-z)	$a_{13}=R_1-C_1+r_1+\pi_2+\pi_3-s_4$	$a_{14}=R_1-C_1-c_1+\pi_2+\pi_3+\pi_4$	$a_{15}=R_1-\lambda C_1+r_1$	$a_{16}=R_1-\lambda C_1-c_1$
		$b_{13}=R_2+\eta_2-\pi_2$	$b_{14}=R_2+\eta_2-\pi_2$	$b_{15}=R_2+\eta_2$	$b_{16}=R_2+\eta_2$
		$c_{13}=R_3-\pi_3$	$c_{14}=R_3-\pi_3$	$c_{15}=R_3$	$c_{16}=R_3$
		$d_{13}=R_4-C_{44}+s_4$	$d_{14}=(1-\beta)R_4-\pi_4$	$d_{15}=R_4-C_{44}$	$d_{16}=(1-\beta)R_4$

4 Equilibrium and Stability Analysis

4.1 Stability analysis of individual players' strategies

4.1.1 Government regulator

The carbon regulatory authority faces a binary strategic choice: strict regulation and lax regulation. To analyze the evolutionary stability of its strategies, we first define the expected payoffs of these two strategies as U_{G1} (for strict regulation strategy) and U_{G2} (for lax regulation strategy), respectively, and the average expected payoff of the government regulator's strategy set as U_G . The payoff under the strict regulation strategy is contingent on the strategic combinations of the other three players. By expanding the probability-weighted payoff function, we derive the replicator dynamics equation (Eq. 2), the stability of which is analyzed using its first-order derivative (Eq. 3).

$$U_{G1} = yzwa_1 + yz(1-w)a_2 + y(1-z)wa_5 + y(1-z)(1-w)a_6 + (1-y)zwa_9 + (1-y)z(1-w)a_{10} + (1-y)(1-z)wa_{13} + (1-y)(1-z)(1-w)a_{14}$$

$$U_{G2} = yzwa_3 + yz(1-w)a_4 + y(1-z)wa_7 + y(1-z)(1-w)a_8 + (1-y)zwa_{11} + (1-y)z(1-w)a_{12} + (1-y)(1-z)wa_{15} + (1-y)(1-z)(1-w)a_{16}$$

$$U_G = xU_{G1} + (1-x)U_{G2} \quad (\text{Eq. 1})$$

$$F(x) = x(U_{G1} - U_G) = x(1-x)(U_{G1} - U_{G2}) = x(1-x)A(y, z, w) \quad (\text{Eq. 2})$$

$$F'(x) = (1-2x)A(y, z, w) \quad (\text{Eq. 3})$$

where $A(y, z, w) = \pi_2 + \pi_3 + \pi_4 - (1-\lambda)C_1 - (\pi_2 + s_2)y - (\pi_3 + s_3)z - (\pi_4 + s_4)w$

An equilibrium point of the replicator dynamics system is evolutionarily stable if and only if two conditions are satisfied (Lyapunov stability criterion for differential equations):

(1) The replicator dynamics equation equals zero ($F(x) = 0$)—the strategy probability no longer changes;

(2) The first derivative at the equilibrium point is negative ($F'(x) < 0$)—small deviations from the equilibrium will revert to the original state.

For Equation (2), the equilibrium points occur in three scenarios:

$x = 0$: the regulatory authority choose lax regulation;

$x = 1$: the regulatory authority choose strict regulation;

The term in brackets equals zero: $\pi_2 + \pi_3 + \pi_4 - (1-\lambda)C_1 - y(\pi_2 + s_2) - z(\pi_3 + s_3) - w(\pi_4 + s_4) = 0$.

We define the critical value of w (upstream supplier's truthful disclosure probability) that satisfies the third scenario as w_1 , which represents the “threshold

probability” where the regulatory authority is indifferent between strict and lax

$$\text{regulation: } w_1 = \frac{\pi_2 + \pi_3 + \pi_4 - (1 - \lambda)C_1 - (\pi_2 + s_2)y - (\pi_3 + s_3)z}{\pi_4 + s_4}.$$

Proposition 1. *The government regulator’s evolutionarily stable strategy (ESS) is characterized by a threshold w_1 , i.e., when $w < w_1$, the authority’s stable strategy is strict regulation; when $w > w_1$, the authority’s stable strategy is lax regulation; and when $w = w_1$, the stability of the strategy cannot be determined.*

Proof: Since $\frac{\partial A(y, z, w)}{\partial w} < 0$, the function $A(y, z, w)$ is monotonically decreasing

with respect to w . When $w < w_1$, $A(y, z, w) > 0$, and given that $F(x)|_{x=1} = 0$ and

$F'(x)|_{x=1} < 0$, the equilibrium point $x=1$ is asymptotically stable. When $w > w_1$, $A(y, z,$

$w) < 0$, and given that $F(x)|_{x=1} = 0$ and $F'(x)|_{x=1} < 0$, the equilibrium point $x=0$ is

asymptotically stable. While when $w = w_1$, and given that $F(x)|_{x=1} = 0$ and

$F'(x)|_{x=1} = 0$, it means the stability of the strategy cannot be determined. $\square$

Proposition 1 reveals a critical threshold for government regulatory intervention. If the supplier’s propensity for truthful disclosure (w) falls below a certain level (w_1), indicating high systemic risk of “greenwashing”, the government authority is compelled to adopt a strict regulatory stance to maintain order. Conversely, when the supplier demonstrate a high level of voluntary compliance ($w > w_1$), the government can optimally scale back its direct oversight, allowing market mechanisms to function with lower administrative costs. This finding underscores the dynamic and conditional nature of optimal regulatory policy.

Based on Proposition 1, the phase diagram for the government regulatory’s strategy of strict regulation is illustrated in Figure 2.

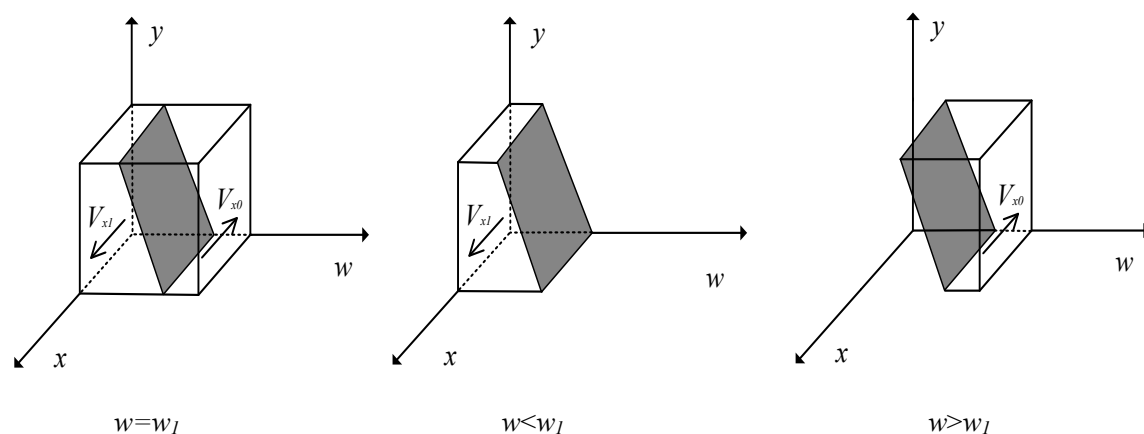


Figure 2. Phase diagram for the government’s strict regulation strategy

As shown in Figure 2, the volume of V_{x1} represents the probability that the governmental carbon regulatory authority adopts a strategy of strict regulation, whereas the volume of V_{x0} represents the probability that it adopts a lax regulation strategy. Let $t_1 = -[\pi_2 + \pi_3 + \pi_4 - (1 - \lambda)C_1 - (\pi_3 + s_3)z]$, then

$w_1 = [t_1 + (\pi_2 + s_2)y] / (\pi_4 + s_4)$. So we obtain:

$$V_{x1} = \int_0^1 \int_0^1 \frac{[t_1 + (\pi_2 + s_2)y]}{(\pi_4 + s_4)} dy dx = \frac{[(\pi_2 + s_2 + \pi_3 + s_3) - 2(\pi_2 + \pi_3 + \pi_4 - 2(1 - \lambda)C_1)]}{2(\pi_4 + s_4)} \quad (\text{Eq. 4})$$

$$V_{x0} = 1 - \frac{[(\pi_2 + s_2 + \pi_3 + s_3) - 2(\pi_2 + \pi_3 + \pi_4 - 2(1 - \lambda)C_1)]}{2(\pi_4 + s_4)} \quad (\text{Eq. 5})$$

Corollary 1.1. *The probability that the governmental carbon regulatory authority adopts strict regulation decreases as the regulatory cost C_1 increases and the reward–penalty intensity (s_2, s_3, s_4) strengthens.*

Proof: Taking the first-order partial derivatives of V_{x1} with respect to π_2, π_3, π_4, s_2

, s_3, s_4 , and C_1 yields: $\frac{\partial V_{x1}}{\partial \pi_2} > 0, \frac{\partial V_{x1}}{\partial \pi_3} > 0, \frac{\partial V_{x1}}{\partial \pi_4} > 0, \frac{\partial V_{x1}}{\partial s_2} < 0, \frac{\partial V_{x1}}{\partial s_3} < 0, \frac{\partial V_{x1}}{\partial s_4} < 0$ and

$\frac{\partial V_{x1}}{\partial C_1} < 0$. It can be observed that V_{x1} is positively correlated with π_2, π_3, π_4 , and

negatively correlated with s_2, s_3, s_4 , and C_1 . $\square$

According to Corollary 1.1, the strategic choice of the governmental carbon regulatory authority is primarily shaped by its regulatory costs and the effectiveness of the reward–penalty scheme it administers. Specifically, when the cost of strict regulation is excessively high, the authority is less inclined to adopt strict oversight, even if such leniency may incur higher ex-post governance costs. Conversely, when the government offers substantial subsidies and rewards to other participants while progressively intensifying penalties, the entire low-carbon supply chain evolves toward outcomes that enhance both ecological and economic performance. Under such circumstances, the authority no longer needs to closely monitor the focal firm’s trust-repair efforts or the disclosure behavior of upstream supplier. Consequently, the probability of choosing strict regulation declines over time and ultimately stabilizes at a low level.

Corollary 1.2. *The governmental carbon regulatory authority prefers strict regulation when its regulatory cost C_1 is below the critical threshold C_0 ; however, once C_1 exceeds C_0 , its strategy shifts to lax regulation.*

Proof: From the expression for $A(y,z,w)$, we have $\frac{\partial A(y,z,w)}{C_1} < 0$. Let the threshold

be defined as $C_0 = \frac{[\pi_2 + \pi_3 + \pi_4 - (\pi_2 + s_2)y - (\pi_3 + s_3)z - (\pi_4 + s_4)w]}{(1 - \lambda)}$. When $C_1 > C_0$,

$A(y,z,w) < 0$ holds, resulting in $F(x)|_{x=1} = 0$ and $F'(x)|_{x=1} < 0$. Thus, the government carbon regulatory department chooses the “lax regulatory” strategy. Conversely, when $C_1 < C_0$, $A(y,z,w) > 0$ holds, resulting in $F(x)|_{x=1} = 0$ and $F'(x)|_{x=1} < 0$, indicating the department selects the “strict regulatory” strategy. $\square$

This corollary establishes regulatory cost C_1 as the pivotal factor in the government’s strategic choice. When the regulatory burden is relatively low, the authority is more inclined to enforce strict oversight to curb opportunistic behaviors and strengthen carbon governance. However, as C_1 increases and exceeds the critical threshold, the marginal cost of strict enforcement becomes disproportionately high. Consequently, the authority rationally transitions to a lax regulatory approach, even though such relaxation may result in higher governance costs for the supply chain in the long run.

4.1.2 Focal firm

The downstream focal firm also faces a binary strategic choice: trust repair and trust breakdown. Let U_{C1} and U_{C2} denote the expected payoffs from trust repair and trust breakdown, respectively, and let U_C be the average expected payoff of the focal firm’s strategy set. The analytical procedure follows the same methodological framework established in above Section 4.1.1 for the government regulator; therefore, we present a condensed derivation below while focusing on the focal firm’s distinctive payoff structure and equilibrium conditions.

$$U_{C1} = xzw b_1 + xz(1-w)b_2 + (1-x)zw b_3 + (1-x)z(1-w)b_4 + x(1-z)wb_5 + x(1-z)(1-w)b_6 + (1-x)(1-z)wb_7 + (1-x)(1-z)(1-w)b_8$$

$$U_{C2} = xzw b_9 + xz(1-w)b_{10} + (1-x)zw b_{11} + (1-x)z(1-w)b_{12} + x(1-z)wb_{13} + x(1-z)(1-w)b_{14} + (1-x)(1-z)wb_{15} + (1-x)(1-z)(1-w)b_{16}$$

$$U_C = yU_{C1} + (1-y)U_{C2} \tag{Eq. 6}$$

$$F(y) = y(U_{C1} - U_C) = y(1-y)(U_{C1} - U_{C2}) = y(1-y)B(x, z, w) \tag{Eq. 7}$$

$$F'(y) = (1-2y)B(x, z, w) \tag{Eq. 8}$$

where $B(x, z, w) = \alpha R_2 - \eta_2 - C_2 + (\pi_2 + s_2)x + ew + r_2z$

According to the stability theorem for differential equations, the focal firm's decision reaches a stable state requires the following two conditions are satisfied simultaneously: $F(y) = 0$ and $F'(y) < 0$. Setting $B(x, z, w) = 0$ yields the critical threshold for the supplier's truthful-disclosure probability: $w_2 = \frac{[\eta_2 + C_2 - \alpha R_2 - (\pi_2 + s_2)x - r_2 z]}{e}$.

Proposition 2. *The focal firm's ESS is determined by the threshold w_2 : when $w < w_2$ the ESS is trust breakdown; when $w > w_2$ the ESS is trust repair; when $w = w_2$ stability is indeterminate.*

Proof. The proof is essentially the same as that of Proposition 1.

Proposition 2 highlights that the firm's shift toward Trust Repair critically hinges on supplier transparency (w). When the supplier is likely to misreport ($w < w_2$), the benefits of trust repair are eroded by information uncertainty, making trust breakdown the rational short-run choice. Conversely, a credible disclosure environment ($w > w_2$) assures the firm that investments in green procurement and verification will generate tangible returns, incentivizing the firm to adopt and sustain a proactive trust-repair strategy. This result underscores that the focal firm's transition toward sustainable procurement is not merely an internal decision, but is critically dependent on the informational ecosystem of its supply base.

Based on Proposition 2, the phase diagram for the focal firm's strategy of trust repair is illustrated in Figure 3.

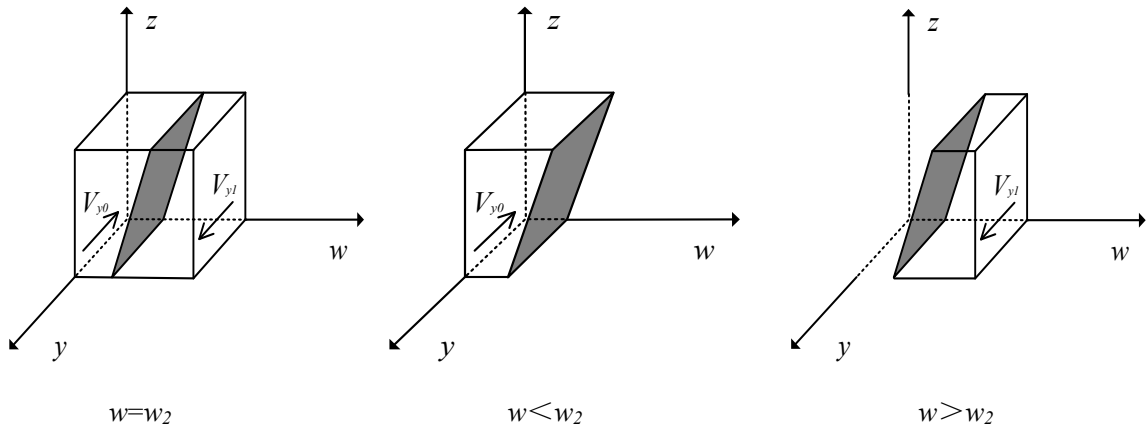


Figure 3. Phase diagram of the focal firm's trust repair strategy

As shown in Figure 3, the volume of V_{y1} represents the probability that the focal firm adopts a strategy of trust repair, whereas the volume of V_{y0} represents the probability that it adopts a trust breakdown strategy. So we obtain:

$$V_{y0} = \int_0^1 \int_0^1 w_2 dz dy = \frac{[2(C_{43} + C_{44} + \pi_4 + s_4) - (\pi_2 + s_2 + r_2) - (\pi_3 + s_3)]}{4(\pi_4 + s_4)} \quad (\text{Eq. 9})$$

$$V_{y1} = 1 - V_{y0} = 1 - \frac{[2(C_{43} + C_{44} + \pi_4 + s_4) - (\pi_2 + s_2 + r_2) - (\pi_3 + s_3)]}{4(\pi_4 + s_4)} \quad (\text{Eq. 10})$$

Corollary 2.1. *The probability that the focal firm chooses trust repair strategy decreases when the cost of supplier truthful disclosure under its own trust-repair (C_{43})*

or under trust-breakdown (C_{44}) increases, and increases with the penalty for trust breakdown (π_2) and government rewards for trust repair (s_2), or the green premiums from third-party certification (r_2) rises.

Proof. The proof is essentially the same as that of Corollary 1.1.

Corollary 2.1 reflects a strategic trade-off for the focal firm between trust repair and trust breakdown is governed by a trade-off between the expected returns from trust repair and the compliance costs of supplier transparency. Higher supplier disclosure costs (C_{43} or C_{44}) erode the economic incentive for trust repair, pushing the firm toward trust breakdown, even at the risk of future penalties. Conversely, stronger penalties for non-compliance (π_2), larger government rewards (s_2), and greater green premiums from third-party certification (r_2) amplify the financial and reputational returns of trust repair, motivating the firm to adopt proactive low-carbon procurement and reduce reliance on external regulatory pressure.

Corollary 2.2. *The focal firm chooses trust repair only when its market-share amplification coefficient α exceeds a critical threshold α^* , defined as*

$\alpha^* = [\eta_2 + C_2 - (\pi_2 + s_2)x - ew - r_2z] / R_2$. *This threshold increases when the risk gain from trust breakdown (η_2) or the cost of trust repair (C_2) rises, and decreases when the penalty for trust breakdown (π_2), governmental rewards for trust repair (s_2), environmental benefits of truthful disclosure (e), or green premiums from third-party certification (r_2) strengthen.*

Proof. The proof is essentially the same as that of Corollary 1.2.

Corollary 2.2 identifies the market-share effect α as the pivotal driver of the focal firm's trust-repair decision. Trust repair becomes economically viable only when α exceeds the threshold α^* , in which case the additional revenue from an expanded green market is sufficient to cover the costs and risks of trust repair, motivating the firm to invest in verification and transparent disclosure. Conversely, when $\alpha < \alpha^*$, the firm rationally reverts to trust breakdown, prioritizing short-term cost savings over long-term carbon credibility. In addition, the threshold α^* is not fixed: it increases with higher trust-repair costs (C_2) or greater short-term gains from inaction (η_2), and decreases with stronger government incentives (s_2, π_2), environmental benefits (e), or green premiums from third-party certification (r_2). Policy and market measures that visibly raise α or lower α^* —such as enhanced green labelling, consumer-driven demand, or targeted subsidies—can thus jointly steer the firm toward sustained trust-based procurement and long-term carbon trust repair. Therefore, trust repair serves as the pivotal mechanism that transforms coercive oversight into a market-driven, reputation-supported equilibrium.

4.1.3 Third-party certifier

The expected payoffs of the third-party certifier when choosing to participate in or not to participate in carbon-footprint verification are denoted by U_{T1} and U_{T2} , respectively, and the mean expected payoff is U_T . The corresponding replicator dynamic and its first derivative are given in Eqs. (12) and (13).

$$\begin{aligned}
U_{T1} &= xywc_1 + xy(1-w)c_2 + (1-x)ync_3 + (1-x)y(1-w)c_4 + x(1-y)wc_9 \\
&\quad + x(1-y)(1-w)c_{10} + (1-x)(1-y)wc_{11} + (1-x)(1-y)(1-w)c_{12} \\
U_{T2} &= xywc_5 + xy(1-w)c_6 + (1-x)ync_7 + (1-x)y(1-w)c_8 + x(1-y)wc_{13} \\
&\quad + x(1-y)(1-w)c_{14} + (1-x)(1-y)wc_{15} + (1-x)(1-y)(1-w)c_{16} \\
U_T &= zU_{T1} + (1-z)U_{T2} \tag{Eq.11}
\end{aligned}$$

$$F(z) = z(U_{T1} - U_T) = z(1-z)(U_{T1} - U_{T2}) = z(1-z)C(x, y, w) \tag{Eq.12}$$

$$F'(z) = (1-2z)C(x, y, w) \tag{Eq.13}$$

$$\text{Where } C(x, y, w) = (\pi_3 + s_3)x + r_3y - C_3$$

Similarly, according to the stability theorem for differential equations, the third-party certifier's decision reaches a stable state requires the following two conditions are satisfied simultaneously: $F(z) = 0$ and $F'(z) < 0$.

Proposition 3. *The third-party certifier's ESS is characterized by two critical thresholds x_3 and y_3 . Specifically, when $x < x_3$ and $y < y_3$, the agency's stable strategy is not to participate in carbon footprint verification; when $x > x_3$ and $y > y_3$, its stable strategy is to participate in carbon footprint verification; and when $x = x_3$ and $y = y_3$, the stability of the strategy cannot be determined. The thresholds are given by*

$$x_3 = \frac{[C_3 - r_3y]}{(\pi_3 + s_3)}, \quad y_3 = \frac{[C_3 - (\pi_3 + s_3)x]}{r_3}.$$

Proof. Since $\partial C(x, y, w) / \partial x > 0$, the function $C(x, y, w)$ (which reflects the net payoff difference of the third-party certifier between “participating” and “not participating”) is a monotonically increasing function of x . When $x < x_3$, $C(x, y, w) < 0$, it satisfies

$$F(z)|_{z=0} = 0 \quad \text{and} \quad F'(z)|_{z=0} < 0, \text{ thus } z=0 \text{ (not participating) is evolutionarily stable.}$$

When $x > x_3$, $C(x, y, w) > 0$, it satisfies $F(z)|_{z=1} = 0$ and $F'(z)|_{z=1} < 0$, so the

equilibrium point $z = 1$ is stable. When $x = x_3$, $F(z) = 0$ and $F'(z) = 0$, the stability of the

strategy cannot be determined. Since $\partial C(x, y, w) / \partial y > 0$, $\partial C(x, y, w) / \partial w > 0$, the effect

of the critical threshold y_3 on strategy stability of the third-party certifier can be proved analogously. $\square$

Proposition 3 indicates that the third-party certifier's participation decision hinges jointly on the strength of governmental regulation and the focal firm's trust-repair behaviour. When the probability of strict regulation x and the probability

of trust repair y are both below their respective thresholds (x_3, y_3) , external regulatory pressure and market-based demand for verification are both weak, so the agency lacks sufficient economic and reputational incentives and rationally chooses not to participate in carbon footprint verification. Once the regulator's enforcement intensity exceeds x_3 and the focal firm's trust-repair probability surpasses y_3 , the combined effect of government rewards (s_3) , penalties (π_3) and certification fees (r_3) makes participation more profitable than non-participation. In this regime, the agency actively engages in carbon footprint certification and provides credible verification reports, thereby enhancing its own market reputation and reinforcing the supply chain's overall carbon trust.

Based on Proposition 3, the phase diagram for the third-party certifier's strategy of participation decision in carbon-footprint verification is illustrated in Figure 4.

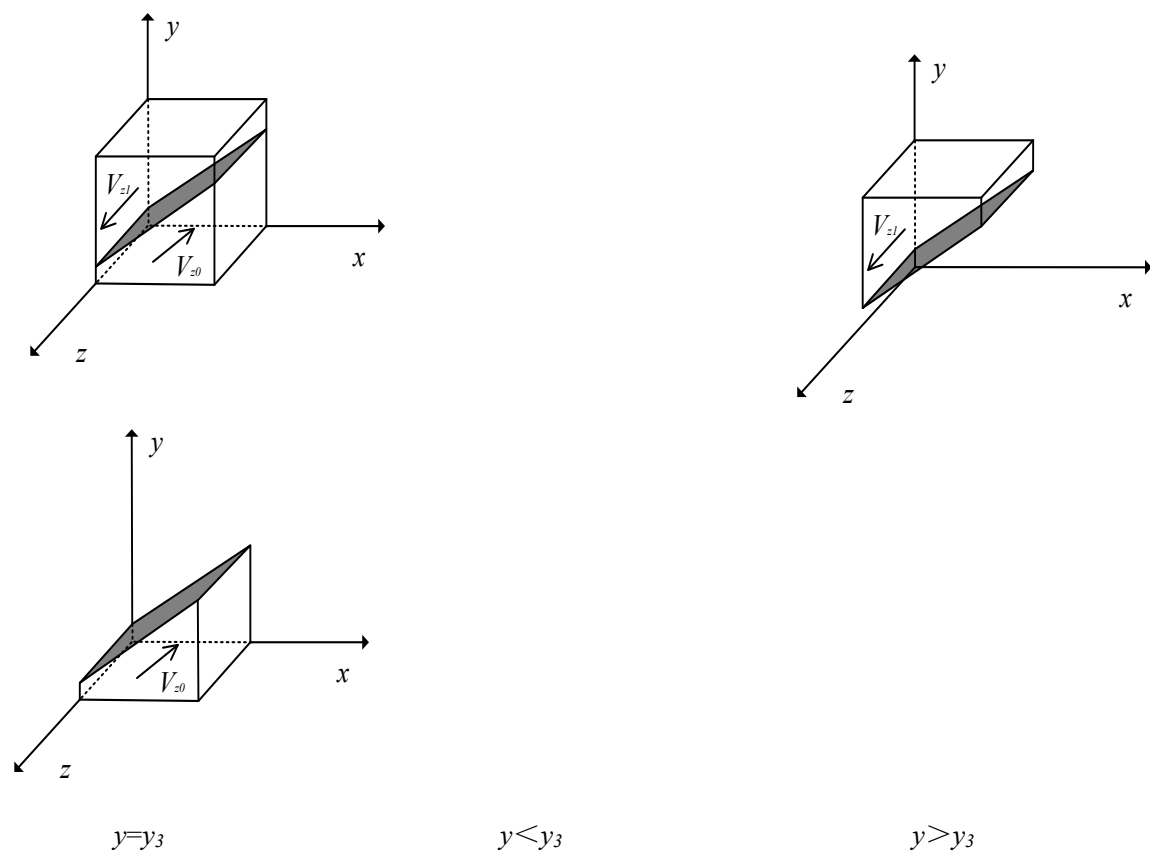


Figure 4. Phase diagram of the third-party certifier's carbon-footprint verification strategy

The volumes of the V_{z0} and V_{zl} segments represent the probabilities that the third-party certifier chooses “non-participation” and “participation” in carbon footprint verification detection, respectively. So we obtain:

$$V_{zl} = \int_0^1 \int_0^1 y_3 dx dz = [1 + (s_3 - C_3) / 2(r_3 + s_3)] / 2 \quad (\text{Eq.14})$$

$$V_{z0} = 1 - V_{zl} = 1 - [1 + (s_3 - C_3) / 2(r_3 + s_3)] / 2 \quad (\text{Eq.15})$$

Corollary 3.1. *The probability that third-party certification agencies evolve toward “participation” ($z=1$) increases with the regulator’s strict-supervision probability x , the focal firm’s trust-repair probability y , the government reward s_3 and the verification fee r_3 paid by the focal firm, and decreases with the agency’s certification cost C_3 . In particular, a decline in y must be offset by a higher x to sustain third-party participation.*

Proof. The proof is essentially the same as that of Corollary 1.1.

According to Corollary 3.1, the participation decision of third-party certification agencies is jointly driven by regulatory pressure and market-based demand. Stricter government supervision (for example, higher penalties π_3 or larger rewards s_3) raises the risk and opportunity cost of staying out of the verification market, while the core firm’s trust-repair behaviour (such as paying a higher verification fee r_3) directly improves the economic return from participation. Together, these factors expand the effective market for certification services and encourage agencies to provide credible carbon-footprint verification. Conversely, when the core firm’s willingness to repair trust weakens and the associated demand for certification falls, the regulator must compensate for the weaker market incentives by increasing the strict-supervision probability x ; otherwise, third-party certifier may withdraw from participation, undermining the reliability of carbon information along the supply chain. This highlights that, especially in the early stages of carbon trust governance, regulatory oversight and market demand should function as complementary drivers of third-party engagement.

Corollary 3.2. *There exists a threshold S_0 which is defined as $S_0 = s_3 + \pi_3 = \frac{[C_3 - r_3 y]}{x}$*

for the regulator’s total reward-and-penalty level S such that third-party agencies participate in carbon-footprint verification if and only if $S > S_0$. The threshold S_0 increases with the agencies’ effective participation cost and with the opportunity gains from non-participation, which rise as the contributions of the other three players to the low-carbon supply chain diminish, thereby requiring a higher S from the regulator to induce participation.

Proof. The proof is essentially the same as that of Corollary 1.2.

According to Corollary 3.2, third-party participation in carbon-footprint verification is not unconditional; it requires the regulator’s incentive intensity to exceed a critical level. When other supply chain actors (such as the focal firm and the supplier) underinvest in low-carbon efforts, certification agencies face higher effective verification costs (for example, greater difficulty in collecting reliable information) and lower cooperation returns (for example, weaker demand for certification). At the same time, the potential gains from non-participation, such as saving costs or avoiding reputational exposure, become more attractive. In this situation, the regulator must compensate for missing market incentives by strengthening the reward-and-penalty scheme, for instance by increasing subsidies s_3 or fines π_3 ; otherwise, the supply of certification services will shrink and carbon trust will struggle to emerge along the supply chain. This implies that, in the early stages of

low-carbon transition or when market mechanisms malfunction, the government needs to act as an “incentive backstop”, using robust policy tools to steer the development of third-party certification systems.

4.1.4 Upstream supplier

The expected payoffs for the upstream supplier choosing truthful carbon-footprint disclosure and false disclosure are U_{S1} and U_{S2} , respectively, and the average expected payoff is U_S . The replicator dynamics equation for their behavioral strategy and its first derivative are given by Equations (17) and (18), respectively.

$$U_{S1} = xyzd_1 + (1-x)yzd_3 + xy(1-z)d_5 + (1-x)y(1-z)d_7 + x(1-y)zd_9 \\ + (1-x)(1-y)zd_{11} + x(1-y)(1-z)d_{13} + (1-x)(1-y)(1-z)d_{15}$$

$$U_{S2} = xyzd_2 + (1-x)yzd_4 + xy(1-z)d_6 + (1-x)y(1-z)d_8 + x(1-y)zd_{10} \\ + (1-x)(1-y)zd_{12} + x(1-y)(1-z)d_{14} + (1-x)(1-y)(1-z)d_{16}$$

$$U_S = wU_{S1} + (1-w)U_{S2} \quad (\text{Eq.16})$$

$$F(w) = w(U_{S1} - U_S) = w(1-w)(U_{S1} - U_{S2}) = w(1-w)D(x, y, z) \quad (\text{Eq.17})$$

$$F'(w) = (1-2w)D(x, y, z) \quad (\text{Eq.18})$$

where $D(x, y, z) = [(s_4 + \pi_4)x + (C_{44} - C_{43})y + (C_{41} - C_{42})yz + C_{42}z + \beta R_4 - C_{44}]$.

It is evident that the key determinants influencing the upstream supplier's choice of truthful disclosure strategy include the market reputation loss coefficient resulting from false disclosure, the rewards and penalties imposed by government regulatory authorities, and the costs associated with truthful disclosure for the upstream supplier. Similarly, according to the stability theorem for differential equations, the upstream supplier's decision reaches a stable state requires the following two conditions are satisfied simultaneously: $F(y) = 0$ and $F'(y) < 0$.

Proposition 4. *The upstream supplier's ESS is characterized by a critical thresholds x_4 . Specifically, when $x > x_4$, the supplier's stable strategy is to truthfully disclose their carbon footprint; when $x < x_4$, its stable strategy is to falsely disclose; and when $x = x_4$, the stability of the strategy cannot be determined. The thresholds are given by*

$$x_4 = \frac{[C_{44} - (C_{44} - C_{43})y - (C_{41} - C_{42})yz - C_{42}z - \beta R_4]}{(s_4 + \pi_4)}.$$

Proof. The proof is essentially the same as that of Proposition 1.

Proposition 4 shows that when the probability of strict government regulation is low, the focal firm's willingness to repair trust is weak, and the participation rate of third-party certifier is limited, the probability that the upstream supplier will be detected for false reporting declines while the short-term gains from non-compliance rise, so the supplier tends to choose a false-disclosure strategy. Conversely, when the regulator's enforcement intensity exceeds the threshold x_4 and both the focal firm and

the third-party certifier raise their probabilities of trust repair and certification participation, the upstream supplier will face higher expected fines, stronger incentives from government rewards, and greater pressure to protect their market reputation. In this case, they are more likely to truthfully disclose their carbon footprint, thereby driving the evolution of supply chain carbon trust from a “false equilibrium” toward a “true equilibrium”.

Based on Proposition 4, the phase diagram for the upstream supplier’s strategy of truthfully disclose their carbon footprint is illustrated in Figure 5.

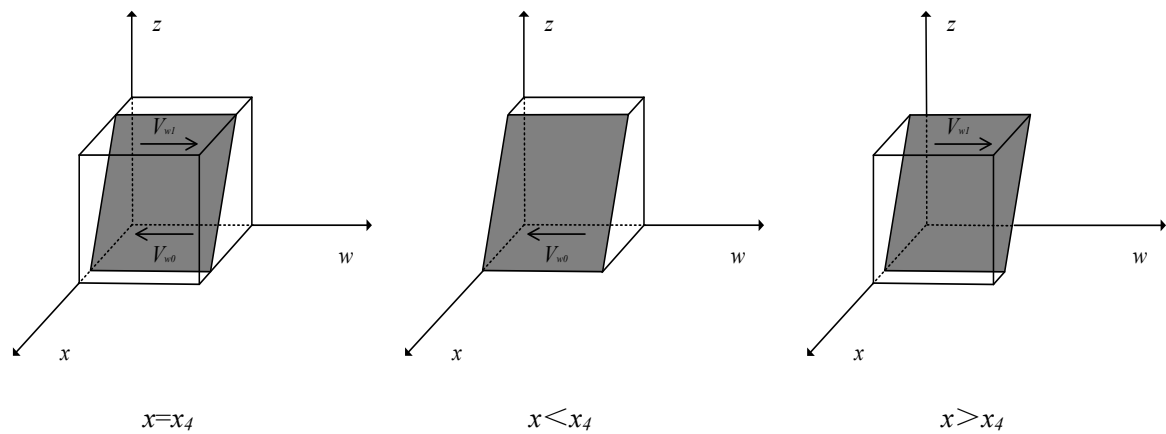


Figure 5. Phase diagram of the upstream supplier’s carbon-footprint disclosure strategy

Corollary 4.1. *The upstream supplier’s strategic choice is primarily shaped by its own cost-benefit trade-offs and external reward-penalty mechanisms. The probability of the supplier choosing truthful carbon footprint disclosure increases with higher benefits from truthful disclosure, higher costs of false disclosure, and stronger penalties for false disclosure imposed by the other three players. Conversely, this probability decreases as the cost of truthful disclosure or the benefits from false disclosure rise.*

Proof. The proof is essentially the same as that of Corollary 1.1.

Corollary 4.1 shows that the supplier’s carbon disclosure decisions essentially reflect a fundamental trade-off between the benefits of compliance and the opportunity costs. Higher returns to truthful disclosure, greater expected losses from false disclosure, and stronger joint sanctions from the regulator, the focal firm and the third-party certifier form a reinforcing incentive network that encourages the supplier toward honest behaviour. In this sense, effective low-carbon governance depends less on supplier’s environmental awareness per se than on whether institutional design can make truthful disclosure economically more attractive than opportunism. Conversely, when the operational and verification costs of truthful disclosure are high or the short-term gains from misreporting are substantial, the supplier may still lean toward opportunistic strategies even under formal regulation. This implies that policymakers should jointly optimise cost-sharing arrangements and penalty schemes within the supply chain, thereby reducing the operational burden of truthful disclosure and

squeezing the profit space of false disclosure, so as to guide supplier toward stable low-carbon disclosure norms.

Corollary 4.2. *There exists a threshold s'_4 for the government's reward s_4 such that the supplier chooses truthful carbon-footprint disclosure only if $s_4 > s'_4$. The threshold s'_4 rises with the cost of truthful disclosure and with weaker contributions from the focal firm, the third-party certifier, and market-based reputation sanctions; in these cases, the regulator must increase the reward level s_4 to sustain truthful disclosure as an evolutionarily stable strategy.*

Proof. The proof is essentially the same as that of Corollary 1.2.

Corollary 4.2 indicates that the regulator's reward scheme exerts a critical but threshold-dependent influence on the supplier's willingness to disclose truthfully: only when the reward level exceeds a psychological threshold can it offset the additional costs and perceived risks of truthful reporting and steer the system toward a "truthful-disclosure" equilibrium. Importantly, this threshold is not fixed but adjusts with changing conditions in supply chain governance. As the cost of truthful disclosure rises, the focal firm's trust-repair effort and the third-party agency's participation weaken, or the reputational loss from false reporting diminishes, supplier's reliance on and expectations about government rewards increase accordingly. In such cases, the carbon regulator needs the capacity to dynamically adjust the reward intensity s_4 ; if the reward scheme is too rigid and fails to track these shifts, the incentive mechanism may lose effectiveness and lead to either persistent false disclosure or inefficient allocation of public funds. This finding underscores that government incentive policies in low-carbon supply chain governance must remain flexible and adaptive rather than static "one-shot" designs.

4.2 Stability Analysis of System Strategy Combinations and Equilibrium Paths

To reveal the ESS formation conditions and the corresponding evolutionary process for the supply chain carbon trust repair problem induced by information asymmetry, this section models and analyzes the dynamics of the four-player evolutionary game. The game involves the government regulatory agency, the focal firm, the third-party certifier, and the upstream supplier, and is based on the replicator dynamic system defined in Section 3.3. The asymptotic stability of the system is determined by analyzing the stability of the Replicator Dynamic Equations (RDEs). We first obtain the potential steady states by setting the growth rate of each strategy to zero ($F(x)=0$, $F(y)=0$, $F(z)=0$, $F(w)=0$). According to the stability theorem for multi-group evolutionary games, a stable solution must be an ESS, which is a refinement of the Strict Nash Equilibrium. We therefore apply Lyapunov's First Method to analyze the stability of the 16 pure-strategy equilibrium points, with the results summarized in Table 4. Specifically, the stability of any equilibrium point $E^*(x^*, y^*, z^*, w^*)$ is governed by the signs of the eigenvalues of the corresponding Jacobian matrix J .

The Jacobian matrix J is given by:

$$J = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} & \frac{\partial F(x)}{\partial z} & \frac{\partial F(x)}{\partial w} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} & \frac{\partial F(y)}{\partial z} & \frac{\partial F(y)}{\partial w} \\ \frac{\partial F(z)}{\partial x} & \frac{\partial F(z)}{\partial y} & \frac{\partial F(z)}{\partial z} & \frac{\partial F(z)}{\partial w} \\ \frac{\partial F(w)}{\partial x} & \frac{\partial F(w)}{\partial y} & \frac{\partial F(w)}{\partial z} & \frac{\partial F(w)}{\partial w} \end{bmatrix}$$

Table 4. Pure Strategy Equilibria of the Four-Player Game

Equilibrium Point	Eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$	Sign Notation	Stability and Conditions
$E_1(0, 0, 0, 0)$	$\Delta_1 + (\lambda - I)C_1, \alpha R_2 - \Delta_2, -C_3, \beta R_4 - C_{44}$	$\times \times - \times$	ESS(Condition 1)
$E_2(0, 0, 0, 1)$	$\pi_2 + \pi_3 - s_4 + (\lambda - I)C_1, \alpha R_2 + e - \Delta_2, -C_3, C_{44} - \beta R_4$	$\times \times - \times$	ESS(Condition 2)
$E_3(0, 0, 1, 0)$	$\pi_2 + \pi_4 - s_3 + (\lambda - I)C_1, \alpha R_2 + r_2 - \Delta_2, C_3, \beta R_4 + C_{42} - C_{44}$	$\times \times + \times$	Instable
$E_4(0, 0, 1, 1)$	$\pi_2 - s_3 - s_4 + (\lambda - I)C_1, e + \alpha R_2 + r_2 - \Delta_2, C_3, C_{44} - C_{42} - \beta R_4$	$\times \times + \times$	Instable
$E_5(0, 1, 0, 0)$	$\pi_3 + \pi_4 - s_2 + (\lambda - I)C_1, \Delta_2 - \alpha R_2, r_3 - C_3, \beta R_4 - C_{43}$	$\times \times \times \times$	/
$E_6(1, 0, 0, 0)$	$(I - \lambda)C_1 - \Delta_1, \alpha R_2 + \pi_2 + s_2 - \Delta_2, \Delta_3 - C_3, \beta R_4 + \Delta_4 - C_{44}$	$\times \times \times \times$	/
$E_7(1, 0, 0, 1)$	$(I - \lambda)C_1 - \pi_2 - \pi_3 + s_4, \alpha R_2 + e + \pi_2 + s_2 - \Delta_2, \Delta_3 - C_3, C_{44} - \beta R_4 - \Delta_4$	$\times \times \times \times$	/
$E_8(0, 1, 0, 1)$	$\pi_3 - s_2 - s_4 + (\lambda - I)C_1, \Delta_2 - \alpha R_2 - e, r_3 - C_3, C_{43} - \beta R_4$	$\times \times \times \times$	/
$E_9(0, 1, 1, 0)$	$\pi_4 - s_2 - s_3 + (\lambda - I)C_1, \Delta_2 - \alpha R_2 - r_2, C_3 - r_3, \beta R_4 + C_{41} - C_{43}$	$\times \times \times \times$	/
$E_{10}(1, 0, 1, 0)$	$s_3 - \pi_2 - \pi_4 + (I - \lambda)C_1, \alpha R_2 + \pi_2 + s_2 - \Delta_2 + r_2, C_3 - \Delta_3, \beta R_4 + \Delta_4 + C_{42} - C_{44}$	$\times \times \times \times$	/
$E_{11}(1, 1, 1, 0)$	$(I - \lambda)C_1 + s_2 + s_3 - \pi_4, \Delta_2 - \alpha R_2 - \pi_2 - s_2 - r_2, -\Delta_3 - r_3, C_{41} + \beta R_4 + \Delta_4 - C_{43}$	$\times \times - \times$	ESS(Condition 3)
$E_{12}(1, 0, 1, 1)$	$(I - \lambda)C_1 - \pi_2 + s_3 + s_4, \alpha R_2 + \pi_2 + s_2 - \Delta_2 + r_2 + e, C_3 - \Delta_3, C_{44} - \beta R_4 - \Delta_4 - C_{42}$	$\times \times \times \times$	/
$E_{13}(1, 1, 0, 1)$	$(I - \lambda)C_1 + s_2 + s_4 - \pi_3, \Delta_2 - \alpha R_2 - e - \pi_2 - s_2, \Delta_3 + r_3 - C_3, C_{43} - \beta R_4 - \Delta_4$	$\times \times \times \times$	/
$E_{14}(1, 1, 0, 0)$	$(I - \lambda)C_1 + s_2 - \pi_3 - \pi_4, \Delta_2 - \alpha R_2 - \pi_2 - s_2, \Delta_3 + r_3, \beta R_4 + \Delta_4 - C_{43}$	$\times \times + \times$	Instable
$E_{15}(0, 1, 1, 1)$	$(\lambda - I)C_1 - s_2 - s_3 - s_4, \Delta_2 - \alpha R_2 - r_2 - e, C_3 - r_3, C_{43} - \beta R_4 - C_{41}$	$- \times \times \times$	ESS(Condition 4)
$E_{16}(1, 1, 1, 1)$	$(I - \lambda)C_1 + s_2 + s_3 + s_4, \Delta_2 - \alpha R_2 - \pi_2 - s_2 - r_2 - e, C_3 - \Delta_3 - r_3, C_{43} - C_{41} - \beta R_4 - \Delta_4$	$+ \times \times \times$	Instable

Note: $\times$ represents positive or negative uncertainty; $\Delta_1 = \pi_2 + \pi_3 + \pi_4$, $\Delta_2 = C_2 + \eta_2$, $\Delta_3 = \pi_3 + s_3$, $\Delta_4 = \pi_4 + s_4$.

Condition 1: $\Delta_1 < (I - \lambda)C_1$, $\alpha R_2 < \Delta_2$, $\beta R_4 < C_{44}$;

Condition 2: $\pi_2 + \pi_3 < s_4 + (I - \lambda)C_1$, $\alpha R_2 + e < \Delta_2$, $C_{44} < \beta R_4$;

Condition 3: $(I - \lambda)C_1 + s_2 + s_3 < \pi_4$, $\Delta_2 < \alpha R_2 + \pi_2 + s_2 + r_2$, $C_{41} + \beta R_4 + \Delta_4 < C_{43}$;

Condition 4: $\Delta_2 < \alpha R_2 + r_2 + e$, $C_3 < r_3$, $C_{43} < \beta R_4 + C_{41}$.

Based on the sign analysis of the eigenvalues of the Jacobian matrix J , four pure-strategy equilibrium points are identified as potential ESS. These four points represent distinct governance paradigms within the supply chain.

$E_1(0, 0, 0, 0)$: The Carbon Trust Deficit State

The stability condition is $\Delta_1 < (I - \lambda)C_1$, $\alpha R_2 < \Delta_2$, $\beta R_4 < C_{44}$. In this scenario, the government's total revenue from strict regulation is lower than the cost savings from lax regulation, the focal firm's trust repair benefits are less than the risk benefits, and the supplier's reputation loss from false disclosure is lower than the true disclosure cost. This equilibrium represents a "full non-compliance" state, where the high regulatory costs, lack of incentives for the focal firm, and low cost of supplier

concealment lead the system trapped in a vicious cycle of “bad money driving out good”, resembling a race to the bottom in trust.

E₂ (0, 0, 0, 1): Supplier-Driven Compliance Equilibrium

The stability condition is $\pi_2 + \pi_3 < s_4 + (1-\lambda)C_1$, $\alpha R_2 + e < \Delta_2$, $C_{44} < \beta R_4$. Here, the government’s penalty revenue is lower than subsidy expenditure, the focal firm’s trust repair benefits are insufficient to cover the repair costs, and the supplier’s true disclosure cost is lower than the reputation loss. This equilibrium relies heavily on the supplier’s unilateral effort and the market’s high sensitivity to reputation, making the stability fragile and highly susceptible to changes in market dynamics.

E₁₁ (1, 1, 1, 0): Coercion-Constrained Trust Equilibrium

The stability condition is $(1-\lambda)C_1 + s_2 + s_3 < \pi_4$, $\Delta_2 < \alpha R_2 + \pi_2 + s_2 + r_2$, $C_{41} + \beta R_4 + \Delta_4 < C_{43}$. In this scenario, the carbon regulatory agency, focal firm, and third-party certifier have formed a governance alliance, but the supplier opt for false disclosure due to high costs of true disclosure (e.g., production line overhaul). This state reveals the technical bottlenecks in carbon footprint tracking, which constrain the system’s evolution. When $C_{43} > C_{41} + \beta R_4 + \Delta_4$, a rational supplier choose false disclosure, and the coordination of the three players alone cannot solve the fundamental issue of the high carbon footprint disclosure costs for the upstream supplier, thus driving their “greenwashing” behavior.

E₁₅ (0, 1, 1, 1): Autonomous Carbon Trust Equilibrium

The stability condition is $\Delta_2 < \alpha R_2 + r_2 + e$, $C_3 < r_3$, $C_{43} < \beta R_4 + C_{41}$. This equilibrium represents the ideal state, where the focal firm transitions its profit model through green premiums and environmental benefits, the third-party becomes sustainable due to economies of scale ($C_3 < r_3$), and the supplier voluntarily comply due to the enhanced reputational value β . In this state, the government carbon regulatory agency assumes a supportive role, forming a positive cycle of “De-emphasizing coercive regulation to empower market mechanisms”. However, this does not imply the absence of regulation, but rather a shift from a “front-line regulator” to a “back-end rule-maker and market enabler”. By establishing standards, certification systems, and information disclosure rules, the government lays the groundwork for market mechanisms to play their role.

In the ideal state, the driving force of governance shifts from external coercion to endogenous incentives. A powerful signal transmission and reputation feedback loop is formed between the four key players. The focal firm sends a strong signal of “green leadership” to the market through trust repair actions, attracting green-preference consumers, thus gaining green premiums and market share. At the same time, it requires the supplier to provide third-party certified carbon data as a cooperation threshold. The supplier must disclose truthfully to secure orders, and their good carbon performance further strengthens the focal firm’s green reputation, enhancing trust across the supply chain. Once this loop is established, it has self-reinforcing characteristics, achieving a market-enterprise “autonomous” trust repair. Since this equilibrium no longer heavily relies on government fiscal rewards and administrative penalties, it is more resilient to external policy fluctuations. For example, fluctuations

in carbon price P_C affect r_2 , but the focal firm can absorb this fluctuation through pricing, innovation, and other means, without easily altering its trust repair strategy.

5. Numerical Experiments

5.1 Parameter setting and benchmark trajectory

To illustrate how key parameters shape evolutionary trajectories and equilibrium convergence in the four-player game, we conduct numerical simulations of the replicator dynamic system using MATLAB R2024b with a focus on the punishment–subsidy intensity and reputation related returns. While the parameter calibration in multi-agent evolutionary games is inherently challenging. Following common parameterization practices in related studies (e.g., Sun, Gao & Li, 2022; Li et al., 2024; Xu et al., 2024; Fan, Fang & Jiang, 2024), we adopt baseline values by referencing representative settings in the literature and then set the remaining parameters to satisfy model feasibility and the analytical stability constraints derived in Section 4, thereby enhancing the scientific rigor and generality of our findings. The specific parameter values are set as follows: $R_1=100$, $C_1=60$, $\lambda=0.3$, $r_1=15$, $c_1=10$, $R_2=80$, $C_2=45$, $s_2=8$, $e=5$, $P_C=65$, $r_2=62$, $\eta_2=12$, $\pi_2=35$, $\alpha=0.6$, $R_3=30$, $C_3=18$, $s_3=10$, $r_3=15$, $\pi_3=30$, $R_4=50$, $C_{41}=12$, $C_{42}=8$, $C_{43}=15$, $C_{44}=10$, $s_4=7$, $\pi_4=25$, $\beta=0.2$.

In the benchmark run, the initial strategy frequencies are set to $(x, y, z, w) = (0.5, 0.5, 0.5, 0.5)$, reflecting an uninformative prior over each player’s strategy adoption. Figure 6 shows that the focal firm’s trust repair probability y , the third-party participation probability z , and the supplier’s truthful disclosure probability w rapidly converge to 1 and remain stable, whereas the government’s strict regulation probability x converges to 0. This trajectory is consistent with the stability conditions of $E_{15}(0,1,1,1)$ reported in Table 4.

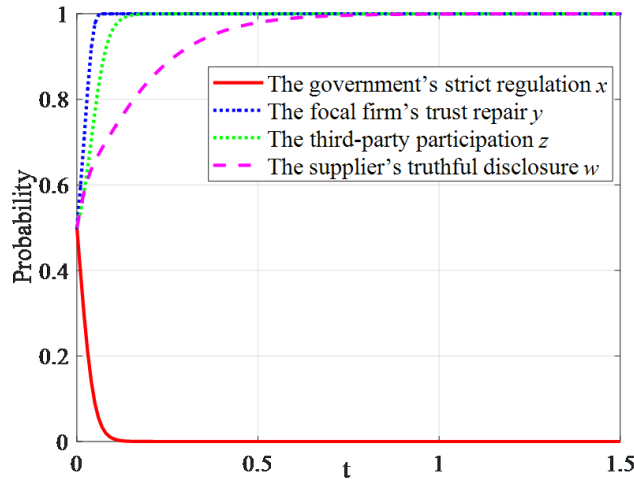


Figure 6. Evolutionary Paths of the Four players' Initial Strategies under Carbon Footprint Information Asymmetry

The results show that, under the baseline parameterization, the system converges endogenously to the ideal equilibrium E_{15} without reliance on strict governmental regulation in the long term. Accordingly, the Government carbon regulatory agency

transitions from a “front-line enforcer” to a “back-end guarantor”. This new role focuses primarily on establishing carbon-footprint accounting standards, ensuring data integrity and security, and maintaining a credible minimum deterrence baseline, thereby enabling market mechanisms to operate autonomously in a transparent and trustworthy environment. This simulation supports the feasibility of the “De-emphasizing coercive regulation to empower market mechanisms” virtuous cycle articulated in Corollary 1.1 and Proposition 2. Further sensitivity analysis confirms that the initial strategy probabilities only affect the speed of convergence to the stable state, without altering the final equilibrium. This provides additional evidence for the global stability of E_{15} .

5.2 Key parameter sensitivity analysis

5.2.1 Sensitivity analysis of focal firm’s strategy parameters

By adjusting the focal firm’s green premium factor k and the market amplification coefficient α associated with trust repair, we examine the focal firm’s strategic evolution more intuitively. We set the multiplier k in $r_2=20+2kP_C$ takes the values of 0.2, 0.4, 0.6 and 0.8 respectively to simulate how carbon price fluctuations affect the focal firm’s willingness to undertake trust repair.

As shown in Figure 7, when the value of k is small, the focal firm’s propensity to engage in trust repair declines markedly and the convergence slows. However, as k increases, r_2 rises accordingly, the evolutionary speed of trust repair accelerates significantly, ultimately stabilizing at $y=1$. These results indicate that the green premium accruing to the focal firm, enabled by the third-party certification agency’s active participation in carbon-footprint verification, constitutes a critical market driver for the focal firm to adopt the trust repair strategy.

Figure 8 further illustrates the effect of the market amplification coefficient α on the focal firm’s strategy. As α increases, the probability that the focal firm chooses trust repair gradually rises, indicating that economies of scale induced by market-share expansion can effectively offset the costs of trust repair, thereby forming a virtuous cycle of “reputation investment–order growth–market expansion”. Overall, the simulation results in Figures. 7–8 underscore the central role of green market mechanisms in supply chain trust repair: by engaging in trust repair, the focal firm can enhance its reputation and market standing, thereby facilitating a market-based and incentive-compatible autonomous governance regime.

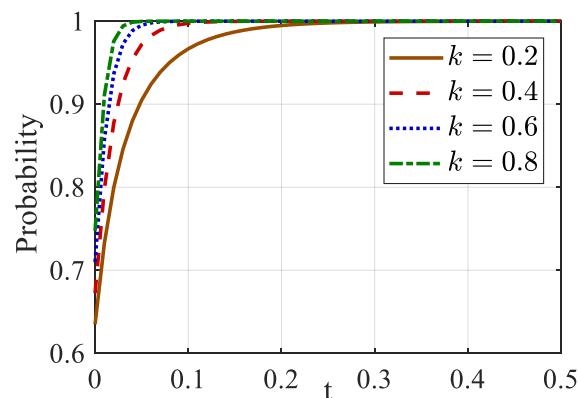


Figure 7. Impact of carbon price fluctuation factor k on the evolutionary trajectory

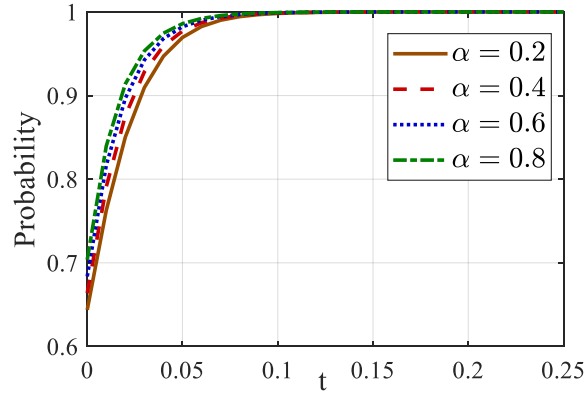


Figure 8. Impact of the market amplification coefficient α of trust repair on the evolutionary trajectory

5.2.2 Sensitivity analysis of third-party certifier's strategy parameters

The sensitivity analysis of the third-party certifier's strategy (participation probability z) confirms the critical role of cost-benefit analysis in its participation decision. The results are consistent with Corollary 3.1, which predicts a negative correlation between certification cost and participation propensity. Specifically, as shown in Figure 9, we set the third-party certifier's participation cost to $C_3=8, 18, 28$ and 38 (million) in turn. The simulation results show that when the cost is low, the third-party certifier can achieve profitability through scale effects (e.g., batch testing), ensuring $z \rightarrow 1$. In contrast, when the cost of participating in carbon footprint testing exceeds the certification revenue, the cost cannot be fully covered even with government rewards. As a result, its participation propensity declines.

Figure 10 depicts the evolutionary trajectories of the third-party's strategy under alternative government reward levels, with $s_3=5, 10, 15$ and 20 (million). The increase in convergence speed when s_3 rises from 5 to 10 is noticeably larger than that when s_3 rises from 15 to 20. This pattern indicates diminishing marginal returns to government rewards; simply increasing the reward amount has limited effectiveness in accelerating strategy evolution. Comparing Figures 9 and 10 suggests that reducing certification costs (e.g., through standardized testing protocols or promoting technological innovation) is more effective than increasing government rewards. This supports the general principle of minimizing institutional friction to enable market functionality.

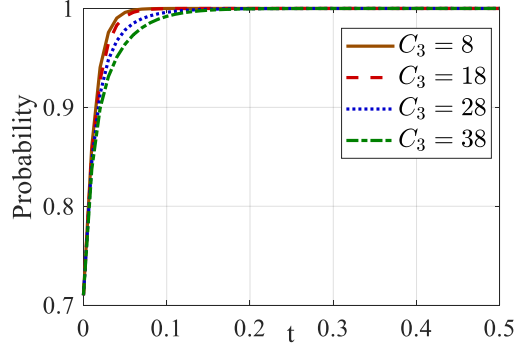


Figure 9. Impact of third-party certification cost C_3 on the evolutionary trajectory

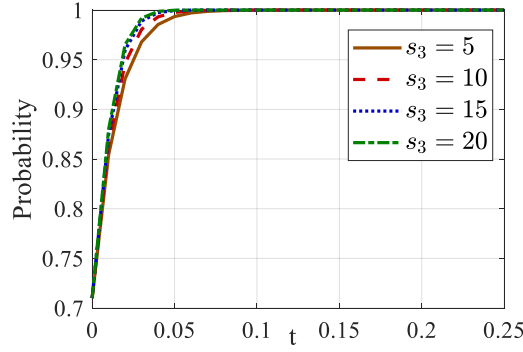


Figure 10. Impact of government reward s_3 on the evolutionary trajectory

5.2.3 Sensitivity analysis of upstream supplier's strategy parameters

This analysis investigates the dynamic stability of the upstream supplier's strategy, truthful disclosure w , under the influence of two critical parameters: government coercive penalty π_4 and the market-driven reputation loss coefficient β . By setting the coercive penalty $\pi_4=15, 25, 40$ and 55 (million), the evolutionary trajectories of the upstream supplier's strategy $w(t)$ are reported in Figure 11. As the government regulator increases the penalty intensity, the system converges substantially faster and the trajectory becomes steeper. This indicates that stronger penalties rapidly raise the cost of false disclosure and shorten the supplier's adjustment cycle.

We then analyze the impact of the reputation loss coefficient, setting $\beta=0.2, 0.4, 0.6, 0.8$ in turn. The simulation results are shown in Figure 12. A clear threshold effect is observed: When $\beta \leq 0.4$, $w(t)$ evolves toward 0, suggesting that low reputation losses are insufficient to curb the supplier's opportunistic behavior. In contrast, when $\beta \geq 0.6$, $w(t)$ converges to 1 rapidly within $t=10$, and the system approaches the ideal equilibrium, implying that high reputation losses impose a strong market discipline. Crucially, when the penalty is set at a medium level $\pi_4=25$ and $\beta \geq 0.6$, the supplier still tends to truthful disclosure. This finding empirically validates that the reputation mechanism can partially substitute for administrative penalties. This strongly supports the efficacy of the "De-emphasizing coercive regulation to empower market mechanisms" autonomous trust repair paradigm, highlighting the strategic value of cultivating a robust market reputation system in low-carbon governance.

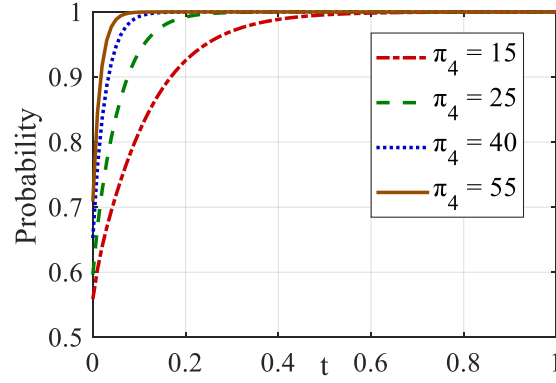


Figure 11. Impact of government penalty intensity π_4 on the evolutionary trajectory

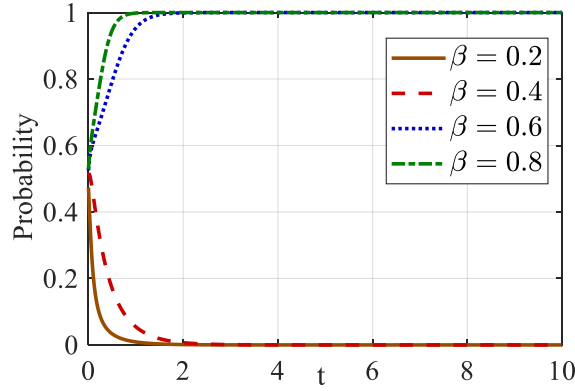


Figure 12. Impact of the supplier's reputation-loss coefficient β on the evolutionary trajectory

5.3 Impact of government regulation on players' evolutionary strategies

The strategic role of the government regulator is contingent on its strategic choice ($x=0$ or $x=1$). It is pivotal in articulating the governance mechanism's transformation. Figure 13 illustrates the system's evolutionary paths under different government boundary strategies. When $x=0$ (lax regulation), the system exhibits non-unique stable strategies, i.e., two potential stable equilibria: $E_1(0, 0, 0, 0)$ and $E_{15}(0, 1, 1, 1)$. This result indicates that, in the absence of effective regulation, market mechanisms may fail to spontaneously converge to the ideal state, highlighting the risk of market failure. When $x=1$ (strict regulation), the system tends to converge to $E_{16}(1, 1, 1, 1)$. Under the strict governmental regulation, the system's convergence speed is significantly accelerated, and all players' strategies stabilize faster.

However, the analysis further reveals a dynamic transition: once the system achieves stability under strict regulation, the government's role will gradually shift from a "front-line enforcer" to a "back-end guarantor". The system ultimately stabilizes at $E_{15}(0, 1, 1, 1)$. In this phase, market mechanisms, such as the focal firm's green premium, the certifier's service revenue, and the supplier's reputational discipline, can become the dominant drivers, thereby helping sustain the long-run autonomous trust repair equilibrium with less reliance on persistent coercive intervention.

This comparative analysis suggests that the government's role is dynamic and contingent. Strict regulation ($x=1$) is essential for initiating the system and guaranteeing initial convergence. Yet, the long-term stable state relies on the strategic regulatory retrenchment of the government, which allows market-based incentives to sustain the equilibrium, thereby achieving the desired market-centric governance.

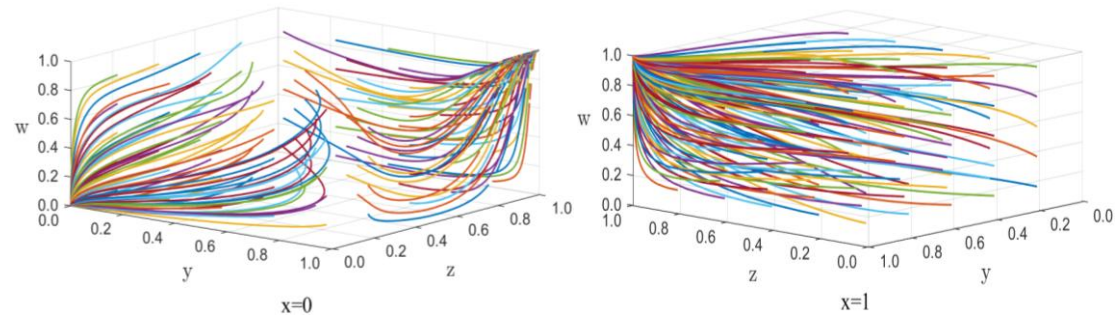


Figure 13. Impact of government regulatory mechanism on player' strategies

5.4 Impact of trust repair on players' evolutionary strategies

The strategic choice of the focal firm regarding trust repair y is pivotal to the stability and efficiency of the entire system, highlighting its role as the market catalyst. As shown in Figure 14, when $y=0$, meaning the focal firm does not engage in trust repair, the system exhibits a clear evolutionary dilemma. All players tend to adopt passive strategies, necessitating increased government regulation to compensate for the deficiency of the market mechanism. While when $y=1$, the system converges to the unique stable strategy $E_{15}(0,1,1,1)$. This demonstrates that the focal firm's trust repair strategy generates significant synergies across the supply chain, which are specifically reflected in the following: (1) Signal transmission mechanism: The Focal firm's green procurement behavior sends a clear low-carbon development signal to the Upstream Supplier. This signal is amplified through investment in blockchain traceability and green reputation, effectively attracting the third-party certification agency to actively participate in carbon footprint verification. (2) Reputation incentive mechanism: The upstream supplier proactively improves the quality of carbon footprint disclosure to gain the focal firm's trust, forming a dual incentive system of "trust + reputation." This mechanism is evident in the rapid convergence of w in Figure 14, indicating that the supplier's initial opportunistic behavior is quickly corrected. (3) Economies of scale: As the volume of verification services increases, third-party certification agency reduces per-unit costs accordingly, thereby enhancing their willingness to participate.

This comprehensive mechanism highlights the critical role of trust repair in facilitating market-driven, reputation-based incentives and enhancing the overall efficiency of supply chain governance.

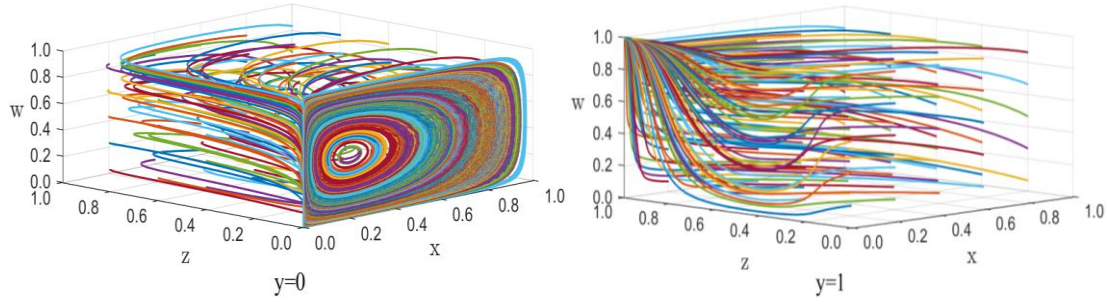


Figure 14. Impact of trust repair mechanism on player' strategies

6 Conclusions and Implications

6.1 Main conclusions

This study utilizes a four-player Evolutionary Game Theory model to analyze the formation and stability of carbon trust within a low-carbon supply chain under information asymmetry. Our numerical experiments and stability analysis yield four primary conclusions, focusing on the conditions required for a sustainable, market-centric governance regime.

First, autonomous carbon trust is condition-dependent. Within the baseline and tested parameter ranges, the system endogenously converges to the ideal equilibrium $E_{15}(0,1,1,1)$. In this regime, the government's strict regulatory probability decreases ($x \rightarrow 0$), while market mechanisms drive the other three players converge to full adoption of trust repair, participation, and truthful disclosure ($y=1, z=1, w=1$). Importantly, this transition is conditional rather than automatic: when market incentives and reputational discipline are weak relative to verification and disclosure costs, the system may remain trapped in the carbon trust deficit state $E_1(0, 0, 0, 0)$. This underscores that market mechanisms are not automatically effective but require specific foundational strengths to emerge.

Second, the focal firm is the pivotal mechanism enabler. The focal firm's trust repair ($y=1$) functions as a pivotal catalyst for escaping the all-violation equilibrium $E_1(0,0,0,0)$ and moving toward the desirable state E_{15} . This catalytic role operates through three reinforcing channels: a signal transmission mechanism that strengthens low-carbon signals along the supply chain, a reputation incentive mechanism that motivates suppliers' truthful disclosure, and an economies-of-scale effect that improves the certifier's participation incentives.

Third, reputation acts as a conditional substitute for coercion. The disciplinary power of the market exhibits a critical threshold effect. Sensitivity results for the upstream supplier indicate that reputation loss (β) can partially substitute for administrative penalties (π_4) once β exceeds a critical level, accelerating convergence to truthful disclosure even under moderate penalty intensity. This substitution, however, depends on the transparency and salience of reputational consequences and may be weakened when disclosure costs are prohibitively high.

Finally, governance efficiency requires institutional optimization. Policy efficiency is highest when focusing on institutional optimization rather than solely

increasing fiscal incentives. For the third-party certifier, the findings indicate a diminishing marginal return to government fiscal incentives (s_3). Consequently, reducing the cost of verification (C_3), e.g., through standardized testing or technological innovation, is a more effective and sustainable lever for ensuring the third party's continued participation ($z=1$) than merely increasing reward amounts.

6.2 Theoretical contributions

This study contributes to the literature on supply chain governance and evolutionary game theory in three ways.

First, we extend the application of evolutionary game theory by introducing a four-player dynamic framework that endogenizes the credibility of carbon information. This approach moves beyond static or single-agent analyses and successfully captures the co-evolution of regulatory policy, trust repair, third-party verification, and supplier disclosure, providing a more comprehensive analytical lens than previous two- or three-player models that often treat information reliability as exogenous. This provides a novel framework for understanding the evolution of regulatory paradigms in multi-player systems.

Second, we propose and empirically validate the concept of Strategic Regulatory Retrenchment. We challenge the prevailing notion that government regulator must act as perpetual “front-line enforcer”. Our findings theoretically correct the regulator's role by demonstrating that their highest strategic utility is achieved by strategically withdrawing to a “back-end guarantor” role after market mechanisms have been firmly established to sustain the E_{15} equilibrium. This provides a theoretical basis for dynamic, cost-efficient regulatory scaling, thereby paving the way for market self-governance.

Third, we advance research on endogenous compliance by identifying the synergistic roles of market-based incentives and reputational mechanisms in sustaining carbon trust repair. It indicates that sustainable carbon compliance is not coercively enforced but strategically earned and maintained through market signaling and self-reinforcing incentives. The analysis highlights threshold effects and substitution/complimentarity patterns between administrative enforcement and market discipline, providing mechanism-based predictions that are transferable across institutional contexts.

6.3 Policy implications

Based on the evolutionary analysis, we propose the following policy implications for fostering sustainable carbon trust repair in supply chains. A key overarching insight is policy sequencing: strong, rule-establishing interventions are often necessary to initiate convergence, while the long-term stability relies more on lowering institutional frictions and amplifying market discipline.

(1) Strategic Reshaping of Regulatory Functions.

Rather than relying on persistent high-intensity enforcement, the government regulator should prioritize the foundational conditions for market trust: unified carbon-footprint accounting standards, credible verification rules, data integrity and

security, and a minimum deterrence baseline for misconduct. This approach maintains regulatory credibility while reducing routine administrative intervention.

(2) Strengthen focal-firm incentives that trigger trust repair.

Policymakers must prioritize policies designed to maximize the market returns to the focal firm's trust repair strategy (e.g., via green finance instruments tied to verified supply-chain performance, or procurement preferences for verifiably low-carbon products). Such measures directly increase the focal firm's green premium (r_2) potential helps sustain the self-reinforcing governance cycle identified in the model.

(3) Cultivating High-Discipline Reputation Systems.

Given the existence of a critical reputation threshold, policy should improve transparency and the visibility of non-compliance through standardized disclosure, credible publication of verification outcomes, and linkages between environmental performance and reputational/credit consequences. When reputational penalties are salient and credible, they can reduce reliance on costly administrative penalties while maintaining compliance.

6.4 Limitations and future research

This study provides a robust framework for analyzing the evolutionary dynamics of carbon trust, yet it is subject to several limitations that suggest avenues for future research. First, players are modeled with bounded rationality under stylized payoff structures and binary strategies; future work could explore richer behavioral updating rules and heterogeneous agent types. Second, the analysis focuses on a single-tier structure; extending the framework to multi-tier supply networks would clarify how trust deficits and governance signals propagate upstream. Third, future research could examine heterogeneity in certifier credibility and supplier disclosure costs, as well as the interaction between dynamic carbon pricing and equilibrium selection. These extensions would further enhance the external validity of the proposed governance transition mechanism.

Reference

1. Aftab, J., Abid, N., Sarwar, H., & Veneziani, M. (2022). Environmental ethics, green innovation, and sustainable performance: Exploring the role of environmental leadership and environmental strategy. *Journal of Cleaner Production*, 378, 134639.
2. Bachmann, R., Gillespie, N., & Priem, R. L. (2015). Repairing trust in organizations and institutions: Toward a conceptual framework. *Organization Studies*, 36(9), 1123–1142.
3. Butt, A. S., Alghababsheh, M., Sindhwani, R., & Gwalani, H. (2025). Role of supplier engagement to reduce Scope 3 emissions in circular supply chains. *Business Strategy and the Environment*, 34(1), 598–611.
4. Cai, X., Li, D., & Wang, C. (2024). Auto parts quality certification and manufacturer regulation: An evolutionary game theory perspective. *Managerial and Decision Economics*, 45(4), 2459–2472.
5. Chen, Y., Ping, L., & Liang, F. H. (2023). Industry reputation crisis and firm certification: A co-evolution perspective. *Journal of Business Ethics*, 186(4), 761–780.

6. Connelly, B. L., Certo, S. T., Reutzel, C. R., DesJardine, M. R., & Zhou, Y. S. (2025). Signaling theory: state of the theory and its future. *Journal of management*, 51(1), 24-61.
7. Dahlmann, F., & Roehrich, J. K. (2019). Sustainable supply chain management and partner engagement to manage climate change information. *Business Strategy and the Environment*, 28(8), 1632–1647.
8. Deng, J., Hung, M., & Wang, S. (2024). *The effect of mandatory carbon disclosure along global supply chains* (Working paper). Hong Kong University of Science and Technology.
9. Eghbali, R., Rasti-Barzoki, M., & Safarzadeh, S. (2024). An evolutionary game-theoretic approach to analyse the green innovation chain dynamics under government policies. *Technology in Society*, 76, 102568.
10. Fan, W. J., Fang, Y., & Jiang, R. B. (2024). An analysis of optimal equilibrium in the carbon trading market-From a tripartite evolutionary game perspective. *International Review of Financial Analysis*, 96, 103629.
11. Fan, W., Wu, X., & He, Q. (2024). Digitalization drives green transformation of supply chains: A two-stage evolutionary game analysis. *Annals of Operations Research*. Advance online publication. <https://doi.org/10.1007/s10479-024-06050-0>
12. Gao, Y. L., Gong, B., Liu, Z., Tang, J., & Wang, C. (2023). The behavioural evolution of the smart electric vehicle battery reverse supply chain under government supervision. *Industrial Management & Data Systems*, 123(10), 2577-2606.
13. Hu, X., Deng, J., & Li, Y. (2024). Supply chain digitalization, green technology innovation and corporate energy efficiency. *Energy Economics*. Advance online publication.
14. Ioannou, I., Kassinis, G., & Papagiannakis, G. (2023). The impact of perceived greenwashing on customer satisfaction and the contingent role of capability reputation. *Journal of Business Ethics*, 185(2), 333–347.
15. Li, J., Li, S., Zhang, Y., & Tang, X. (2024). Network evolutionary game analysis of green credit: A perspective of carbon emissions trading. *Managerial and Decision Economics*, 45(3), 1343-1362.
16. Ma, J., Li, Q., Zhao, Q., Liou, J., & Li, C. (2024). From bytes to green: The impact of supply chain digitization on corporate green innovation. *Energy Economics*, 139, 107942.
17. Ni, J., Xu, Y., Shi, J., & Li, J. (2024). Product innovation in a supply chain with information asymmetry: Is more private information always worse? *European Journal of Operational Research*, 314, 229–240.
18. Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.
19. Qian, J., & Zhou, Y. (2025). Environmental regulation and green technology innovation: An evolutionary game analysis between government and high energy-consuming enterprises. *Computational Economics*. Advance online publication.
20. Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, 57(7), 2117–2135.
21. Sun, H., Gao, G., & Li, Z. (2022). Evolutionary game analysis of enterprise carbon emission regulation based on prospect theory. *Soft Computing*, 26(24), 13357-13368.
22. Taj, S. A. (2016). Application of signaling theory in management research: Addressing major gaps in theory. *European Management Journal*, 34(4), 338-348.

23. Tu, J.-C., Cui, Y., Liu, L., & Yang, C. (2024). Perceived greenwashing and its impact on the green image of brands. *Sustainability*, 16(20), 9009.
24. Wang, Z., Zhao, L., Shao, Y., & Wen, X. (2023). Reputation compensation for incentive alignment in a supply chain with trade credit under information asymmetry. *Annals of Operations Research*, 331, 581–604.
25. Xu, J., Tong, B., Wang, M., & Yin, S. (2024). How to systematically reduce the carbon emissions of the manufacturing industry? Evidence from four-party evolutionary game analysis. *Environmental Science and Pollution Research*, 31(2), 2614-2639.
26. Yi, H., & Cui, C. (2019). Coping with functional collective action dilemma: Functional fragmentation and administrative integration. *Public Management Review*, 21(7), 1052-1075.
27. Zhang, F., Wu, L., Liu, W., Ding, K., Hui, J., Leng, J., & Zhou, X. (2022). Evolutionary game-based incentive models for sustainable trust enhancement in a blockchained shared manufacturing network. *Advanced Engineering Informatics*, 54, 101791.
28. Zhang, L., Li, X., & Lu, Y. (2025). Leveraging environmental regulation: How green innovation moderates the relationship between carbon information disclosure and firm value. *Sustainability*, 17(6), 2597.
29. Zhang, Z., Shi, K., Gao, Y., & Feng, Y. (2025). How does environmental regulation promote green technology innovation in enterprises? A policy simulation approach with an evolutionary game. *Journal of Environmental Planning and Management*, 68(5), 979–1008.