

Article

ESG Rating Disagreement as a Greenwashing Signal: Asymmetric Effects of Digital Transformation Through Disclosure and Performance Channels

İsmail Ögütçen ¹  and Ümit Yılmaz ^{2,*} 

¹ Department of Finance, Banking and Insurance, Bigadiç Vocational School, Balıkesir University, Balıkesir 10440, Türkiye; ismailogutcen@balikesir.edu.tr

² Quality Coordination Office, Bursa Technical University, Bursa 16310, Türkiye

* Correspondence: umit.yilmaz@btu.edu.tr

Abstract

This study examines whether ESG rating disagreement is a leading indicator of corporate greenwashing and how digital transformation (DTI) moderates this relationship through disclosure and performance channels. Using 8111 firm-year observations from Chinese A-share companies (2012–2022), we employ two-way fixed-effects panel regression complemented by Bayesian-optimised machine learning models interpreted through SHAP. Aggregate rating disagreement is a strong and robust predictor of greenwashing. Channel decomposition reveals asymmetric DTI moderation: the disclosure channel amplifies greenwashing risk as digitally advanced firms expand reporting capacity to widen the gap between disclosed and actual ESG performance (bloom_DTI: $\beta = +0.2471$, $p < 0.01$), while the performance channel attenuates greenwashing risk as digital operational monitoring translates substantive performance into a measurable reduction (hua_DTI: $\beta = -0.2804$, $p < 0.01$). This pattern is robust across ownership structure, pollution intensity, and region. Machine learning analysis confirms the econometric findings and reveals nonlinear threshold effects invisible to panel regression. This asymmetric channel mechanism contributes to the ESG rating divergence literature and has implications for disclosure regulation and ESG-based investment screening.

Keywords: ESG rating disagreement; greenwashing; digital transformation; panel regression; machine learning; SHAP; China



Academic Editors: Bruce Morley,
Gomaa Mohamed Agag and
Mohamed A. Khashan

Received: 31 May 2026

Revised: 18 June 2026

Accepted: 28 June 2026

Published: 4 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Environmental, social, and governance (ESG) ratings have become indispensable to sustainable investment, shaping capital allocation decisions for trillions of dollars in global assets [1]. Yet a well-documented paradox undermines their utility: agencies assessing the same company routinely reach divergent conclusions. Berg, Kölbel and Rigobon [1] demonstrated that pairwise correlations among six major ESG rating providers range from 0.38 to 0.71, which is starkly lower than the near-unity agreement observed among credit rating agencies. This divergence, decomposed into measurement (56%), scope (38%), and weighting (6%) contributions [1], generates an information environment in which a firm may simultaneously appear as an ESG leader to one rater and an average performer to another.

While the emergence of ESG rating agencies provides non-financial benchmarks for sustainable investments, the severe lack of convergence among them has created

an opaque and incoherent ESG ecosystem. Recent literature indicates that ESG rating disagreement is not merely a product of methodological differences, but can be a direct manifestation of corporate hypocrisy, in which firms decouple their idealised rhetoric from their actual practices to manage stakeholder impressions [2]. This rating divergence severely misleads investors and has detrimental consequences across the capital market. Empirically, ESG rating disagreement has been shown to weaken ESG's role in reducing the cost of equity capital [3]. Furthermore, it amplifies information asymmetry, increasing the risk of stock price crashes by shaking investor confidence [4]. Beyond financial markets, the confusion created by disparate ratings actively suppresses corporate green innovation [5] and undermines corporate energy efficiency by diverting resources away from substantive environmental improvements [6]. In this context, treating ESG rating disagreement as a strategic greenwashing signal rather than mere statistical noise is critical.

An underexplored implication of this divergence is its potential use as a greenwashing signal. Previous literature suggests that when a firm receives markedly higher scores on disclosure-oriented metrics than on performance-oriented ones, the gap reflects substantive decoupling, communicating sustainability more than delivering it [7,8]. Building on this conceptual framework, we operationalise this intuition for our specific context by distinguishing Bloomberg ESG (representing disclosure breadth) from HuaZheng ESG (which we utilise to capture substantive environmental and social outcomes), constructing a globally standardised rating disagreement measure ($RD = \text{bloom_g} - \text{hua_g}$) to capture the disclosure–performance wedge at the firm level.

Simultaneously, digital transformation (DTI) has emerged as a significant governance mechanism. By enhancing information processing capacity, improving internal monitoring, and reducing information asymmetry, DTI has been shown to inhibit ESG greenwashing [9,10]. However, whether DTI operates symmetrically across disclosure and performance channels remains unexplored. Firms that digitise rapidly may improve their disclosure sophistication (widening Bloomberg-type scores) without a proportional improvement in substantive performance, or digital operational monitoring may predominantly improve HuaZheng-type scores. Disentangling these channels is essential for both policy design and investment practice.

Against this backdrop, this study addresses the following research question: Does disagreement in ESG ratings between disclosure-oriented and performance-oriented rating providers constitute a systematic greenwashing signal, and how does digital transformation asymmetrically moderate this relationship through disclosure and performance channels? This paper makes three interrelated contributions. First, we formalise ESG rating disagreement as a greenwashing leading indicator by decomposing it into Bloomberg-type disclosure and HuaZheng-type performance components, and estimating their asymmetric effects using a two-way fixed-effects panel model with channel-specific DTI interaction terms. Second, we identify an asymmetric two-channel DTI moderation pattern in which performance-channel attenuation ($\text{hua_DTI} = -0.2804$) marginally outweighs disclosure-channel amplification ($\text{bloom_DTI} = +0.2471$) in absolute magnitude. Third, we complement the econometric analysis with a machine learning layer comprising four gradient-boosting and neural architectures optimised via Bayesian search, and apply SHAP to identify nonlinear threshold effects and heterogeneous firm profiles that panel regression cannot capture. The analysis employs the publicly available dataset from Zhao [11], constituting a novel empirical contribution. China provides an ideal laboratory for this investigation for three reasons. First, Chinese A-share markets feature pronounced ESG rating divergence, driven by the coexistence of international providers (Bloomberg) and domestic outcome-focused agencies (HuaZheng) [12,13]. Second, China's Digital Economy Development Strategy has generated substantial cross-firm variation in DTI adoption over

the 2012–2022 period, providing the identification leverage needed for channel decomposition. Third, the publicly available Zhao [11] dataset enables full replication while covering a sufficiently long panel for within-firm fixed-effects estimation.

Unpacking the aggregate inhibitory finding of Li, Miao, and Xu [9], our channel decomposition reveals that DTI operates asymmetrically: amplifying greenwashing risk through the disclosure channel while attenuating it through the performance channel, with the net effect determined by the relative strength of these opposing forces at the firm level. The performance channel marginally outweighs the disclosure channel in absolute magnitude, but the disclosure channel effect becomes economically significant at high DTI adoption levels.

The remainder of the paper proceeds as follows. Section 2 reviews the literature and develops hypotheses. Section 3 describes data and methodology. Section 4 presents results. Section 5 discusses findings. Section 6 concludes.

2. Literature Review and Hypotheses Development

2.1. ESG Rating Disagreement, Information Asymmetry, and Greenwashing

The landmark decomposition by Berg, Kölbel, and Rigobon [1] reveals that ESG rating divergence stems primarily from measurement differences in how agencies operationalise nominally identical attributes, supplemented by scope disagreements over which attributes to include. Divergence is not innocuous: it restricts bank loan availability [14], elevates financing costs [12], reduces institutional holdings [15], and increases real earnings management [16]. In the Chinese A-share context, ESG rating divergence is structurally embedded and particularly pronounced [13,17]. This becomes especially evident when contrasting domestic agencies such as HuaZheng, which apply outcome-focused frameworks, with international providers such as Bloomberg, which emphasise breadth of disclosure.

Hu, Hua, Liu, and Wang [7] introduced the “green fog” metaphor to describe how rating disagreement enables strategic greenwashing, in which firms exploit informational ambiguity to claim unwarranted ESG legitimacy. Subsequent panel evidence from 16,653 Chinese firm-year observations confirms a robust positive association between ESG rating divergence and the propensity to greenwash [8]. The underlying mechanism is information asymmetry: when ratings diverge, the market signal is corrupted, reducing disciplinary pressure on firms that disclose selectively or symbolically. Furthermore, recent evidence suggests that such rating disagreement acts as a precursor to potential ESG risks, often manifesting as lower future ESG performance [18]. A critical dimension neglected in this literature is the directionality of disagreement: disagreement driven by a firm scoring high on disclosure-oriented ratings but low on performance-oriented ratings carries qualitatively different greenwashing implications than the reverse. We exploit the Bloomberg–HuaZheng architectural contrast to construct directional hypotheses. The operationalisation of greenwashing as the standardised Bloomberg–HuaZheng score differential follows an established strand of the rating disagreement literature. Following Hu, Hua, Liu, and Wang [7], who explicitly frame the divergence between disclosure-oriented ratings (Bloomberg) and performance-oriented ratings (Wind) as a disclosure-versus-performance wedge, we conceptualise the Bloomberg–HuaZheng divergence similarly: higher Bloomberg scores relative to HuaZheng scores signal symbolic rather than substantive ESG commitment. Zhang, Zheng, and Meng [8] adopt a structurally analogous construction using Chinese A-share data, operationalising ESG-washing precisely as this directional divergence, with high disclosure scores relative to performance scores. The present study extends this approach by standardising scores globally before differencing, ensuring that cross-firm comparisons reflect genuine disclosure–performance gaps rather than scale artefacts. While this measure captures the disclosure–performance wedge at the

firm level, it does not directly verify misconduct through regulatory records; its validity rests on the theoretical and empirical consensus that persistent divergence between disclosure and performance metrics constitutes a reliable greenwashing signal in the Chinese institutional context [7,8].

Hypothesis 1a (H1a): *Higher Bloomberg ESG scores (disclosure orientation) are positively associated with greenwashing risk.*

Hypothesis 1b (H1b): *Higher HuaZheng ESG scores (performance orientation) are negatively associated with greenwashing risk.*

Hypothesis 1c (H1c): *Digital transformation moderates both the disclosure-channel and performance-channel associations asymmetrically, such that the moderation effects differ in sign.*

Digital transformation captures the adoption of digital technologies, including big data analytics, artificial intelligence, blockchain, and cloud computing. The textual frequency of digital keywords measures it in annual reports [9,10].

2.2. Digital Transformation as a Governance Mechanism

Digital transformation (DTI) serves as a potent governance mechanism to align corporate environmental responsibility with actual performance. Extant studies demonstrate that digital empowerment significantly reduces Corporate Environmental Responsibility (CER) decoupling and ESG greenwashing by enabling dynamic resource reconfiguration, real-time monitoring, and strengthening internal controls [19–21]. Additionally, DTI mitigates greenwashing by driving innovation in green technologies and reducing information asymmetry through enhanced online platform interactions between enterprises and investors [19,21,22]. The adoption of big data and artificial intelligence increases transparency, elevating the reputational cost of symbolic disclosures, and alleviating the financial constraints that traditionally motivate greenwashing behaviours [20,23,24]. Furthermore, the positive impact of DTI on ESG outcomes is often more pronounced in non-high-tech industries, where digital upgrading overcomes legacy inefficiencies [25,26].

Specifically, Li, Miao, and Xu [9] established that DTI significantly inhibits ESG greenwashing in Chinese listed firms, with mechanisms running through improved internal controls and alleviated financing constraints. Xu and Li [10] corroborated this using a fraud risk factor framework, showing that digitalisation reduces greenwashing by mitigating motivations, reducing opportunities, and enhancing exposure. However, the channel-level analysis is still missing. On the disclosure channel, digital technologies expand data collection and reporting capacity, potentially inflating Bloomberg-type scores without commensurate performance improvements, a “data abundance without substance” effect. On the performance channel, advanced digital monitoring (IoT-based emissions tracking, AI-assisted supply chain auditing) directly improves substantive environmental outcomes, narrowing the Bloomberg–HuaZheng gap. We hypothesise asymmetric channel effects based on the mechanistic logic that operational improvements, being less reversible and more objectively verifiable, generate stronger performance-channel attenuation than disclosure-channel amplification in absolute magnitude.

Hypothesis 2a (H2a): *Digital transformation amplifies the greenwashing-enhancing effect of disclosure-oriented scores ($bloom_DTI > 0$): digitally advanced firms with high Bloomberg scores exhibit greater greenwashing risk as expanded reporting capacity widens the disclosure–performance gap.*

Hypothesis 2b (H2b): *Digital transformation strengthens the greenwashing-attenuating effect of performance-oriented scores ($hua_DTI < 0$): digitally advanced firms with high HuaZheng scores exhibit lower greenwashing risk as digital operational monitoring translates substantive performance into measurable greenwashing reduction.*

2.3. Machine Learning Approaches to Greenwashing Detection

Machine learning approaches to ESG greenwashing are nascent but growing. Zeng et al. [27] demonstrate that XGBoost with Improved Hunter-Prey Optimisation (IHPO) achieves $R^2 = 0.979$ in greenwashing prediction, and Jianfeng and Tiantian [28] show that XGBoost with SHAP outperforms Random Forest, SVM, and LightGBM on Chinese listed firm data. Critically, SHAP, grounded in Shapley values from cooperative game theory [29], provides theoretically principled, post hoc explanations that decompose predictions additively across features, satisfying the efficiency, symmetry, and dummy axioms. This renders SHAP values interpretable in the same economic register as regression coefficients, enabling a principled dialogue between econometric and machine learning findings.

3. Materials and Methods

3.1. Dataset

The empirical analysis draws on the publicly available dataset deposited by Zhao [11] in the Mendeley Data repository. The dataset covers Chinese A-share-listed companies on the Shanghai and Shenzhen exchanges from 2012 to 2022. After excluding financial sector firms, ST/*ST-designated companies, and firm-years with missing values on key variables, the final sample comprises 8111 firm-year observations from 1265 unique firms, yielding an unbalanced panel. The dependent variables *esgdis* and *esgper* are drawn directly from this dataset, in which Bloomberg (disclosure-oriented) and HuaZheng (performance-oriented) ESG scores have been standardised at the sector-year level. For the channel decomposition, we construct *bloom_g* and *hua_g* as globally standardised (zero-mean, unit-variance) versions of the winsorised raw scores, thereby enabling cross-firm comparability of the interaction terms.

The 1265 unique firms span 65 CSRC industry codes, with the top three sectors being Computer, Communication, and Other Electronic Equipment Manufacturing (C39, 9.6%), Chemical Raw Materials and Products Manufacturing (C26, 7.4%), and Electrical Machinery and Apparatus Manufacturing (C27, 7.2%). State-owned enterprises (SOEs) account for 44.7% of firms. Eastern-region firms represent 46.6% of the sample. Firms classified as polluting under MEE guidelines constitute 47.3% of the sample. Big-4-audited firms account for 12.1% of the sample, consistent with the broader pattern of international auditor concentration among larger listed firms. The sample grows from 543 firm-years in 2012 to a peak of 1039 in 2021, reflecting China's expanding ESG disclosure requirements over the period.

3.2. Variable Construction

The primary outcome variable, the greenwashing score (GWS), is defined as the difference between globally standardised Bloomberg ESG scores (*bloom_g*) and HuaZheng scores (*hua_g*), with both scores standardised to zero mean and unit variance after winsorising at the 1st and 99th percentiles. Positive values indicate that a firm projects a stronger disclosure-based ESG image than its substantive performance warrants, the core operational definition of greenwashing employed in this study. A binary indicator ($GWS_{binary} = 1$ if $GWS > \text{sample median}$) is constructed for robustness analysis, yielding a near-balanced split (49.7% positive) suitable for classification tasks. Of the 8111 firm-years, 33.2% exhibit

an economically positive gap ($GWS > 0$); the median-based binary definition therefore captures firms with above-median disclosure–performance divergence relative to peers rather than firms with strictly positive gaps.

The channel decomposition separates *bloom_g* (global standardised Bloomberg score) and *hua_g* (global standardised HuaZheng score) into distinct regressors that capture the disclosure and performance dimensions independently. Their interaction with DTI ($bloom_DTI = bloom_g \times DTI$ and $hua_DTI = hua_g \times DTI$) constitutes the main channel-identification variables. The rating disagreement measure $RD = bloom_g - hua_g$ is included as an aggregate divergence control in Specification 2.

Control variables are drawn from Zhao [11] and follow the broader ESG-greenwashing literature: firm size (log total assets), leverage, ROA, revenue growth, Tobin’s Q, largest shareholder concentration, institutional ownership, board size, board independence, SOE dummy, Big4 auditor dummy, firm age, and provincial GDP growth.

Each control is included on theoretical and empirical grounds. Firm size (log total assets) proxies for disclosure capacity and market visibility; larger firms face greater stakeholder scrutiny and possess resources to invest in ESG signalling [7]. Leverage captures financial constraint and agency costs that incentivise symbolic ESG disclosure [9]. ROA controls for financial performance, as profitable firms face lower pressure to greenwash but may have more resources for substantive ESG investment. Revenue growth and Tobin’s Q capture growth opportunities and market valuation effects on ESG behaviour. Largest shareholder concentration (Top1) and institutional ownership (INST) proxy for internal governance mechanisms and external monitoring intensity [15]. Board size and board independence reflect governance quality, which affects disclosure decisions. SOE and Big4 dummies control for ownership-type and audit quality effects on reporting credibility. Firm age captures reporting experience and reputation capital. Provincial GDP growth controls for local economic conditions affecting ESG investment capacity.

Table 1 presents descriptive statistics for all variables. The GWS continuous variable has a mean of -0.4804 and a median of -0.5255 , indicating that the average firm-year exhibits a negative disclosure–performance gap; only 33.2% of firm-years show an economically positive gap ($GWS > 0$). DTI has a mean of 1.4316 with substantial right skew (median = 1.099), reflecting the concentration of highly digitalised firms in the upper tail. Rating disagreement (RD) has a mean of zero by construction and a standard deviation of 1.288, with a right-skewed distribution (max = 5.459) indicating that large positive disagreement, firms scoring far higher on disclosure than performance, is more extreme than negative disagreement. The near-equal binary GWS split (49.7% positive) confirms that the median-based classification yields a balanced sample suitable for classification tasks.

Table 1. Descriptive statistics (N = 8111 firm-year observations, 2012–2022).

Variable	N	Mean	Std	Min	P25	Median	Max
GWS (greenwashing score)	8111	−0.4804	1.2277	−3.2771	−1.3250	−0.5255	2.8090
GWS binary	8111	0.4971	0.500	0.000	0.000	1.000	1.000
ESG Disclosure (esgdis)	8111	−0.0067	0.9413	−1.4020	−0.7070	−0.3356	2.8386
ESG Performance (esgper)	8111	0.4748	0.9915	−1.7099	−0.2557	0.4612	3.0042
Rating Disagreement (RD)	8111	0.0000	1.2878	−3.6709	−0.8636	−0.0693	5.4587
Digital Transformation (DTI)	8111	1.4316	1.3310	0.0000	0.0000	1.0986	5.0370
Firm Size	8111	23.3861	1.2578	20.6769	22.4864	23.3036	26.3978
Leverage	8111	0.4817	0.1894	0.0792	0.3390	0.4957	0.8654
ROA	8111	0.0489	0.0583	−0.1415	0.0165	0.0404	0.2204
Tobin’s Q	8111	1.9019	1.3191	0.8270	1.0971	1.4421	8.3116
State Ownership (SOE)	8111	0.5310	0.4991	0.0000	0.0000	1.0000	1.0000

3.3. Panel Regression Framework

Three nested specifications are estimated for each of three dependent variables, *esgdis* (disclosure scores), *esgper* (performance scores), and *GWS* (greenwashing). The primary specification (Spec-3) is:

$$Y_{it} = \alpha + \beta_1 DTI_{it} + \beta_2 bloom_DTI_{it} + \beta_3 hua_DTI_{it} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (1)$$

where $bloom_DTI_{it} = bloom_git \times DTI_{it}$ and $hua_DTI_{it} = hua_git \times DTI_{it}$ denote the channel-specific interaction terms; μ_i and λ_t denote firm and year fixed effects, respectively, and X contains 13 firm-level controls. Standard errors are clustered by both firm and year (two-way clustering) to account for within-firm serial correlation and cross-sectional dependence simultaneously [30]. Driscoll-Kraay kernel-robust standard errors (bandwidth selected by the $T^{1/4}$ rule) are additionally reported as a robustness check, providing the most conservative inference under arbitrary cross-sectional and temporal dependence structures. Estimation uses the PanelOLS estimator from the `linearmodels` package in Python 3.12.13.

Diagnostic tests justify the fixed-effects specification. The F-test for joint significance of entity fixed effects rejects pooled OLS for all three dependent variables (*esgdis*: $F = 11.41$, $p < 0.001$; *esgper*: $F = 10.94$, $p < 0.001$; *GWS*: $F = 2.46$, $p < 0.001$). The Hausman test rejects random effects in favour of fixed effects for all three outcomes (*esgdis*: $\chi^2(14) = 39.40$, $p < 0.001$; *esgper*: $\chi^2(14) = 47.81$, $p < 0.001$; *GWS*: $\chi^2(14) = 40.27$, $p < 0.001$), confirming that individual effects are correlated with the regressors. The Wooldridge test detects first-order panel autocorrelation, and the Modified Wald test detects panel heteroscedasticity across all dependent variables (both $p < 0.001$), motivating the use of two-way clustered and Driscoll-Kraay standard errors. Im-Pesaran-Shin (IPS) panel unit root tests confirm the stationarity of the key continuous variables. *DTI* is stationary in levels ($W\text{-bar} = -15.30$, $p < 0.001$). For *bloom_g*, a first-difference IPS test rejects the null hypothesis of $I(2)$ ($W\text{-bar} = -14.43$, $p < 0.001$), confirming that *bloom_g* is at most $I(1)$ and validating level regressions. All remaining continuous variables are similarly confirmed stationary in levels or at most $I(1)$.

For the binary *GWS* specification, a Linear Probability Model (LPM) with two-way fixed effects is employed. Binary logit with high-dimensional fixed effects suffers from the incidental parameters problem, which introduces asymptotic bias in the slope estimates as the number of fixed effects grows with the sample [31]. While bias-corrected logit estimators exist, the LPM with clustered standard errors yields consistent average marginal effects under the same two-way fixed-effects structure without requiring distributional assumptions about the error term and remains the standard approach in large-panel settings with binary outcomes [30,31]. Coefficient estimates from the LPM are therefore interpreted as linear probability differences rather than odds ratios.

Specification 3 is estimated without *bloom_g* and *hua_g* as standalone main-effect regressors (see Table 2). The channel decomposition requires this exclusion: because *GWS* is constructed as $bloom_g - hua_g$, entering these components freely as main effects would conflate their algebraic contribution to the dependent variable with the moderating role of *DTI*, rendering the interaction coefficients uninterpretable as channel-identification estimates. The specification, therefore, loads identification entirely onto *bloom_DTI* and *hua_DTI*, isolating how digital transformation differentially amplifies or attenuates each channel. This mirrors moderated regression designs in which constituent terms of a theoretically motivated composite are absorbed into the interaction structure rather than entered as unrestricted regressors.

Table 2. Panel regression results.

Variable	esgdis S1	esgdis S2	esgdis S3	esgper S1	esgper S2	esgper S3	gws S1	gws S2	gws S3
DTI	0.0232 (1.174)	0.0285 * (1.775)	−0.0187 (−1.051)	0.0583 *** (3.132)	0.0519 *** (3.293)	0.0170 (0.979)	−0.0356 (−1.397)	−0.0241 * (−1.766)	−0.0354 (−1.449)
RD		0.4369 *** (15.632)			−0.5622 *** (−15.240)			0.9889 *** (25.869)	
RD × DTI		−0.0261 *** (−3.392)			0.0155 ** (1.575)			−0.0396 *** (−3.812)	
bloom_g			-			-			-
hua_g			-			-			-
bloom_DTI			0.2474 *** (20.951)			−0.0035 (−0.502)			0.2471 *** (11.105)
hua_DTI			−0.0059 (−1.459)			0.2736 *** (23.518)			−0.2804 *** (−20.978)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Two-way	Two-way	Two-way	Two-way	Two-way	Two-way	Two-way	Two-way	Two-way
Within-R ²	0.0229	0.3031	0.3021	0.0220	0.4445	0.3864	0.0050	0.7366	0.3182
N	8111	8111	8111	8111	8111	8111	8111	8111	8111

t-statistics in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by firm and year (two-way clustering). Spec-3 excludes bloom_g and hua_g as standalone regressors by design, not due to insignificance; see Section 3.3 for rationale. Channel identification is carried entirely by the bloom_DTI and hua_DTI interaction terms.

3.4. Machine Learning Pipeline

Four supervised learning architectures were employed: XGBoost [32] and LightGBM [33] as gradient boosting methods with level-wise and leaf-wise tree-growth strategies; Random Forest [34] as an independent bagging-based benchmark; and a Multilayer Perceptron (MLP) [35] as a deep architecture capturing higher-order feature interactions. The four-model ensemble reflects a deliberate design: gradient-boosting methods from different families enable cross-validation of nonlinear findings, while the RF and MLP serve as independent structural references.

All models were optimised via Bayesian hyperparameter search using Optuna (Tree-structured Parzen Estimator sampler) [36], with identical budgets: 50 trials per model, 5-fold cross-validation, and seed = 42. For tree-based models, the common search space included n_estimators (100–500), max_depth (3–8), learning_rate (0.01–0.30, log scale), subsample (0.6–1.0), and L1/L2 regularisation (10^{-8} –10.0, log scale). Random Forest replaced structural regularisation parameters (min_samples_split, min_samples_leaf) with the learning rate, thereby constituting the architectural equivalent of regularisation in bagging models. MLP search space covered layer count (1–3), units per layer (32–256), and learning rate (10^{-4} – 10^{-1}). Regression models minimised RMSE; classification models maximised AUC-ROC. The 20 input features were standardised using StandardScaler. Optimised hyperparameters for all four models are reported in the Supplementary Material (Table S1). To verify that predictive performance is not driven by the overlap between bloom_g, hua_g, RD, and the GWS construction, a strict orthogonal ablation was conducted, excluding all components algebraically related to GWS—specifically bloom_g, hua_g, RD, and their interaction terms (bloom_DTI, hua_DTI, RD_DTI)—retaining only DTI and the 13 firm-level controls (14 features total).

3.5. SHAP Interpretability

The SHAP framework [29] was applied via the TreeExplainer algorithm to the best-performing tree-based model to decompose individual predictions into additive feature contributions satisfying the local accuracy, missingness, and consistency properties. Global importance is assessed via mean absolute SHAP values. Beeswarm plots reveal directional

effects for all features simultaneously. Dependence plots expose nonlinear threshold dynamics for the key interaction terms *bloom_DTI* and *hua_DTI*. Subgroup SHAP analysis compares feature importance profiles between SOE and non-SOE firms, complementing the regression-based heterogeneity tests.

3.6. Robustness Strategy

Six robustness exercises complement the main analysis. First, one-period-lagged predictor variables ($t - 1$) are used in place of contemporaneous values to mitigate simultaneity bias. Second, VIF analysis assesses multicollinearity among interaction terms. Third, the binary GWS specification estimated via LPM serves as an alternative to nonlinear models. Fourth, two-way clustered standard errors ($\text{firm} \times \text{year}$) and Driscoll-Kraay kernel-robust standard errors are estimated for all main specifications to address cross-sectional dependence and heteroscedasticity, both confirmed by diagnostic tests. Fifth, a residualised specification orthogonalises *bloom_g* and *hua_g* with respect to DTI and controls before constructing interaction terms; VIF values fall below 2.6 under this specification. Sixth, uniform sector-year standardisation is applied to *bloom_g* and *hua_g*, and keyword density (*dig*) replaces DTI as an alternative operationalisation. Heterogeneity analyses partition the sample by SOE status, pollution intensity (MEE classification), and geographic region (eastern vs. central-western China), and report the coefficients of interest (*bloom_DTI*, *hua_DTI*) for each subgroup.

4. Results

4.1. Exploratory Analysis

Figure 1 presents annual trends in mean GWS, DTI, and rating disagreement from 2012 to 2022. DTI rises steadily from approximately 0.6 in 2012 to 1.9 in 2022. Rating disagreement remains negative through 2017, then turns sharply positive after 2018, coinciding with China's accelerated digital economy policies and the broader expansion of ESG reporting requirements. This divergence in trajectories provides descriptive motivation for the channel analysis.

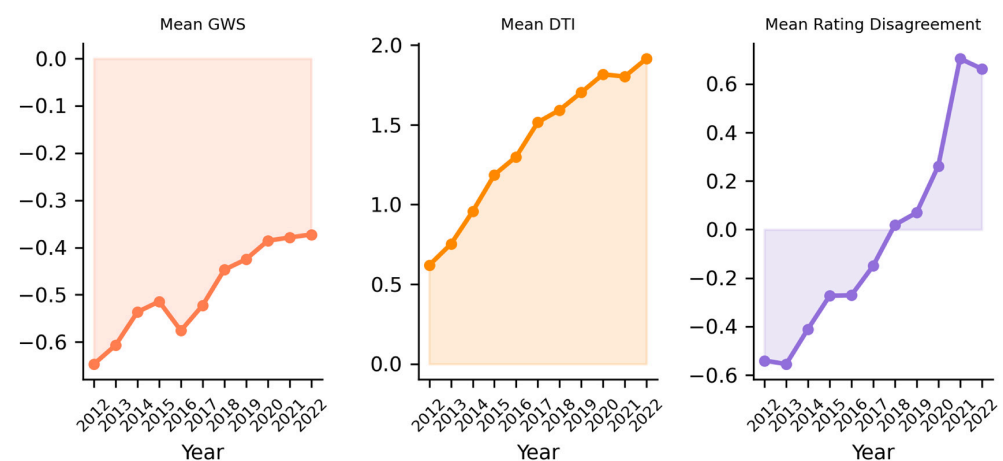


Figure 1. Annual trends 2012–2022.

Figure 2 presents the distributions of key variables (upper two rows) and GWS distributional evidence (bottom row). The GWS distribution has a slight right skew (skewness = +0.23) and a negative mean (-0.480), indicating that the average firm-year exhibits a negative disclosure–performance gap. Only 33.2% of firm-years exhibit an economically positive gap ($\text{GWS} > 0$); the remaining 66.8% have negative gaps. The binary GWS variable ($\text{GWS} = 1$ if $\text{GWS} > \text{sample median}$) captures firms with above-median

disclosure–performance divergence, yielding a near-balanced 50/50 split by construction. The negative overall mean is driven by large-magnitude negative outliers: the mean of the negative-gap group (−1.16) substantially exceeds the mean of the positive-gap group (+0.88) in absolute value, pulling the overall mean negative despite the near-balanced binary split. The bottom row of Figure 2 further illustrates this structure: the histogram confirms the left-skewed distribution with median split, the quantile plot shows that negative values dominate below the 60th percentile, and the annual trend panel reveals that the share of economically positive gaps has increased from approximately 30% in 2012 to near 50% by 2022, coinciding with China’s accelerating ESG reporting requirements.

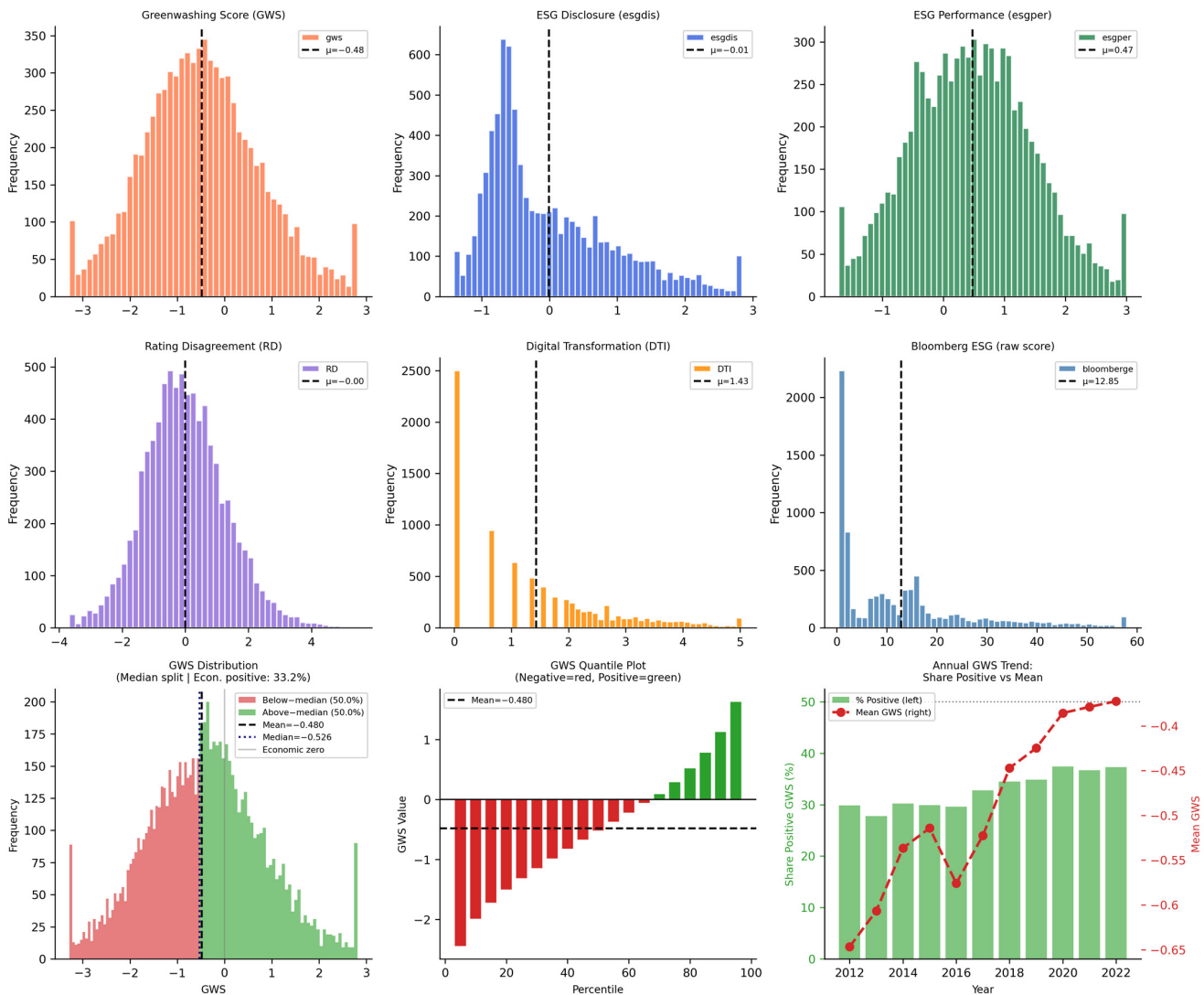


Figure 2. Distribution of key variables and GWS distributional evidence.

Figure 3 presents the correlation matrix for key variables. GWS is positively correlated with bloom_g ($r = 0.61$) and negatively correlated with hua_g ($r = -0.58$), consistent with the directional hypotheses. RD and GWS share a correlation of 0.73, confirming that the aggregate disagreement measure is a strong predictor of the channel-decomposed greenwashing score. Figure 4 presents two complementary visualisations of the RD–GWS relationship across DTI levels. The left panel displays scatter plots with fitted trend lines per DTI tercile; all three terciles show a strong positive RD–GWS association, with low-DTI firms exhibiting a marginally steeper gradient. The right panel plots GWS distributions by RD quartile for low- and high-DTI firms, showing that low-DTI firms tend toward higher

GWS in the lower RD quartiles (Q1–Q2), while the gap narrows at higher disagreement levels (Q3–Q4). These patterns provide descriptive context for the regression-based channel decomposition that follows.

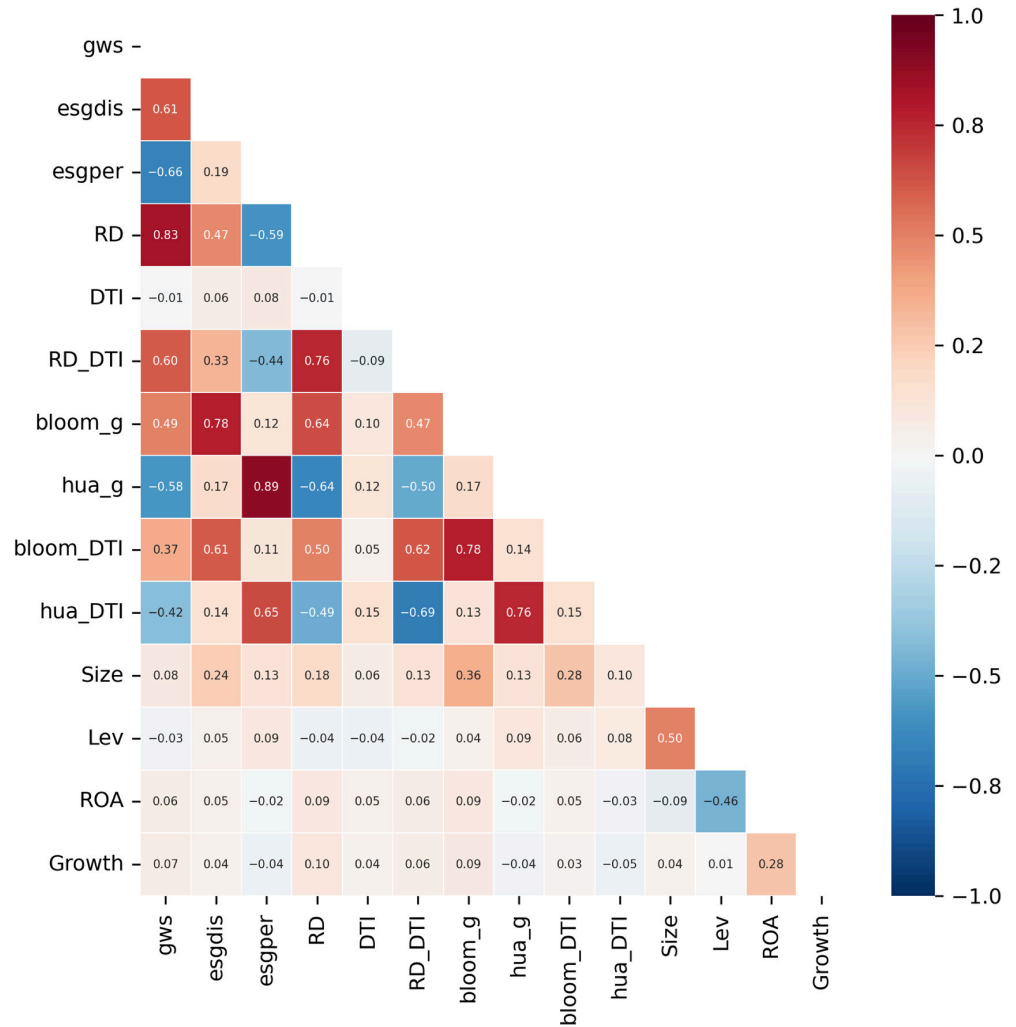


Figure 3. Correlation matrix for key variables.

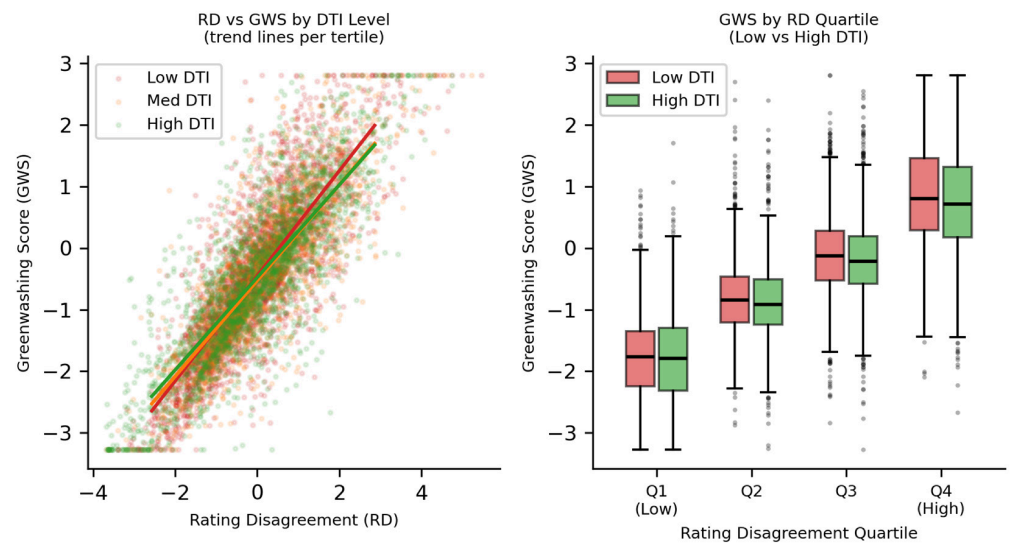


Figure 4. Rating disagreement and greenwashing.

4.2. Panel Regression Analysis

Table 2 reports panel regression results. Specification 2 achieves high within- R^2 across all dependent variables (esgdis: 0.303, esgper: 0.445, GWS: 0.737), confirming that aggregate rating disagreement (RD) is a powerful predictor of greenwashing. Specification 3 yields lower within- R^2 (esgdis: 0.302, esgper: 0.386, GWS: 0.318) but serves a distinct purpose: by replacing RD with its Bloomberg and HuaZheng interaction components, it isolates the asymmetric mechanisms through which DTI operates on each channel. The two specifications are therefore complementary: Spec-2 establishes predictive power, Spec-3 identifies the mechanism, as visualised in Figure 5. The high within- R^2 of Spec-2 (GWS: 0.737) partly reflects the shared algebraic construction of RD and GWS; it should therefore be read as a predictive benchmark motivating the channel decomposition in Spec-3, not as an autonomous causal estimate.

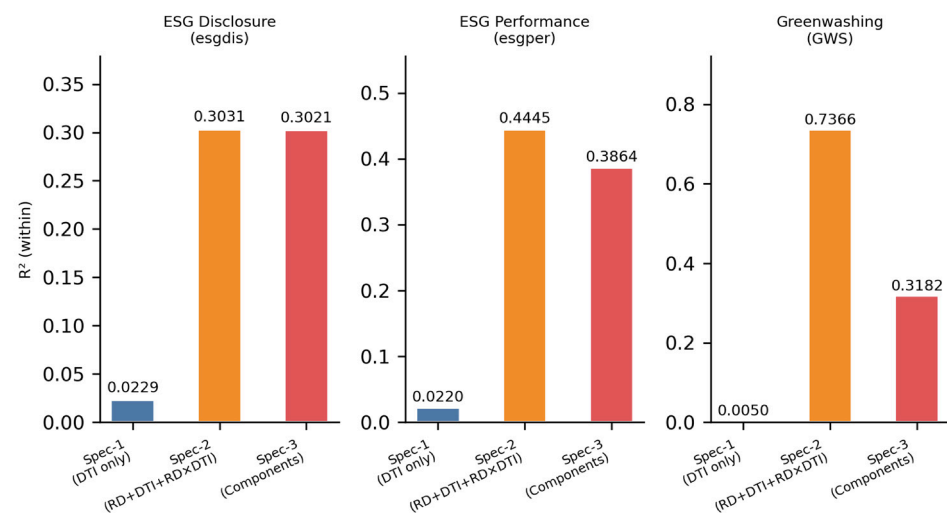


Figure 5. Panel regression within- R^2 by model specification.

Consistent with Hypotheses H1a and H1b, the double dissociation observed in Spec-3 provides direct evidence that Bloomberg and HuaZheng scores operate through distinct channels. In the esgdis equation, bloom_DTI is strongly positive ($\beta = +0.2474$, $p < 0.01$) while hua_DTI is insignificant, confirming that DTI primarily amplifies disclosure-oriented scoring. In the esgper equation, hua_DTI is strongly positive ($\beta = +0.2736$, $p < 0.01$) while bloom_DTI is insignificant, confirming that DTI primarily strengthens performance-oriented scoring. This pattern establishes that disclosure-oriented ratings (Bloomberg) and performance-oriented ratings (HuaZheng) respond to digital transformation through separate mechanisms rather than a common DTI effect, consistent with the predictions of H1a, H1b, and H1c regarding directionally asymmetric associations.

Consistent with Hypotheses H2a and H2b, the interaction terms in Spec-3 reveal that DTI moderates through two asymmetric channels. For the disclosure channel, bloom_DTI is positive and significant in the GWS equation ($\beta = +0.2471$, $p < 0.01$): firms with simultaneously high Bloomberg scores and advanced digital transformation exhibit greater greenwashing risk, consistent with digitally enabled disclosure capacity widening the gap between reported and actual ESG performance. For the performance channel, hua_DTI is negative and significant ($\beta = -0.2804$, $p < 0.01$): firms with simultaneously high HuaZheng scores and advanced digital transformation exhibit lower greenwashing risk, consistent with digital operational monitoring translating substantive ESG performance into measurable reductions in greenwashing. The opposing signs of bloom_DTI and hua_DTI establish the asymmetric channel mechanism as the structural driver of DTI's heterogeneous effect on greenwashing.

These asymmetric channel findings extend the aggregate inhibitory effect documented by Li, Miao, and Xu [9], revealing the distinct disclosure and performance mechanisms through which DTI reduces greenwashing rather than treating it as a uniform dampening effect. The control variable patterns offer interpretive insight: firm size is consistently positive across specifications, confirming that larger firms face greater stakeholder scrutiny and invest more in ESG signalling; leverage is positive, consistent with financially constrained firms engaging in symbolic disclosure to manage lender perceptions; Big-4 auditor presence is negative, indicating that audit quality disciplines greenwashing behaviour.

4.3. Robustness Tests

Lagged specifications ($t - 1$ predictors) preserve the sign and significance of all key coefficients: bloom_DTI_lag ($\beta = +0.0788, p < 0.01$) and hua_DTI_lag ($\beta = -0.0383, p < 0.01$) in the GWS equation, alleviating simultaneity concerns. VIF analysis confirms manageable multicollinearity for all control variables (VIF range: 1.14–2.26; DTI = 1.15). The main regressors (bloom_g, hua_g, RD, and their DTI interactions) yield infinite VIFs when entered jointly into a single design matrix, because $RD = bloom_g - hua_g$ generates perfect linear dependence by construction. This infinite VIF arises by construction from the nested specification design: Spec-2 and Spec-3 are estimated separately, each maintaining full column rank, and the pooled diagnostic matrix is the only context in which this dependence manifests. Multicollinearity, therefore, poses no threat to the validity of the reported estimates. The LPM binary GWS specification corroborates the main findings: bloom_DTI is positive and significant ($\beta = +0.0798, p < 0.01$), and hua_DTI is negative and significant ($\beta = -0.0980, p < 0.01$), consistent with the continuous GWS results and ruling out specification artefacts arising from the dependent variable construction.

Two-way clustered SE (firm \times year) and Driscoll-Kraay SE produce identical point estimates with marginally wider standard errors; bloom_DTI (DK: $t = +11.06, p < 0.001$) and hua_DTI (DK: $t = -28.49, p < 0.001$) remain significant under both approaches, confirming that cross-sectional dependence and heteroscedasticity do not inflate inference. Under the residualised specification, bloom_g_res ($\beta = +0.935, p < 0.001$) and hua_g_res ($\beta = -1.035, p < 0.001$) dominate the GWS equation as expected; bloom_DTI_res ($\beta = -0.045, p = 0.010$) and hua_DTI_res ($\beta = +0.039, p = 0.011$) are significant but sign-reversed relative to Spec-3, reflecting that the residualised interactions capture the DTI-specific variation in bloom_g and hua_g after orthogonalisation, a distinct economic quantity from the raw interaction terms. VIF values fall below 2.6 under the residualised specification, confirming that collinearity does not affect the main estimates. Uniform sector-year standardisation yields bloom_DTI_sy = +0.336 ($p < 0.001$) and hua_DTI_sy = -0.339 ($p < 0.001$), virtually identical in sign, magnitude, and significance to the main model, confirming that the mixed standardisation choice does not affect the conclusions. dig and DTI are correlated ($r = 0.707$) but not identical; the alternative DTI operationalisation using keyword density (dig) yields consistent Spec-2 results but attenuated and insignificant Spec-3 channel interactions, suggesting that the log-count DTI formulation captures disclosure-relevant variation more precisely than keyword density.

4.4. Heterogeneity Analysis

Figure 6 visualises the bloom_DTI and hua_DTI coefficients for the GWS equation across the three dimensions of heterogeneity. The asymmetric channel pattern, positive bloom_DTI and negative hua_DTI, is consistent across all subgroups, confirming that the disclosure and performance channel mechanisms are not confined to specific ownership types, pollution intensities, or geographic regions. Among ownership groups, bloom_DTI is positive and significant in both SOE ($\beta = +0.2500, p < 0.01$) and non-SOE firms ($\beta = +0.2430,$

$p < 0.01$), while hua_DTI is negative and significant in both SOE ($\beta = -0.2920, p < 0.01$) and non-SOE firms ($\beta = -0.2707, p < 0.01$). The magnitude differences between ownership groups are modest, and confidence intervals overlap substantially, indicating that the channel asymmetry operates similarly regardless of state ownership status. However, private firms face comparatively greater external pressure from institutional investors and lenders to signal ESG quality [14,15]. Among pollution-intensity groups, the same pattern holds: bloom_DTI is positive in both polluting ($\beta = +0.2401, p < 0.01$) and non-polluting firms ($\beta = +0.2577, p < 0.01$), while hua_DTI is negative in both groups (polluting: $\beta = -0.3121, p < 0.01$; non-polluting: $\beta = -0.2577, p < 0.01$). Among regional groups, eastern firms exhibit bloom_DTI = +0.2328 ($p < 0.01$) and central-western firms exhibit bloom_DTI = +0.2656 ($p < 0.01$), with correspondingly negative hua_DTI in both regions. Across all three dimensions, the primary finding is the robustness of the asymmetric channel mechanism rather than heterogeneity in its magnitude.

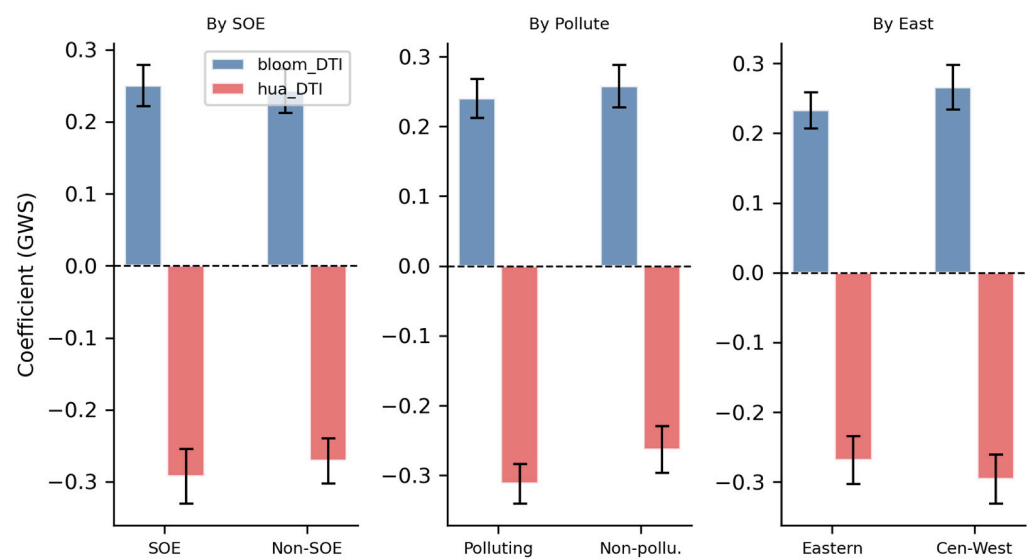


Figure 6. Heterogeneity analysis—bloom_DTI and hua_DTI coefficients (GWS) by subgroup.

4.5. Machine Learning Performance

Table 3 summarises 5-fold cross-validated performance metrics for all four Bayesian-optimised models. XGBoost achieves the highest regression R^2 (0.808 ± 0.004) and classification AUC-ROC (0.945 ± 0.005), followed closely by LightGBM. Random Forest and MLP trail modestly but remain strong predictors with AUC-ROC exceeding 0.93. The consistent superiority of gradient boosting over bagging confirms that sequential error correction captures structured nonlinearities in the greenwashing data-generating process. To establish the incremental predictive contribution of DTI and firm-level governance characteristics independent of rating construction, Table 4 reports a strict orthogonal ablation in which all rating-derived features are excluded. Under this orthogonal specification, XGBoost achieves $R^2 = 0.221$ and AUC = 0.713, confirming that firm size, leverage, audit quality, ownership concentration, and DTI together constitute meaningful and independent predictors of greenwashing risk. DTI ranks ninth in ablation SHAP importance (mean $|\text{SHAP}| = 0.078$), demonstrating that digital transformation retains predictive relevance even when the disclosure–performance wedge is entirely removed from the feature set.

Table 3. Machine learning model performance.

Model	R ² (GWS)	R ² _{std}	RMSE	RMSE _{std}	AUC-ROC	Accuracy
XGBoost	0.8076	(0.0043)	0.5380	(0.0058)	0.9452	0.8692
LightGBM	0.8070	(0.0084)	0.5388	(0.0101)	0.9434	0.8659
Random Forest	0.7459	(0.0030)	0.6184	(0.0079)	0.9328	0.8558
MLP	0.7709	(0.0091)	0.5869	(0.0057)	0.9324	0.8554

Table 4. Full model vs. strict orthogonal ablation—Performance comparison (5-fold CV).

Model	Full R ²	Ablation R ²	ΔR ²	Full AUC-ROC	Ablation AUC-ROC	ΔAUC-ROC
XGBoost	0.8076	0.2209	−0.5867	0.9452	0.7134	0.2318
LightGBM	0.8070	0.1982	−0.6088	0.9434	0.6952	0.2482
Random Forest	0.7459	0.1080	−0.6379	0.9328	0.6635	0.2693
MLP	0.7709	0.1217	−0.6492	0.9324	0.6475	0.2849

Ablation excludes bloom_g, hua_g, RD, bloom_DTI, hua_DTI, and RD_DTI—retaining only DTI and 13 firm-level controls (14 features)—same hyperparameters as full model.

4.6. SHAP Analysis

Figure 7 presents global SHAP importance rankings from the XGBoost model. Rating disagreement (RD) ranks first in mean absolute SHAP value, followed by bloom_g, hua_g, Size, and GDP, confirming that the disclosure–performance wedge and its components are the dominant drivers of greenwashing heterogeneity. Notably, the DTI interaction terms (bloom_DTI, RD_DTI) appear at lower rankings, indicating that their contribution operates through the rating-disagreement channel already captured by RD.

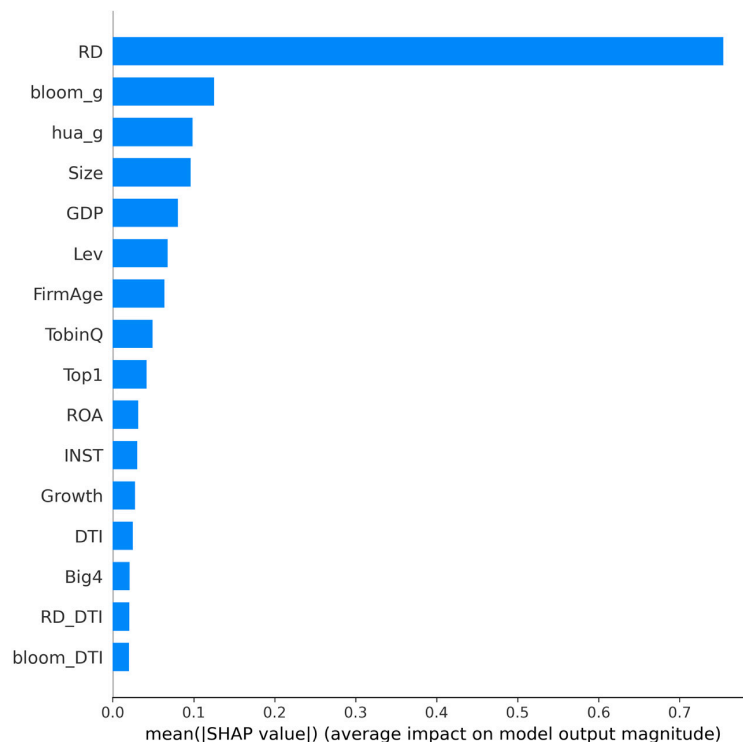


Figure 7. SHAP global feature importance (XGBoost—GWS regression).

Figure 8 (beeswarm plot) reveals the directional structure underlying the importance rankings: bloom_DTI exerts predominantly positive SHAP contributions (high values push GWS up), consistent with the disclosure-channel amplification effect. hua_DTI contributions are predominantly negative, consistent with the performance-channel attenuation

effect. Firm size and SOE status occupy the middle ranks, consistent with the heterogeneity findings. Figure 9 (dependence plots) reveals nonlinear dynamics in the key interaction terms: bloom_DTI SHAP values become increasingly positive at higher DTI levels, indicating that the disclosure-channel amplification effect strengthens as firms advance digitally. Similarly, hua_DTI SHAP contributions shift toward more negative values at high DTI, consistent with the performance channel attenuation deepening at advanced digitalisation levels. This concentration of effects among digitally advanced firms would be missed by linear models constrained to constant marginal effects.

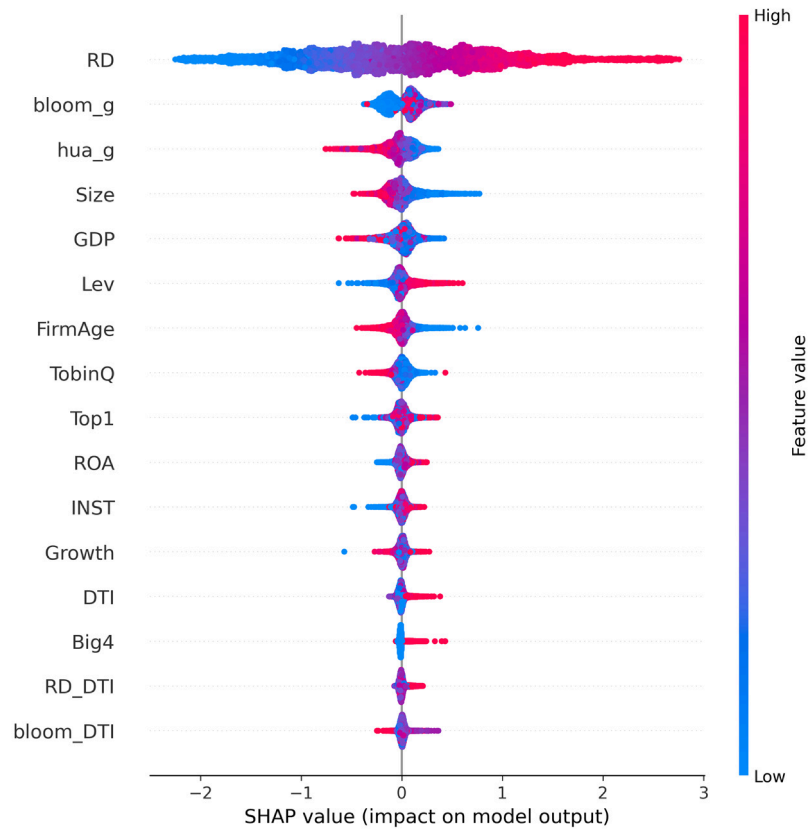


Figure 8. SHAP beeswarm plot (XGBoost—GWS regression).

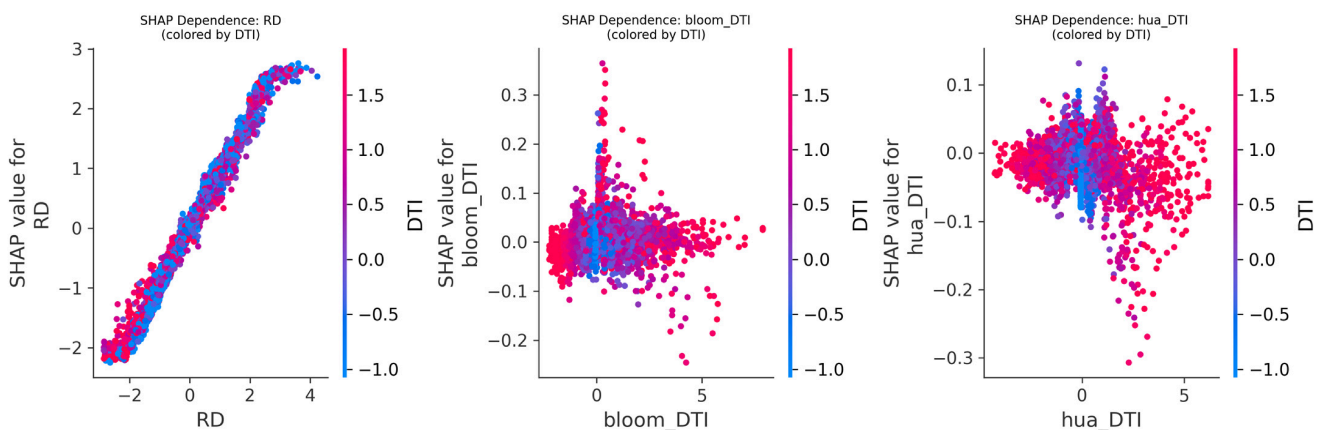


Figure 9. SHAP dependence plots.

Figure 10 (subgroup SHAP analysis) compares feature importance profiles between SOE and non-SOE firms. Bloom_DTI SHAP values are present across both ownership groups, with marginally larger magnitudes in non-SOE firms, consistent with the regression finding of a slightly stronger disclosure-channel effect among private firms. In contrast,

hua_DTI SHAP magnitudes are comparable across ownership types. This machine learning evidence corroborates and sharpens the regression-based heterogeneity finding, confirming that the disclosure-driven greenwashing response to digitalisation is not a regression artefact but a genuine feature of the data across ownership structures.

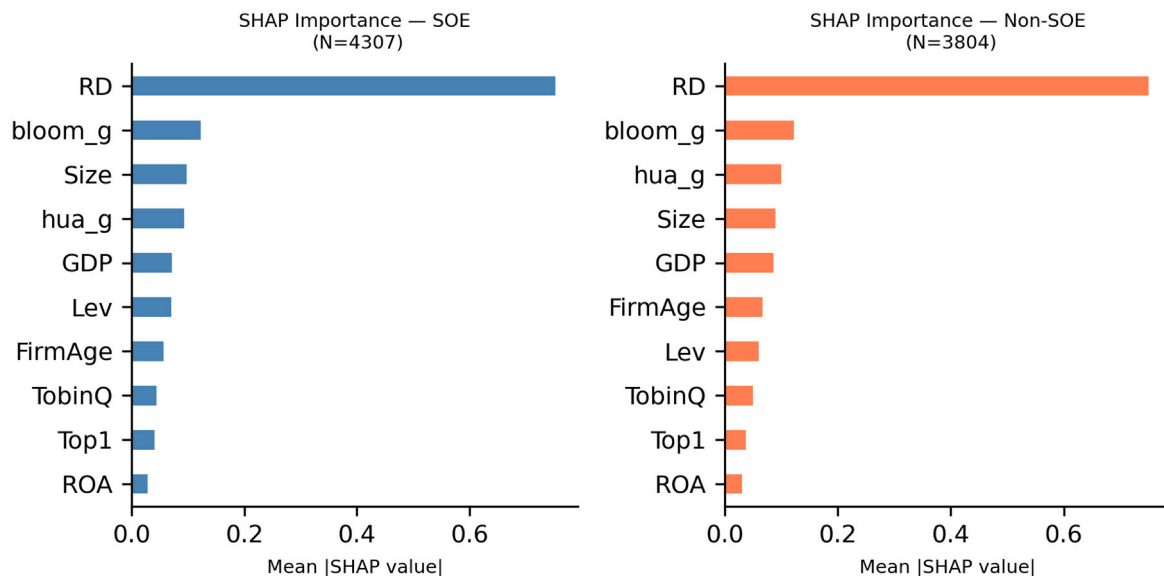


Figure 10. SHAP subgroup analysis.

5. Discussion

The asymmetric channel mechanism uncovered here has several implications. First, it challenges the common practice of treating ESG ratings as interchangeable signals. The Bloomberg–HuaZheng divergence is not random noise; it carries systematic information about a firm’s disclosure-versus-performance profile that predicts greenwashing risk at the firm-year level. Regulators seeking to combat greenwashing should attend not only to the level of ESG scores but to their cross-provider dispersion and directional composition.

Second, the finding that DTI operates through two asymmetric channels reveals a more nuanced picture than prior literature has captured. Li, Miao, and Xu [9] and Xu and Li [10] document net inhibitory effects of DTI on greenwashing; our channel decomposition reveals the distinct mechanisms underlying this aggregate finding. In the disclosure channel, bloom_DTI is positive and significant ($\beta = +0.2471, p < 0.01$), indicating that digitally advanced firms with high Bloomberg-type scores exhibit greater greenwashing risk: expanded digital reporting capacity widens the disclosure–performance gap by enabling more sophisticated ESG communication without commensurate substantive improvements [24]. In the performance channel, hua_DTI is negative and significant ($\beta = -0.2804, p < 0.01$), indicating that digitally advanced firms with high HuaZheng-type scores exhibit lower greenwashing risk: digital operational monitoring, IoT-based emissions tracking, and AI-assisted supply chain auditing translate substantive environmental performance more directly into measurable greenwashing reduction [19,21]. The net effect of DTI on greenwashing is therefore determined by the relative strength of these opposing channel forces at the firm level, rather than a uniform inhibitory mechanism as prior aggregate estimates suggest. In high-pressure institutional environments, digital reporting platforms may be leveraged for symbolic ESG signalling, such as ESG-washing or carbon washing, rather than genuine sustainability improvement [8,24], underscoring the need for regulators to monitor the alignment of ESG disclosures with independently verifiable performance metrics, particularly in sectors with high digital adoption [24]. Firms in early digitalisation stages may predominantly experience performance-channel benefits as digital monitoring systems

first improve substantive ESG outcomes, while firms at advanced digitalisation stages face growing disclosure-channel risk as their expanded reporting capacity increasingly outpaces substantive performance gains, a nonlinear dynamic confirmed by the SHAP dependence plots in Figure 9.

The alternative DTI operationalisation using keyword density (dig) yields attenuated and insignificant channel interaction terms in Spec-3, while Spec-2 results are broadly consistent. This divergence suggests that the log-count formulation of DTI captures threshold and scale dynamics in disclosure intensity more precisely than keyword density because the latter normalises the absolute volume of digital references. The finding underscores that DTI measurement choices matter for channel-level inference and warrants a cautious interpretation of studies that rely on alternative digitalisation proxies.

Third, the findings on heterogeneity have immediate policy relevance. The asymmetric channel pattern, positive bloom_DTI and negative hua_DTI, is consistent across SOE and non-SOE firms, polluting and non-polluting sectors, and eastern and central-western regions, with overlapping confidence intervals across subgroups. This cross-group robustness suggests that the disclosure-amplification and performance-attenuation mechanisms are structural features of digitalisation's interaction with ESG reporting rather than context-specific phenomena. Nevertheless, private firms face comparatively greater external pressure from institutional investors and lenders to signal ESG quality [14,15], and SOEs operate under direct government oversight that may partially substitute for market-based disciplinary mechanisms, differences that regulators should account for when designing DTI-aware ESG disclosure policy.

Fourth, the SHAP threshold analysis implies that linear models underestimate greenwashing risk for digitally advanced firms. Early-warning systems built on gradient-boosting architectures detect high-risk profiles that linear regression misses, with direct applications in ESG auditing, regulatory surveillance, and portfolio risk management. Importantly, because SHAP values satisfy the same additive decomposition property as linear regression coefficients, they permit econometrically informed interpretation: the machine learning predictions are not black-box artefacts but structured function approximations of the same underlying causal process.

6. Conclusions

This paper addresses whether ESG rating disagreement constitutes a systematic greenwashing signal and how digital transformation asymmetrically moderates this relationship through disclosure and performance channels. Using a panel dataset of 8111 Chinese firm-year observations and a hybrid econometric-machine learning framework, we establish three findings. First, aggregate rating disagreement (RD) is a strong predictor of greenwashing (Spec-2 within- $R^2 = 0.737$), confirming that the Bloomberg-HuaZheng divergence carries systematic information about firms' disclosure-performance decoupling. Second, channel decomposition reveals that DTI moderates greenwashing asymmetrically: the disclosure channel (bloom_DTI $\beta = +0.2471$, $p < 0.01$) amplifies greenwashing risk as digitally advanced firms leverage expanded reporting capacity to widen the disclosure-performance gap, while the performance channel (hua_DTI $\beta = -0.2804$, $p < 0.01$) attenuates greenwashing risk as digital operational monitoring translates substantive ESG performance into measurable reduction. Third, the asymmetric channel pattern is robust across ownership structure, pollution intensity, and geographic region, indicating that it reflects a structural feature of digitalisation's interaction with ESG reporting rather than a subgroup-specific phenomenon.

The machine learning analysis, combining four Bayesian-optimised models (best XGBoost: $R^2 = 0.808$, AUC-ROC = 0.945) with SHAP interpretability, confirms the econo-

metric findings, reveals nonlinear thresholds invisible to panel regression, and identifies rating disagreement and its components as the dominant predictors of cross-firm greenwashing heterogeneity. Together, these results underscore that ESG rating disagreement contains actionable information for regulators and investors, and that the impact of digital transformation on greenwashing is more nuanced and channel-dependent than aggregate estimates suggest.

The study makes three scientific contributions to the literature. First, it formalises the Bloomberg–HuaZheng divergence as a directional greenwashing signal, extending the ESG rating disagreement literature [1,7,8] by demonstrating that the direction, rather than merely the magnitude, of disagreement carries predictive power. Second, it decomposes DTI's aggregate inhibitory effect [9,10] into competing channel mechanisms, providing a theoretically motivated explanation for heterogeneous firm-level responses to digitalisation. Third, it demonstrates methodological complementarity between two-way fixed-effects econometrics and gradient-boosting ML with SHAP, showing that nonlinear threshold effects identified by machine learning are consistent with, and extend, the linear channel decomposition.

Limitations and Future Research

The analysis is subject to several limitations. The sample is confined to Chinese A-share firms (2012–2022), and the Bloomberg–HuaZheng pairing reflects a specific disclosure-versus-performance contrast that may not generalise to other market settings or combinations of rating providers. The greenwashing proxy is constructed from rating divergence rather than verified misconduct records, introducing measurement imprecision; external validation against MEE environmental penalties or media-verified ESG controversies was not possible due to data availability constraints, and construct validity therefore rests on the theoretical and empirical consensus established by Hu, Hua, Liu, and Wang [7] and Zhang, Zheng, and Meng [8]. Alternative ESG rating pairs (Wind, Refinitiv, MSCI) were unavailable, precluding cross-provider replication. The DTI index relies on textual keyword frequency, which may not fully capture the quality or strategic orientation of digitalisation. Causal identification rests on two-way fixed effects and lagged specifications rather than a natural experiment; both a natural experiment with an exogenous disclosure shock and a quasi-experimental design exploiting exogenous variation in digital policy exposure, such as broadband rollout or Digital Economy Pilot Zone designation, would further strengthen causal claims. Future research should replicate the channel decomposition with alternative provider pairings, extend the framework to non-Chinese markets, and explore whether the asymmetric channel mechanism generalises across different institutional and regulatory environments.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18136800/s1>, Table S1: Optimized hyperparameters from bayesian search (Optuna TPE, 50 trials, 5-fold CV, seed = 42).

Author Contributions: Conceptualization, I.Ö. and Ü.Y.; methodology, Ü.Y.; software, Ü.Y.; validation, I.Ö. and Ü.Y.; formal analysis, Ü.Y.; investigation, I.Ö. and Ü.Y.; resources, I.Ö.; data curation, Ü.Y.; writing—original draft preparation, I.Ö. and Ü.Y.; writing—review and editing, I.Ö. and Ü.Y.; visualization, Ü.Y.; supervision, I.Ö.; project administration, I.Ö. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is publicly available on Mendeley Data at doi: 10.17632/vst753v2m6.1.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Berg, F.; Kölbel, J.F.; Rigobon, R. Aggregate Confusion: The Divergence of ESG Ratings. *Rev. Financ.* **2022**, *26*, 1315–1344. [[CrossRef](#)]
2. He, R.; Chen, H.; Zhu, X. Corporate hypocrisy and ESG rating divergence. *Corp. Soc. Responsib. Environ. Manag.* **2025**, *32*, 1122–1146. [[CrossRef](#)]
3. Mio, C.; Fasan, M.; Scarpa, F.; Costantini, A.; Fitzpatrick, A.C. Unveiling the consequences of ESG rating disagreement: An empirical analysis of the impact on the cost of equity capital. *Account. Eur.* **2025**, *22*, 177–200. [[CrossRef](#)]
4. Li, Y. ESG rating disagreements and stock price crash risk: Can distraction, regional digital economy and corporate intelligence maturity play important roles? *Asian J. Account. Res.* **2024**, *10*, 126–149. [[CrossRef](#)]
5. Li, J.; Ye, Y.; Li, J. Corporate behavior under confusion: ESG disagreement and green innovation. *Asia-Pac. J. Account. Econ.* **2026**, *33*, 208–227. [[CrossRef](#)]
6. Wang, Y.; Wang, Y.; Chen, Z.; Shen, H. The flip side of ESG: Do ESG rating discrepancies undermine corporate energy efficiency?—A comprehensive assessment using double machine learning. *Appl. Econ. Lett.* **2026**, *33*, 1683–1693. [[CrossRef](#)]
7. Hu, X.; Hua, R.; Liu, Q.; Wang, C. The green fog: Environmental rating disagreement and corporate greenwashing. *Pac.-Basin Financ. J.* **2023**, *78*, 101952. [[CrossRef](#)]
8. Zhang, Z.; Zheng, X.; Meng, X. Corporate ESG-washing strategies responding to external rating divergence: Moderating effects of normative institutional pressures. *Humanit. Soc. Sci. Commun.* **2026**, *13*, 927. [[CrossRef](#)]
9. Li, Z.; Miao, S.; Xu, L. Digital transformation and environmental, social, and governance greenwashing: Evidence from China. *J. Environ. Manag.* **2024**, *365*, 121460. [[CrossRef](#)] [[PubMed](#)]
10. Xu, J.; Li, R. How digital transformation curb greenwashing: Insights from fraud risk factor theory. *PLoS ONE* **2026**, *21*, e0339282. [[CrossRef](#)] [[PubMed](#)]
11. Zhao, X. Inhibition Mechanisms of Corporate Greenwashing: Evidence from Chinese A-Share Listed Companies during Digital Transformation. *Mendeley Data* **2025**, *V1*. [[CrossRef](#)]
12. Zhu, P.; Liu, J. How ESG Rating Divergence Undermines Financial Flexibility and Sustainable Resilience. *Sustainability* **2025**, *17*, 9607. [[CrossRef](#)]
13. Deng, J.; Lin, W.; Huang, J.; Cai, Y.; Wang, W. Research on the impact of the ESG rating divergence of manufacturing firms on carbon emission intensity. *PLoS ONE* **2025**, *20*, e0323929. [[CrossRef](#)] [[PubMed](#)]
14. Qin, J.; Wang, M. ESG rating disagreement and bank loan availability: Evidence from China. *PLoS ONE* **2025**, *20*, e0317191. [[CrossRef](#)] [[PubMed](#)]
15. Liu, L.; Wang, H.; Chen, S.; Chen, Z. Has the divergence in ESG ratings affected institutional shareholding? *Financ. Res. Lett.* **2025**, *83*, 107678. [[CrossRef](#)]
16. Li, G.; Cheng, Y. Impact of environmental, social, and governance rating disagreement on real earnings management in Chinese listed companies. *Glob. Financ. J.* **2024**, *62*, 101015. [[CrossRef](#)]
17. Zhu, C.; Yang, C. Divergences among ESG rating systems: Evidence from financial indexes. *Econ. Anal. Lett.* **2024**, *3*, 46–51. [[CrossRef](#)]
18. Liu, M. Quantitative ESG disclosure and divergence of ESG ratings. *Front. Psychol.* **2022**, *13*, 936798. [[CrossRef](#)] [[PubMed](#)]
19. Liu, J.-Y.; Lei, Q.; Zhang, Y.-J. Unveiling the green truth: The impact of digital transformation on the decoupling of corporate environmental responsibility. *J. Ind. Ecol.* **2025**, *29*, 2387–2403. [[CrossRef](#)]
20. Wang, S.; Lai, Y.; Zhang, S. Greening by digitization? Exploring the effect of enterprise digital transformation on greenwashing. *Bus. Strategy Environ.* **2024**, *33*, 6616–6639. [[CrossRef](#)]
21. Sun, Z.; Sun, X.; Wang, W.; Wang, W. Digital transformation and greenwashing in environmental, social, and governance disclosure: Does investor attention matter? *Bus. Ethics Environ. Responsib.* **2025**, *34*, 81–102. [[CrossRef](#)]
22. Guo, X.; Pang, W. The impact of digital transformation on corporate ESG performance. *Financ. Res. Lett.* **2025**, *72*, 106518. [[CrossRef](#)]
23. Ren, H. ESG rating disagreement and corporate digital transformation. *Financ. Res. Lett.* **2025**, *75*, 106903. [[CrossRef](#)]
24. Wang, Y.; Lyu, C. The Dark Side of the Carbon Emissions Trading System and Digital Transformation: Corporate Carbon Washing. *Systems* **2025**, *13*, 619. [[CrossRef](#)]
25. Pu, T. E, S, and G, not just ESG: Exploring the Digital Transformation–Environmental, Social, and Governance Performance Nexus through Patent-Based Evidence. *Sage Open* **2025**, *15*, 21582440251384402. [[CrossRef](#)]

26. Zhang, C.; Wang, Y. Are ESG scores affected by the quality of non-financial disclosures? *Int. Rev. Econ. Financ.* **2024**, *93*, 1431–1458. [[CrossRef](#)]
27. Zeng, F.; Wang, J.; Zeng, C. An optimized machine learning framework for predicting and interpreting corporate ESG greenwashing behavior. *PLoS ONE* **2025**, *20*, e0316287. [[CrossRef](#)] [[PubMed](#)]
28. Jianfeng, Z.; Tiantian, Q. Interpretable predictive model for listed companies ESG greenwashing based on XGBoost and SHAP. *Sci. Rep.* **2026**, *16*, 12899. [[CrossRef](#)] [[PubMed](#)]
29. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, CA, USA; Curran Associates, Inc.: Red Hook, NY, USA, 2017. Available online: <https://dl.acm.org/doi/10.5555/3295222.3295230> (accessed on 30 May 2026).
30. Angrist, J.D.; Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*; Princeton University Press: Princeton, NJ, USA, 2009. [[CrossRef](#)]
31. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*; The MIT Press: Cambridge, MA, USA, 2010. Available online: <http://www.jstor.org/stable/j.ctt5hhcfr> (accessed on 30 May 2026).
32. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
33. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 3149–3157. Available online: <https://dl.acm.org/doi/10.5555/3294996.3295074> (accessed on 30 May 2026).
34. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
36. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 2623–2631. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.