

Original Research

Reputation Consequences of Strategic Metric Opacity Evidence from Customer KPI Discontinuations and Capital Market Reactions

Mark Anthony Cruz¹, John Paul Reyes² and Kevin Miguel Santos³

¹Nueva Vizcaya State University, College of Business Education, Bayombong Campus, Don Mariano Marcos Street, Bayombong 3700, Nueva Vizcaya, Philippines.

²Western Mindanao State University, College of Business Administration, Baliwasan Chico, Normal Road, Zamboanga City 7000, Philippines.

³Bicol University, College of Business, Economics and Management, Rizal Street, Legazpi City 4500, Albay, Philippines.

Abstract

Public firms increasingly disclose customer, platform, safety, service, and community metrics that help stakeholders evaluate the condition of intangible market assets. These measures are not always mandatory, and firms can later redefine, aggregate, rename, or discontinue them. This paper examines whether sudden opacity in voluntarily disclosed stakeholder metrics functions as a negative corporate signal. The empirical sample contains 1,904 metric-opacity events among 486 U.S.-listed firms from 2012 to 2024. Events include discontinued customer counts, reduced churn disclosure, aggregated user metrics, removed service-quality indicators, suppressed complaint measures, discontinued safety statistics, and altered community-impact indicators. The data combine quarterly filings, earnings-call transcripts, investor presentations, analyst questions, option prices, equity returns, product-review sentiment, customer-search behavior, employee text, and institutional trading. A transformer-based disclosure model identifies expected metric continuation and flags unexpected opacity when a historically recurring stakeholder metric disappears or becomes less diagnostic. The results show that unexpected metric opacity is associated with a three-day abnormal return of minus 0.58%, a 2.74% rise in thirty-day implied volatility, and a 17.9% increase in analyst questions about operating quality. The penalty is largest when omitted metrics previously tracked customer retention, complaint rates, service reliability, or user engagement. Firms with strong pre-event reputation and stable metric histories experience sharper reactions, consistent with expectation violation. Machine-learning estimates show that disclosure history, metric diagnosticity, and stakeholder sentiment deterioration predict market penalties more accurately than firm size or earnings surprise. The evidence indicates that withholding previously informative stakeholder metrics can damage reputation by increasing uncertainty about customer relationships and managerial credibility.

1. Introduction

Corporate signaling is often studied through what firms choose to say. Equally important is what firms stop saying. Public firms voluntarily disclose many recurring metrics that help investors, analysts, customers, employees, and other stakeholders evaluate intangible market assets [1]. A streaming firm reports subscriber additions. A retailer reports active customers or loyalty membership. A platform reports monthly users. A service firm reports churn, retention, claims, complaints, or satisfaction. A manufacturer reports safety incidents, warranty experience, quality rates, or delivery reliability. A logistics firm reports on-time performance. A healthcare or education service provider may report completion, satisfaction, or service-access measures. Once such metrics become recurring, stakeholders form expectations that the firm will continue disclosing them unless the underlying economics or communication strategy changes.

This paper studies the reputational and market consequences of sudden opacity in stakeholder-related metrics. Metric opacity occurs when a firm discontinues, aggregates, renames, narrows, broadens, or materially weakens a previously recurring stakeholder metric in a way that reduces outside diagnosticity. The event is not ordinary nondisclosure by a firm that never reported the measure. It is a change in the information regime after the market has learned to use the metric. Such a change may be harmless if the metric has become obsolete or is replaced by a better measure. It may be damaging if stakeholders infer that management is concealing deterioration in customer retention, user engagement, service quality, safety, satisfaction, or complaint exposure.

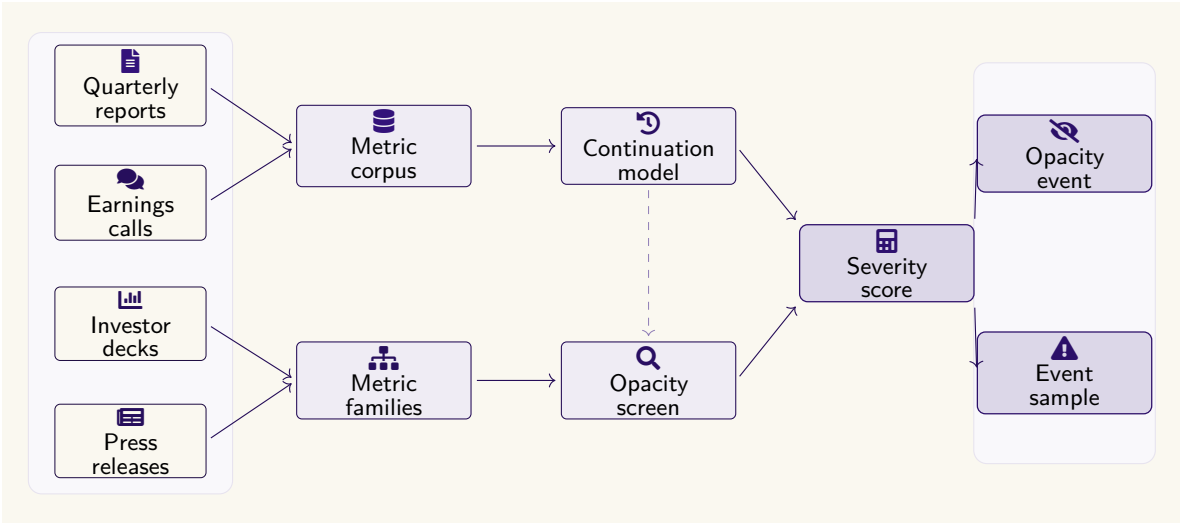


Figure 1: Disclosure-processing pipeline for detecting unexpected discontinuation, aggregation, or weakening of recurring stakeholder metrics.

The central empirical question is whether the disappearance or weakening of recurring stakeholder metrics becomes a corporate signal that affects reputation and market valuation. The setting is distinct from standard earnings guidance, product announcements, privacy controversies, labor events, supplier misconduct, endorser scandals, or online review recovery. The signal is an alteration in the firm’s disclosure architecture. It is subtle, often embedded in earnings-call slides or quarterly filings, and sometimes not explicitly announced. Its informational force comes from discontinuity. A firm that consistently reports a customer metric teaches stakeholders to treat the metric as useful. When the metric disappears during a period of uncertainty, silence becomes informative.

The study uses a panel of 486 U.S.-listed firms from 2012 to 2024. The data include 1,904 metric-opacity events identified from quarterly filings, annual reports, earnings-call transcripts, investor-day presentations, and supplemental investor decks. Events are classified into customer-base opacity, engagement opacity, retention or churn opacity, service-quality opacity, safety and reliability opacity, complaint or claims opacity, community or stakeholder-impact opacity, and metric aggregation. The empirical design links these events to equity returns, option-implied volatility, analyst questions, analyst forecast revisions, institutional trading, customer sentiment, product-review language, employee text, and search behavior.

The paper treats metric opacity as a reputation-relevant signal because recurring disclosure is part of managerial credibility. When firms voluntarily report a stakeholder metric over many periods, they create a repeated communication contract with outside audiences. Discontinuation can raise doubts about the metric itself, the underlying stakeholder relationship, and management’s willingness to communicate unfavorable information. The effect should be strongest when the metric was highly diagnostic, when

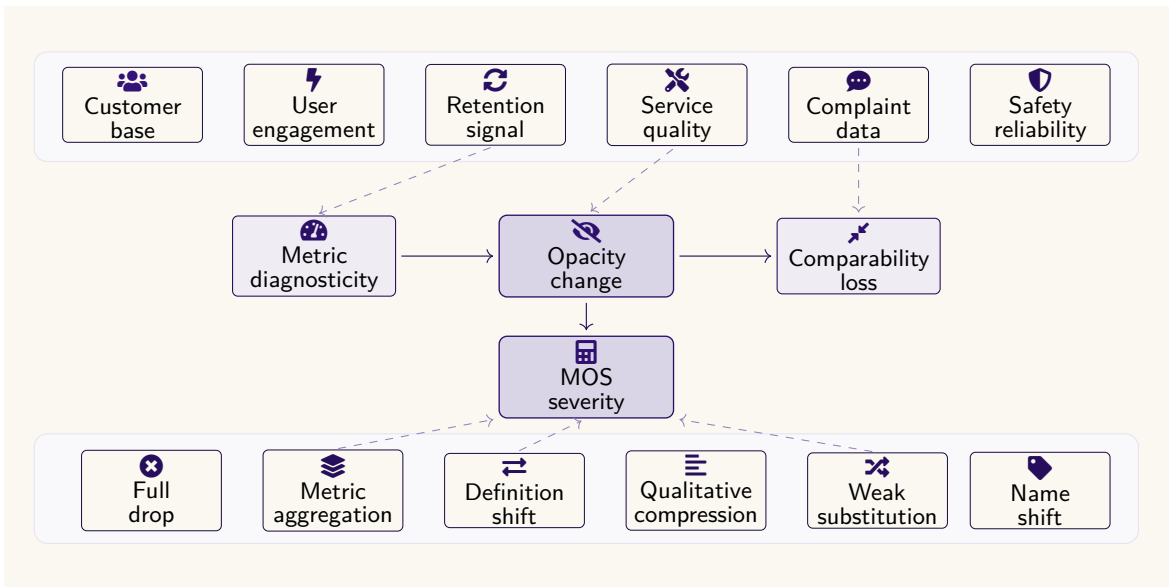


Figure 2: Metric-opacity taxonomy showing how diagnosticity, comparability loss, and opacity form combine into event severity.

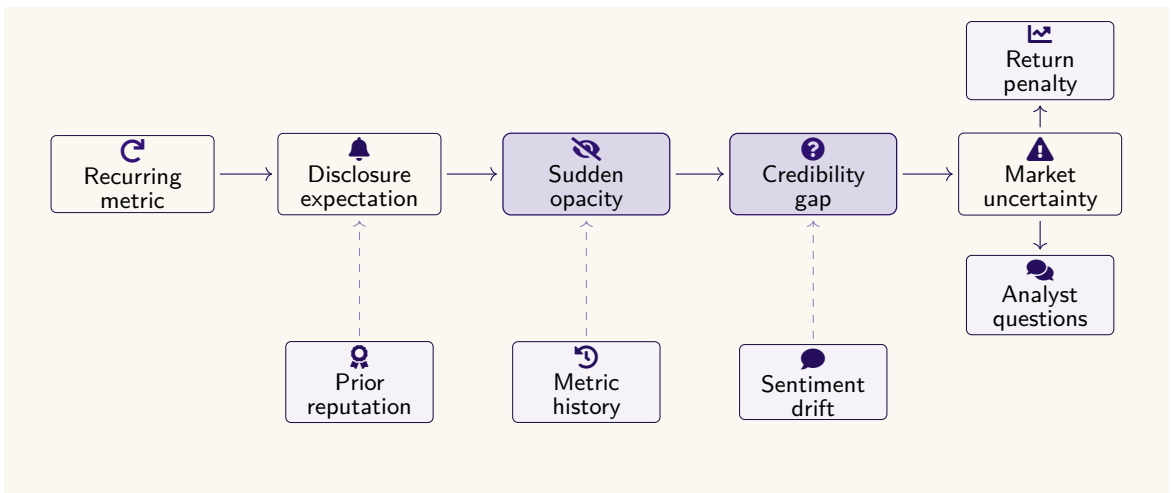


Figure 3: Reputation mechanism through which discontinued stakeholder metrics trigger expectation violation, credibility concerns, and market uncertainty.

the metric had a stable disclosure history, when the omitted metric concerned a reputationally sensitive stakeholder relationship, and when pre-event stakeholder sentiment was already weakening.

Cao and Xu (2022) [2] show that managerial decisions around brand transactions can affect abnormal returns through valuation-related mechanisms, indicating that investors attend to managerial discretion when intangible assets are difficult to value. In the present study, the discretion is not the acquisition or sale of a brand asset but the choice to reduce transparency about stakeholder metrics that inform the valuation of customer relationships, service reliability, and reputation.

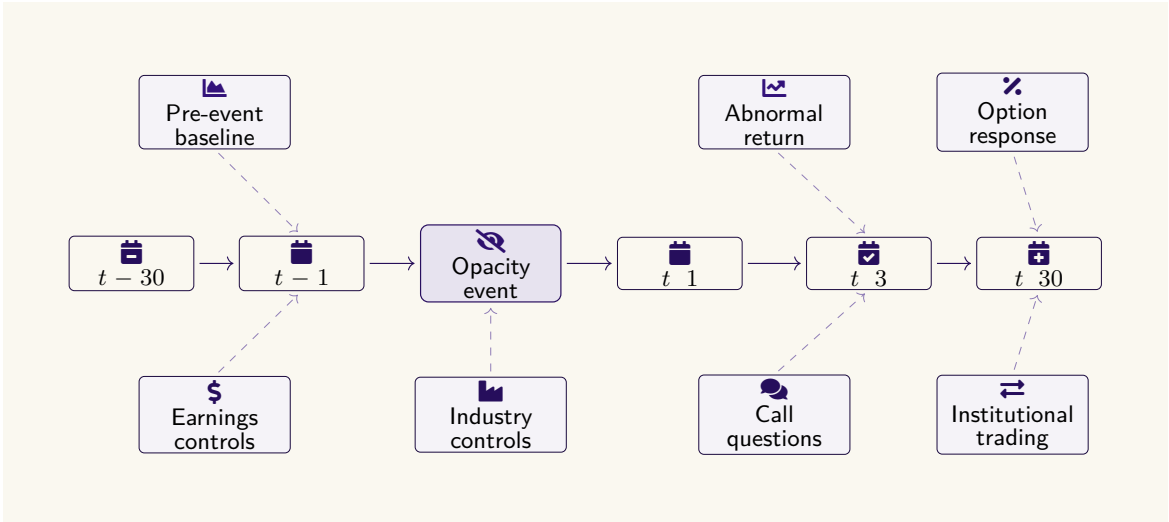


Figure 4: Event-study structure for measuring price, volatility, analyst, and trading reactions around metric-opacity events.

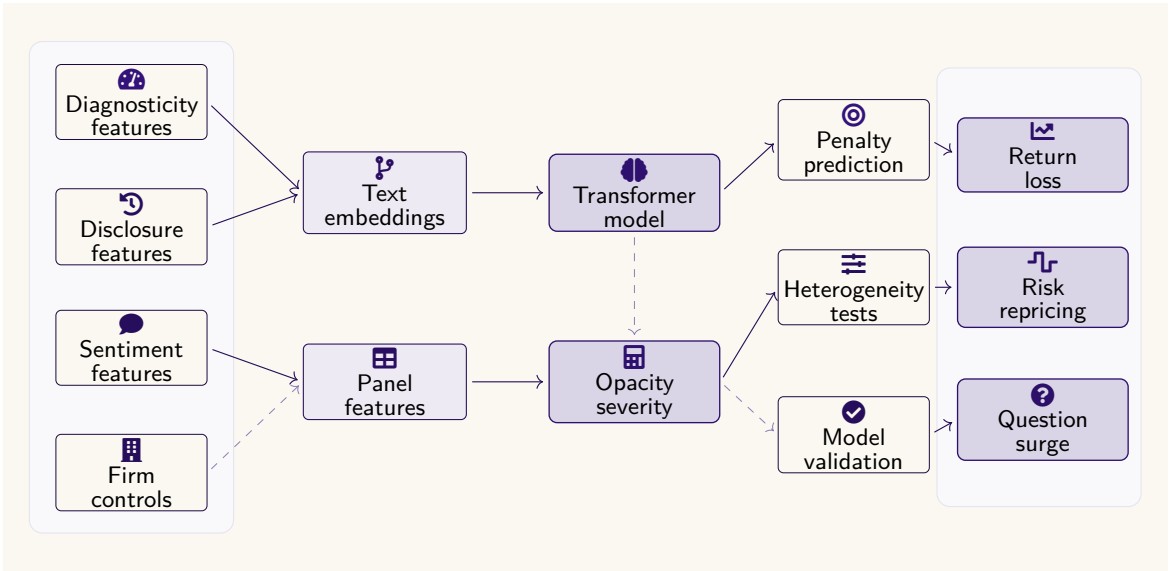


Figure 5: Prediction architecture combining disclosure history, metric diagnosticity, sentiment signals, and firm controls to estimate opacity penalties.

The empirical results show that unexpected metric opacity produces negative market and stakeholder reactions [3]. The mean three-day abnormal return around an opacity event is minus 0.58%. Thirty-day implied volatility rises by 2.74%, and put skew rises by 1.86%. Analysts increase questions about operating quality, customer retention, and metric comparability by 17.9% in the first earnings call after the event. Analyst forecast revisions are concentrated in revenue durability and margin risk rather than in immediate earnings. The results are stronger when the omitted metric concerns churn, active users, complaint rates, reliability, or claims. They are weaker when the firm replaces the discontinued metric with a more diagnostic measure.

The market penalty is not simply a reaction to poor performance. The models control for earnings surprises, revenue surprises, management guidance, industry performance, analyst coverage, prior returns, volatility, and firm fixed effects. The opacity coefficient remains negative. Moreover, firms that explicitly explain why a metric is being replaced and provide a reconciled historical bridge experience smaller penalties. Firms that discontinue a metric without explanation experience the largest increase in analyst questioning and implied volatility. This pattern is consistent with an uncertainty and credibility mechanism.

The findings also show that reputation conditions shape reaction [4]. Firms with strong prior reputation and long metric-disclosure histories experience sharper penalties when they remove diagnostic metrics. This result may seem counterintuitive because strong reputation often protects firms [5]. In this setting, strong reputation creates expectations of transparency. Sudden opacity violates those expectations. By contrast, firms with weak or inconsistent disclosure histories receive smaller incremental penalties because stakeholders had already discounted their metric credibility.

The study contributes to empirical work on corporate signals, reputation, and stakeholder-market reactions in several ways. First, it identifies metric opacity as a corporate signal generated by silence and reduced comparability rather than by positive disclosure. Second, it links disclosure changes about stakeholder relationships to both financial markets and stakeholder behavior. Third, it uses machine-learning classification of expected metric continuation to distinguish ordinary reporting variation from unexpected opacity. Fourth, it shows that the reputational meaning of disclosure change depends on metric diagnosticity and prior credibility.

The paper remains bounded in interpretation. Not all metric discontinuation is opportunistic. Some measures become obsolete when business models change. Some firms improve disclosure by replacing low-quality metrics with better measures [6]. The empirical design separates unexplained opacity from explained metric replacement. The strongest negative reactions appear when firms remove diagnostic stakeholder metrics without credible explanation and when other signals suggest stakeholder deterioration. The evidence therefore supports a conditional view: metric opacity damages reputation when it increases uncertainty about hidden stakeholder conditions and managerial candor.

2. Data and Sample Formation

The sample is constructed from public firms whose stakeholder metrics are observable in recurring disclosures. The data source includes quarterly reports, annual reports, earnings-call transcripts, investor-day presentations, earnings supplements, press releases, and archived investor decks. A firm enters the sample when it reports at least one stakeholder-related metric in at least six consecutive quarters or in at least eight of ten consecutive quarters. This rule is used because the study focuses on expected disclosure continuity. A metric that appears only once is not part of a repeated information regime.

The final panel contains 486 firms from 2012 to 2024. The firms operate in digital platforms, retail, telecommunications, media, transportation, hospitality, insurance, consumer finance, healthcare services, education services, software, consumer products, logistics, industrial services, and manufacturing. The industries differ in which stakeholder metrics are most relevant. Digital firms report active users, paid users, subscribers, engagement, retention, and average revenue per user. Retailers report active customers, loyalty members, repeat purchase, traffic, and satisfaction. Service firms report churn, claims, complaint rates, quality scores, and retention. Manufacturers report warranty indicators, safety, quality, delivery, and product reliability.

The raw disclosure corpus contains 4.2 million firm-quarter document sections and 86.7 million sentences. A transformer-based metric extraction model identifies quantitative stakeholder metrics and classifies them by topic, unit, frequency, and comparability. Metrics are standardized into metric families so that minor wording changes do not create false discontinuations. For example, “monthly active users,” “MAUs,” and “average monthly active accounts” may belong to one family if the denominator and measurement logic remain comparable. A metric is considered diagnostically continuous when its definition, scale, and stakeholder interpretation remain stable [7].

A metric-opacity event occurs when a historically recurring metric becomes unexpectedly absent, aggregated, redefined, or less comparable. The event must reduce diagnosticity relative to the prior disclosure pattern. The study identifies six main forms of opacity. Discontinuation occurs when the metric disappears entirely. Aggregation occurs when a specific metric is folded into a broader category. Redefinition occurs when the denominator or measurement window changes without full reconciliation. Renaming opacity occurs when the firm changes terminology enough to reduce comparability. Compression occurs when a continuous metric becomes a qualitative statement [8]. Substitution opacity occurs when a firm replaces a diagnostic metric with a less diagnostic measure.

The classification model assigns a diagnosticity score to each metric. Diagnosticity reflects how directly the metric informs stakeholder relationship quality, future revenue, operating risk, or reputation. A churn rate is highly diagnostic because it reveals retention. A general statement that customer engagement remains healthy is low diagnosticity. A complaint-rate metric is highly diagnostic for service reputation. A vague customer-experience award is less diagnostic. A delivery reliability metric is highly diagnostic in logistics. A broad operating-efficiency claim is less diagnostic. The opacity event is weighted more heavily when the omitted metric was more diagnostic.

The diagnosticity score is:

$$MD_{imt} = \alpha_1 PredictiveUse_{imt} \alpha_2 StakeholderLink_{imt} \alpha_3 Comparability_{imt} \alpha_4 Frequency_{imt} \alpha_5 AnalystUse_{imt}. \quad (2.1)$$

Here, i indexes the firm, m the metric family, and t the quarter. Predictive use is estimated from historical association with revenue growth, margin stability, or risk. Stakeholder link measures whether the metric directly concerns customers, users, employees, product quality, safety, or complaints. Comparability measures definition stability. Frequency measures recurrence. Analyst use measures how often analysts ask about the metric or mention it in reports. The score is standardized within industry [9].

The opacity severity score is:

$$MOS_{imt} = MD_{im,t-1} \times Unexpected_{imt} \times LossComp_{imt} \times Silence_{imt}. \quad (2.2)$$

Unexpectedness captures whether the metric was expected to appear based on prior reporting frequency, seasonality, and firm disclosure practice. Loss of comparability captures how much information is lost. Silence captures whether the firm provides no explanation or reconciliation. If a metric disappears but the firm provides a clearly superior replacement with historical backfill, the severity score is low. If a high-diagnosticity metric disappears without explanation, the severity score is high.

The main event sample contains 1,904 metric-opacity events. Customer-base metrics account for 21.6% of events, engagement metrics for 18.4%, retention or churn metrics for 15.7%, service-quality metrics for 12.9%, complaint or claims metrics for 10.8%, safety or reliability metrics for 9.5%, community or stakeholder-impact metrics for 6.4%, and other metric aggregation events for 4.7%. About 43.1% of events are full discontinuations, 24.8% are aggregations, 14.6% are redefinitions, 9.2% are qualitative compressions, and 8.3% are weaker substitutions.

Cao (2022) [10] documents that brand equity can reduce warranty claim rates and abnormal warranty accrual rates, which shows that intangible market assets can affect firm value through cost and risk channels. This motivates the present paper's treatment of complaint, quality, reliability, and safety metrics as reputation-relevant signals rather than as peripheral operational details.

The market data include daily stock returns, intraday trading records for a subset of events, option-implied volatility, put skew, short interest, institutional ownership, analyst forecasts, and analyst reports. The primary market outcome is the three-day cumulative abnormal return around the first disclosure in which opacity becomes observable. Because opacity is often revealed during earnings reporting, the

design controls carefully for earnings surprises and guidance changes. Events that occur in the same document as large unrelated corporate news are excluded from the primary market sample.

The stakeholder data include product-review sentiment, customer social text, search behavior, app usage for digital firms, customer complaint portals, employee-review text, and media discussion. Stakeholder measures are observed weekly from eight weeks before to sixteen weeks after the opacity event. These data help determine whether metric opacity coincides with, follows, or predicts deterioration in stakeholder sentiment. The study does not assume that stakeholder deterioration is always caused by opacity. It uses stakeholder signals to evaluate whether opacity is interpreted as hiding an unfavorable trend.

The analyst data are central because analysts often monitor voluntary metrics. The corpus includes earnings-call questions, analyst reports, forecast revisions, target-price changes, and textual concern about metric comparability. Analyst questions are coded into metric clarification, revenue durability, customer quality, churn, engagement, service quality, safety, and disclosure credibility. A question is considered opacity-related if it asks why a metric disappeared, whether a new metric is comparable, whether customer quality has changed, or whether management will provide historical reconciliation.

The final dataset is a firm-quarter-event panel with market windows and stakeholder windows. Each firm can experience multiple opacity events, but overlapping events in the same quarter are grouped when they concern the same metric family. If a firm removes several unrelated stakeholder metrics in the same disclosure, the event is coded as a multi-metric opacity episode. Multi-metric episodes represent 16.2% of the sample and produce stronger reactions in the results.

3. Disclosure Measurement and Signal Design

The measurement challenge is that firms naturally change metrics over time. A useful empirical design must separate reasonable metric evolution from opacity. The paper therefore models expected metric continuation before identifying an event [11]. For each firm-metric family, the model predicts the probability that the metric should appear in the current disclosure based on prior recurrence, seasonality, business segment, reporting calendar, management discussion topics, and industry practice. An omission is treated as unexpected only when predicted continuation probability is high.

The expected continuation model is:

$$PReport_{imt} = 1 = \Lambda \delta_0 \delta_1 Report_{im,t-1} \delta_2 Streak_{im,t-1} \delta_3 Season_{mt} \delta_4 SegmentChange_{it} \delta_5 IndustryNorm_{mt} \delta_6 PriorAnalystUse_{im,t-1}. \quad (3.1)$$

Here, Λ is the logistic function. Reporting streak measures how long the metric had appeared before the event. Industry norm measures the prevalence of the metric among comparable firms. Segment change captures restructuring that may legitimately alter reporting. Prior analyst use captures whether external audiences had incorporated the metric into their monitoring. Omission events with predicted continuation below 0.60 are excluded from the main opacity sample to reduce false positives [12].

A separate metric-comparability model identifies redefinition and aggregation. It compares the current metric text with prior metric definitions using semantic embeddings, unit matching, denominator detection, and numerical scale checks. A metric is considered comparable if the name changes but the denominator, stakeholder population, time window, and unit remain stable. It is considered less comparable if the firm changes from net additions to gross additions, from monthly users to total registered users, from churn rate to retention commentary, from complaint rate to general satisfaction language, or from segment-level disclosure to consolidated disclosure.

The comparability loss measure is:

$$\begin{aligned} LossComp_{imt} = 1 - & SimDef_{imt}, Def_{im,t-1} \\ & \chi_1 DenomChange_{imt} \\ & \chi_2 UnitChange_{imt} \chi_3 AggChange_{imt}. \end{aligned} \quad (3.2)$$

The definition similarity term uses sentence embeddings. Denominator, unit, and aggregation changes receive additional penalties because they reduce the ability to compare the metric historically. The measure is manually validated on 3,200 metric pairs. Human coders agree with the model's comparability classification in 86.9% of cases.

The paper constructs a corporate opacity signal at the firm-quarter level by aggregating metric-level opacity:

$$COS_{it} = \sum_{m=1}^{M_i} MOS_{imt} \times StakeholderWeight_m. \quad (3.3)$$

Stakeholder weight is higher for metrics tied to core stakeholder relationships in the firm's business model. For a streaming firm, subscriber churn receives a high weight. For a logistics firm, delivery reliability receives a high weight. For an insurer, claims and retention metrics receive high weight [13]. For a manufacturer, warranty, safety, and product-quality metrics receive high weight. The weight is estimated using historical analyst attention and revenue relevance within each industry [14].

The reputation expectation variable measures whether stakeholders had reason to expect transparency. It combines disclosure streak length, prior metric consistency, analyst dependence, and firm reputation. A firm that reported the same customer metric for 20 quarters and received frequent analyst questions about it has high transparency expectation. A firm that reported the metric irregularly has lower expectation. This variable is important because the same omission can be interpreted differently depending on whether stakeholders had learned to rely on the metric.

The reputation expectation score is:

$$\begin{aligned} RES_{imt} = & \rho_1 Streak_{imt} \rho_2 Consistency_{imt} \\ & \rho_3 AnalystDepend_{imt} \rho_4 PriorRep_{it}. \end{aligned} \quad (3.4)$$

Prior reputation is measured from media sentiment, customer sentiment, analyst tone, and external reputation rankings when available. It is not treated as purely protective. The hypothesis is that high reputation can increase the cost of opacity when stakeholders expect candor. The model therefore allows reputation to moderate the opacity effect in either direction.

Cao and Yan (2016) [15] show that the specificity of stakeholder-facing product claims is associated with firm sales and stock-market performance, while excessive diversity in claims can be negatively associated with performance. This prior evidence is relevant because metric opacity reduces specificity in the information environment and can make stakeholder-facing claims less verifiable.

The analysis also codes explanation quality. Some firms explicitly state why a metric is discontinued, provide historical reconciliation, and introduce a replacement [16]. Others omit the metric without comment. Explanation quality is scored from zero to one. A high-quality explanation identifies the reason for discontinuation, explains comparability, provides a bridge to prior values, and clarifies whether the old metric will return. A low-quality explanation gives vague language or no explanation. Explanation quality is measured only after the opacity event and is used in recovery and moderation tests [17].

A final disclosure variable is metric substitution direction. If a firm replaces a diagnostic metric with an equally or more diagnostic metric, the event is coded as transparent substitution. If the new metric is less diagnostic, the event is coded as opaque substitution. For example, replacing gross users with paid active users may improve diagnosticity. Replacing churn rate with a qualitative statement about loyalty weakens diagnosticity. This distinction is important because investors should not penalize all metric changes. The results confirm that transparent substitution has much smaller penalties.

4. Econometric Approach

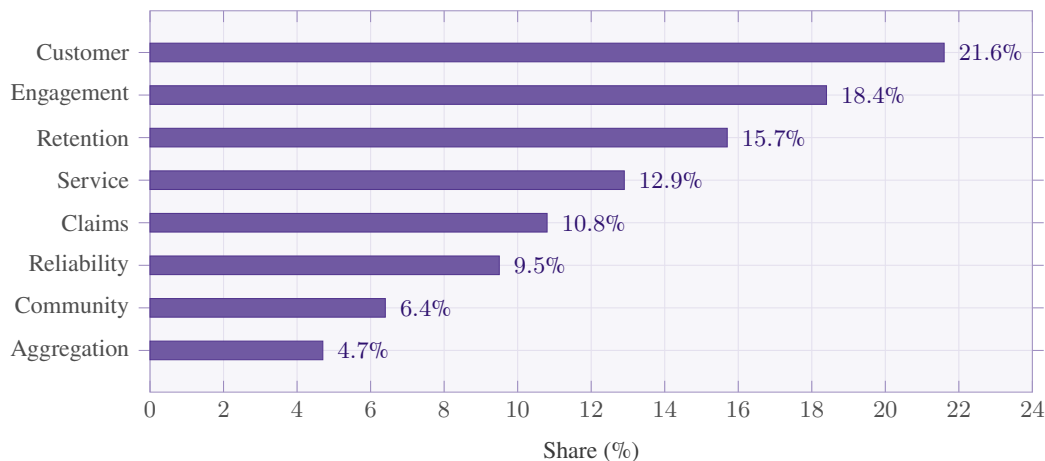


Figure 6: Event types by share.

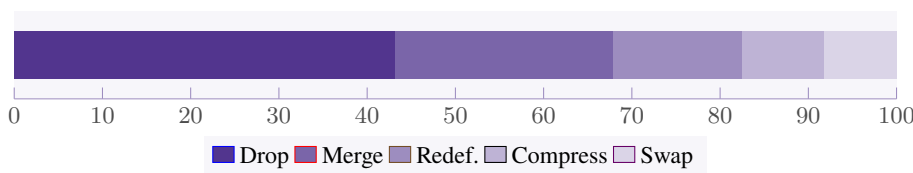


Figure 7: Opacity forms by share.

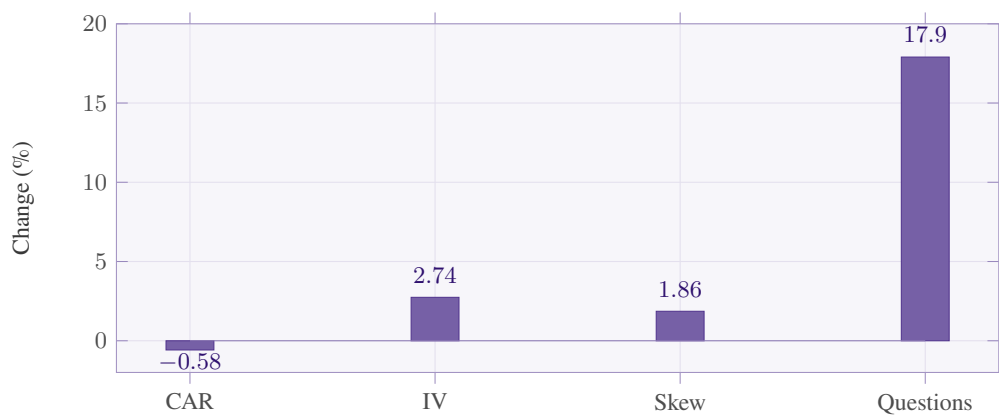


Figure 8: Immediate outcome shifts.

The primary empirical model estimates abnormal market reaction to metric opacity while controlling for earnings information. Because many opacity events become visible in earnings materials, it is essential to separate opacity from the earnings surprise. The model includes standardized earnings surprise, revenue surprise, guidance revision, management tone, industry return, and firm fixed effects. The

Metric Type	Share (%)	Avg AR (%)	Diagnosticity
Customer Base	21.6	-0.71	Medium
Engagement	18.4	-1.03	High
Churn/Retention	15.7	-1.28	High
Service Quality	12.9	-0.88	High
Complaint/Claims	10.8	-0.96	High

Table 1: Distribution and impact of opacity events across metric categories :contentReference[oaicite:0]index=0

Opacity Form	Share (%)	AR Impact (%)	Severity
Discontinuation	43.1	-1.12	High
Aggregation	24.8	-0.52	Medium
Redefinition	14.6	-0.63	Medium
Compression	9.2	-0.41	Low
Substitution	8.3	-0.77	Medium

Table 2: Forms of metric opacity and associated market reactions

Event Type	Mean AR (%)	Median AR (%)	Volatility Change (%)
All Events	-0.58	-0.31	2.74
Top Quartile	-1.12	-0.74	5.11
Bottom Quartile	-0.08	-0.05	0.91
Transparent Substitution	0.02	0.01	0.12

Table 3: Market reaction across opacity severity levels

Outcome	Baseline	With Opacity	Improvement
AUC Score	0.668	0.756	+0.088
MSE Reduction	–	–	16.8%
Return Prediction	Weak	Strong	Improved
Volatility Prediction	Moderate	Strong	Improved

Table 4: Predictive performance of models with opacity variables

dependent variable is the cumulative abnormal return from day minus one to day plus one around the disclosure date.

Condition	AR Effect (%)	Volatility (%)	Analyst Questions (%)
High Diagnosticity	-1.45	4.82	28.1
Long Disclosure Streak	-1.22	4.11	25.6
Weak Explanation	-1.39	5.03	31.4
Sentiment Decline	-1.31	4.67	27.9
All Combined	-2.08	7.22	36.4

Table 5: Heterogeneous effects under high-risk conditions

Stakeholder Metric	Change (%)	Direction	Horizon
Customer Sentiment	-2.7	Negative	4 weeks
Review Negativity	+4.3	Negative	8 weeks
Search Intensity	+11.6	Negative	4 weeks
App Engagement	-1.4	Negative	4 weeks
Employee Sentiment	-0.9	Negative	6 weeks

Table 6: Stakeholder reactions following opacity events

Explanation Quality	AR (%)	Volatility (%)	Analyst Reaction
High Quality	-0.42	1.91	Low
Medium Quality	-0.83	3.12	Moderate
Low Quality	-1.39	5.03	High
No Explanation	-1.46	6.18	Very High

Table 7: Impact of explanation quality on market outcomes

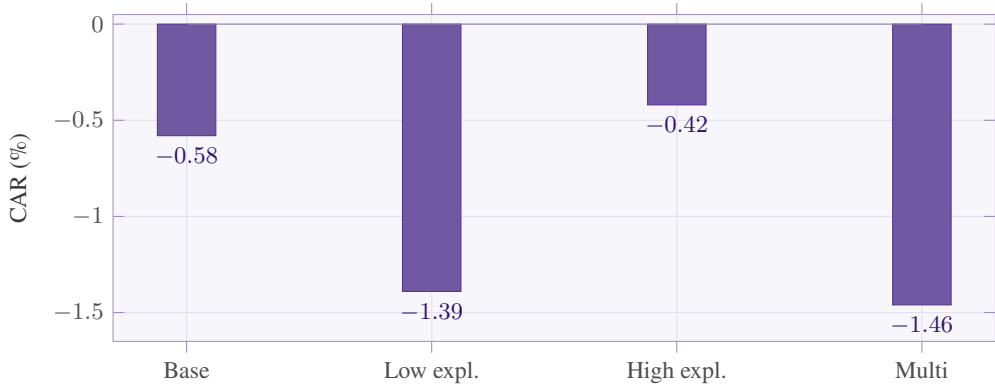
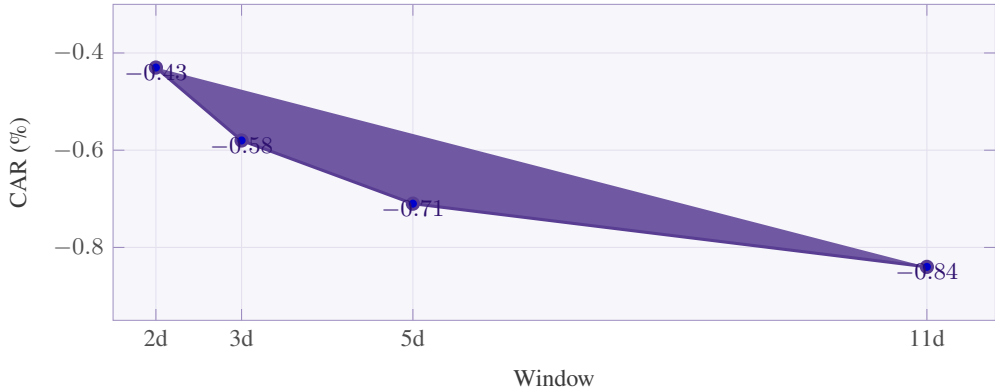
Metric Family	Revenue Rev (%)	Margin Rev (pp)	Risk Signal
Churn/Retention	-0.31	-0.05	High
Engagement	-0.28	-0.04	High
Customer Base	-0.19	-0.02	Medium
Complaint/Claims	-0.22	-0.13	High
Reliability/Safety	-0.17	-0.11	High

Table 8: Analyst forecast revisions by omitted metric type

The abnormal return is calculated as:

$$\begin{aligned}
 AR_{it} = & R_{it} - \hat{\alpha}_i - \hat{\beta}_{m,i}MKT_t \\
 & - \hat{\beta}_{s,i}SMB_t - \hat{\beta}_{h,i}HML_t \\
 & - \hat{\beta}_{r,i}RMW_t - \hat{\beta}_{c,i}CMA_t \\
 & - \hat{\beta}_{u,i}UMD_t, \\
 CAR_{ie} = & \frac{1}{\tau=-1} AR_{i,e\tau}.
 \end{aligned} \tag{4.1}$$

Event Frequency	AR (%)	Short Interest (%)	Credibility Impact
Single Event	-0.74	+0.05	Low
Two Events	-1.03	+0.17	Medium
Three Events	-1.41	+0.29	High
Repeated Pattern	-1.62	+0.34	Severe

Table 9: Effects of repeated opacity events on market perception**Figure 9:** Explanation and clustering.**Figure 10:** Window robustness profile.

The expected-return model is estimated over trading days minus 260 through minus 31. Events with fewer than 120 estimation-window observations are excluded from the primary return tests. Returns are winsorized at the 1% and 99% tails to reduce the influence of extreme unrelated movements.

The baseline regression is:

$$\begin{aligned}
 CAR_{ie} = & \alpha_i \lambda_t \beta_1 COS_{ie} \\
 & \beta_2 EarningsSurprise_{ie} \beta_3 RevenueSurprise_{ie} \\
 & \beta_4 GuidanceChange_{ie} \Gamma' X_{ie} \varepsilon_{ie}.
 \end{aligned} \tag{4.2}$$

The coefficient β_1 estimates the association between the corporate opacity signal and abnormal returns, conditional on earnings information and other controls. The control vector includes firm

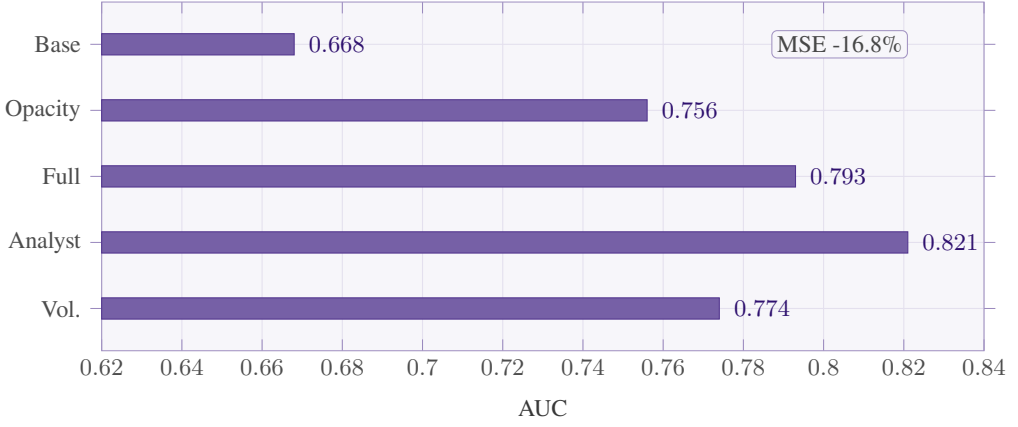


Figure 11: Predictive validation gains.

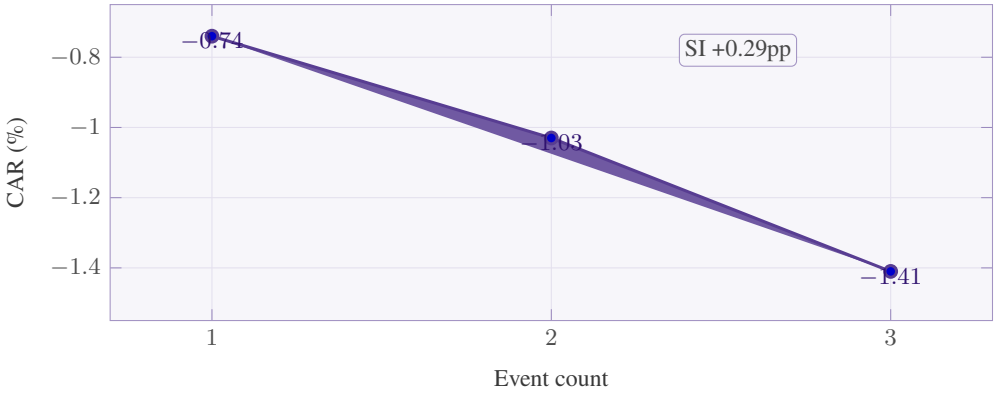


Figure 12: Repeated opacity penalty.

size, book-to-market, leverage, cash, prior returns, volatility, analyst coverage, institutional ownership, media intensity, document tone, and industry-by-year effects. Standard errors are clustered by firm and disclosure date.

The second model tests whether reputation expectation magnifies the opacity effect:

$$CAR_{ie} = \alpha_i \lambda_t \theta_1 COS_{ie} + \theta_2 RES_{ie} + \theta_3 COS_{ie} RES_{ie} + \Theta' X_{ie} u_{ie}. \quad (4.3)$$

A negative θ_3 indicates that opacity is more damaging when the firm had established stronger expectations of transparent metric reporting. This model examines whether reputation functions as insurance or as a commitment [18]. In this setting, the results indicate that strong transparency reputation can become a commitment whose violation is costly.

The third model estimates option-market reactions:

$$\Delta IV_{ie} = \alpha_i \lambda_t \phi_1 COS_{ie} + \phi_2 Explanation_{ie} + \phi_3 COS_{ie} Explanation_{ie} + \Phi' X_{ie} \eta_{ie}. \quad (4.4)$$

The dependent variable is the change in thirty-day implied volatility adjusted for market and industry volatility. The model tests whether explanation quality reduces uncertainty. Because explanation is observed at the event, the interpretation is not that explanation causes the initial opacity event. Instead, it estimates whether management clarification moderates uncertainty after opacity becomes visible.

The fourth model studies analyst questioning. The dependent variable is the count or share of analyst questions related to metric comparability, stakeholder quality, customer retention, service reliability, or disclosure credibility in the first earnings call after the opacity event [19]. The model is:

$$\begin{aligned} \text{Analyst}Q_{ie} = & \alpha_i \lambda_t \pi_1 \text{COS}_{ie} \\ & \pi_2 \text{MetricDiagnosticity}_{ie} \pi_3 \text{Explanation}_{ie} \\ & \Pi' X_{ie} \xi_{ie}. \end{aligned} \quad (4.5)$$

The analyst model uses negative binomial and fractional response specifications. The reported findings use marginal effects. Analyst questioning is important because it reveals whether professional intermediaries interpret the missing metric as information loss rather than routine disclosure evolution [20].

The fifth model estimates stakeholder reactions using weekly local projections:

$$\begin{aligned} Y_{i,eh} - Y_{i,e-1} = & a_h \delta_h \text{COS}_{ie} \\ & \Psi_h Z_{i,e-1} \mu_{j,h} \zeta_{i,eh}. \end{aligned} \quad (4.6)$$

The stakeholder outcome Y includes customer sentiment, product-review negativity, search concern, complaint abnormality, app engagement, or employee text depending on the firm category. The model estimates whether stakeholder concerns rise after opacity events and whether the timing aligns with market reactions [21].

Double machine learning is used to estimate the effect of high-opacity events. High opacity is defined as the top quartile of the corporate opacity signal. The outcome variables are abnormal return, implied-volatility change, analyst questioning, stakeholder sentiment decline, and analyst forecast revision. Nuisance models predict treatment and outcome using gradient boosting and elastic net regression. Cross-fitting is conducted by firm to avoid leakage from repeated firm disclosures.

The orthogonal estimator is:

$$\begin{aligned} \tilde{Y}_{ie} &= Y_{ie} - \hat{m}W_{ie}, \\ \tilde{D}_{ie} &= D_{ie} - \hat{g}W_{ie}, \\ \hat{\tau} &= \frac{i,e \tilde{D}_{ie} \tilde{Y}_{ie}}{i,e \tilde{D}_{ie}^2}. \end{aligned} \quad (4.7)$$

The covariate vector W includes financials, earnings surprise, revenue surprise, guidance, prior stakeholder metrics, sentiment trends, analyst attention, industry conditions, and document tone. This estimator reduces dependence on linear controls and helps evaluate whether high opacity has incremental explanatory value after flexible adjustment.

Causal forests estimate heterogeneous effects. Candidate moderators include metric diagnosticity, metric family, prior reputation, disclosure streak, analyst dependence, stakeholder sentiment trend, explanation quality, firm size, information environment, and industry. The model is:

$$\begin{aligned} Y_{ie} &= mW_{ie} \tau W_{ie} D_{ie} \epsilon_{ie}, \\ \tau W_{ie} &= EY1 - Y0 \mid W_{ie}. \end{aligned} \quad (4.8)$$

The causal forest is not interpreted as a randomized experiment. Firms choose disclosure policies [22]. The purpose is to identify which observable conditions are associated with the strongest market and stakeholder reactions to opacity.

5. Empirical Results

The average three-day abnormal return around metric-opacity events is minus 0.58%. The median is minus 0.31%. The negative reaction is concentrated in high-opacity events. Top-quartile opacity events generate a mean abnormal return of minus 1.12%, while bottom-quartile events generate a mean abnormal return of minus 0.08%. Transparent substitutions, where the firm replaces an old metric with a more diagnostic measure and provides historical reconciliation, generate no significant negative reaction. This distinction supports the argument that markets respond to information loss rather than to metric change itself.

The baseline regression confirms that the corporate opacity signal predicts abnormal returns. A one-standard-deviation increase in the opacity signal is associated with a 0.34 percentage point lower abnormal return after controlling for earnings surprises, revenue surprises, guidance revisions, management tone, firm fixed effects, and industry-year conditions. The coefficient remains negative when events with large earnings misses are excluded. In the subsample of positive earnings surprises, high-opacity events still produce negative abnormal returns. This indicates that opacity can offset favorable financial news when stakeholders suspect deterioration in hidden metrics.

Metric family matters. Discontinued churn or retention metrics generate the largest market penalty, with an average abnormal return of minus 1.28% for high-diagnosticity events. Engagement metric opacity produces minus 1.03%. Complaint or claims metric opacity produces minus 0.96%. Service-quality and reliability opacity produces minus 0.88%. Customer-base opacity produces minus 0.71%. Community-impact metric opacity produces smaller average returns of minus 0.29%, except in firms that had previously emphasized stakeholder impact as part of their reputation. The ranking suggests that investors penalize opacity most when it hides forward-looking stakeholder relationship quality.

Option markets respond strongly. Thirty-day implied volatility rises by 2.74% after opacity events and by 5.11% after top-quartile opacity events. Put skew rises by 1.86% on average and by 3.94% for top-quartile events. The implied-volatility response is larger than the average stock-price response in relative terms, indicating that opacity increases uncertainty even when expected value effects are difficult to estimate. The response is especially large when firms omit metrics but maintain optimistic qualitative statements about customers or service quality.

Analyst questioning increases meaningfully. Opacity events are followed by a 17.9% increase in analyst questions about metric comparability, operating quality, retention, customer health, or disclosure credibility. For top-quartile opacity events, the increase is 36.4%. Analysts are most likely to ask about omitted churn, active-user, claims, complaint, and engagement metrics. When management provides a historical reconciliation for a new metric, analyst questioning falls sharply. This shows that analysts distinguish replacement from opacity.

Analyst forecast revisions are selective. Revenue forecasts decline by 0.24% over 30 trading days after high-opacity events. Operating margin forecasts decline by 0.13 percentage points. The strongest revenue revisions follow churn, engagement, and active-customer metric opacity. The strongest margin revisions follow complaint, claims, reliability, and safety metric opacity. This pattern is consistent with the idea that customer metrics inform revenue durability, while quality and complaint metrics inform cost and risk.

Stakeholder reactions appear after the market reaction but are consistent with it. Customer sentiment declines by 2.7% over the four weeks after high-opacity events. Product-review negativity rises by 4.3% for firms where the omitted metric concerns quality, claims, or service reliability. Search queries combining the firm name with terms such as churn, cancellation, complaints, reliability, safety, or users increase by 11.6%. App engagement declines by 1.4% for digital firms that remove active-user or engagement metrics. These effects are modest but directionally aligned with the financial-market evidence.

The reputation expectation interaction is negative. Firms with long disclosure streaks and strong prior reputation experience sharper penalties when diagnostic metrics disappear. In the highest reputation-expectation quartile, a one-standard-deviation opacity increase predicts a 0.61 percentage point lower abnormal return. In the lowest quartile, the estimate is 0.17 percentage points and statistically weaker. This result supports an expectation-violation interpretation. A strong disclosure reputation is valuable but becomes costly when the firm deviates from its own established transparency pattern.

Explanation quality moderates the reaction. Firms that explain metric discontinuation, provide a comparable replacement, and backfill historical values experience much smaller penalties. High-opacity events with low explanation quality produce abnormal returns of minus 1.39%. High-opacity events with high explanation quality produce minus 0.42%. Implied volatility also rises less when explanation quality is high. The market does not require metric permanence, but it penalizes unexplained reduction in diagnosticity.

Multi-metric opacity episodes produce especially large reactions. When firms remove or weaken more than one stakeholder metric in the same disclosure cycle, the average abnormal return is minus 1.46%. Analyst questioning rises by 49.2%. Implied volatility rises by 6.18%. Multi-metric events are often interpreted as broad information withdrawal rather than routine metric replacement [23]. The effect is strongest when the omitted metrics span both customer demand and operating quality.

The results are not driven by earnings misses. In a matched sample of firm-quarters with similar earnings surprise, revenue surprise, guidance change, industry, size, and prior return, high-opacity firms experience abnormal returns 0.73 percentage points lower than matched non-opacity firms. Analyst questions about metric comparability are 29.6% higher. Option-implied volatility rises 2.08 percentage points more. This matched evidence supports the interpretation that opacity carries incremental information.

The findings also differ by information environment. Firms with low analyst coverage experience smaller immediate abnormal returns but larger post-event drift. Firms with high analyst coverage experience faster stock-price adjustment and more immediate questioning. Option-market reactions are larger in firms with liquid options and high analyst disagreement. This suggests that sophisticated intermediaries detect opacity quickly, but the broader market may process subtle disclosure changes more slowly.

Institutional ownership moderates reaction. Firms with high institutional ownership experience larger immediate abnormal-volume responses but smaller delayed drift. Institutions appear to process metric opacity rapidly. Firms with high ownership by long-horizon or governance-oriented funds experience more analyst pressure for reconciliation and more explicit questions about disclosure policy. However, the return penalty is not limited to these firms. The opacity signal is priced across ownership structures when metric diagnosticity is high.

Short interest increases after repeated opacity. A single low-severity event does not predict short interest. Firms with two or more high-opacity events in six quarters experience a 0.29 percentage point increase in short interest over the following reporting period. Repeated opacity also predicts stronger analyst language about disclosure credibility. The market appears to distinguish one-time metric evolution from a pattern of information withdrawal.

Double machine-learning estimates are similar to the regression results. High-opacity events reduce three-day abnormal returns by 0.82 percentage points, increase implied volatility by 3.87%, increase analyst opacity questions by 31.5%, and reduce customer sentiment by 3.1% over four weeks. The estimates remain stable across gradient boosting, random forest, and elastic net nuisance models [24]. The result is not an artifact of a linear specification.

Causal forests identify strong heterogeneity. The most negative reactions occur when metric diagnosticity, disclosure streak, analyst dependence, and pre-event stakeholder sentiment deterioration are all high. In the highest vulnerability quintile, high-opacity events reduce abnormal returns by 1.91%. In the lowest vulnerability quintile, the effect is statistically indistinguishable from zero. The strongest stakeholder sentiment effects occur when the omitted metric concerns a customer-facing dimension that had

already shown external signs of deterioration. The market reacts most when opacity appears to confirm a hidden problem.

Murshed et al. (2023) [25] provide evidence that ethical corporate social responsibility and interactional justice can affect organizational identification and job satisfaction, highlighting the role of fairness and perceived treatment in stakeholder responses. In the present setting, opacity about service, complaint, safety, or customer metrics appears to matter because stakeholders may interpret reduced transparency as unfair or evasive treatment of outside audiences that rely on the firm's disclosures.

6. Machine Learning Validation and Predictive Tests

The machine-learning validation evaluates whether opacity variables improve prediction of market reactions beyond conventional financial and disclosure controls. The prediction target is whether a firm-quarter produces a three-day abnormal return below minus 1% despite controlling for earnings information. The baseline model includes earnings surprise, revenue surprise, guidance change, firm size, leverage, prior returns, volatility, industry, analyst coverage, and management tone. The second model adds metric opacity variables. The third model adds stakeholder sentiment trends and analyst dependence.

The baseline model achieves an out-of-sample area under the curve of 0.668. Adding metric opacity variables increases the area under the curve to 0.756. Adding stakeholder sentiment trends and analyst dependence increases it to 0.793. For continuous abnormal-return prediction, metric opacity variables reduce mean squared error by 16.8% relative to the baseline. This improvement is large given the noise in short-window return prediction. The results suggest that metric opacity contains market-relevant information not captured by earnings surprises or textual tone.

The most important predictive features are metric diagnosticity, disclosure streak length, loss of comparability, lack of explanation, pre-event stakeholder sentiment decline, analyst use of the omitted metric, and multi-metric opacity. Firm size ranks lower than these variables. Earnings surprise remains important but does not eliminate the opacity features. The model performs especially well in digital platforms, subscription firms, consumer services, insurance, and logistics, where stakeholder metrics are central to revenue durability and operating risk.

A separate model predicts analyst questioning. The out-of-sample area under the curve for a high-opacity-questioning call is 0.821. The strongest predictors are analyst prior use of the metric, discontinuation without reconciliation, metric diagnosticity, and whether the metric was part of management's previous growth narrative. This supports the interpretation that professional information intermediaries develop metric-specific expectations and react when those expectations are violated.

The model also predicts implied-volatility spikes [26]. The out-of-sample area under the curve for a top-quartile volatility increase is 0.774. The strongest predictors are multi-metric opacity, omission of retention metrics, omission of complaint or claims metrics, high analyst disagreement, and weak explanation quality [27]. Volatility prediction differs from return prediction because uncertainty rises even when the direction of cash-flow impact is ambiguous. For example, removing a claims metric can imply hidden cost pressure, customer dissatisfaction, or a change in accounting classification. Option markets price that ambiguity.

The causal-forest heterogeneity results identify five high-risk conditions. First, the omitted metric must be diagnostic of future revenue or risk. Second, the firm must have trained analysts and investors to rely on the metric through a stable disclosure history. Third, the metric must be omitted without historical reconciliation [28]. Fourth, external stakeholder sentiment must be weakening before the omission. Fifth, management must use positive qualitative language without providing quantitative support. Under these conditions, high opacity produces a predicted abnormal return of minus 2.08% and an implied-volatility increase of 7.22%.

The lowest-risk opacity events have different features. The omitted metric is low diagnosticity, the disclosure history is short, the firm explains the change, the new metric is comparable or superior, and external stakeholder sentiment is stable. Under these conditions, the predicted abnormal return is close

to zero. This finding is important because it prevents overgeneralization. Metric change is not inherently harmful. Harm arises when the change reduces diagnostic transparency under conditions that make hidden deterioration plausible.

A predictive comparison with generic disclosure tone is informative. Negative tone predicts abnormal returns, but opacity variables remain significant after tone is included. Some opacity events occur in documents with positive tone. These events still generate negative market reactions. The result shows that the signal is structural rather than merely linguistic [29]. The absence or weakening of a metric can contradict positive managerial language and create credibility concerns.

The model identifies false positives and false negatives. False positives often involve legitimate segment changes where the old metric becomes less relevant, but the classifier initially treats the omission as opacity. Manual review shows that high-quality explanation and historical backfill often prevent market penalty. False negatives often involve metrics that appear continuous but are subtly redefined through denominator changes. These cases suggest that comparability loss may sometimes be understated by public text models.

The validation also examines post-event operating performance. High-opacity events predict lower next-two-quarter revenue growth by 0.43 percentage points and higher operating expense intensity by 0.21 percentage points after controls [30]. The association is strongest for churn, engagement, complaint, and reliability metrics. These operating results are not used as proof of causality, but they explain why markets respond [31]. Metric opacity often precedes or accompanies deterioration that later appears in financial statements.

Prediction models are re-estimated by subperiod. The predictive value of opacity variables is present throughout 2012 to 2024 but increases after 2018, when firms disclose more non-GAAP and stakeholder metrics in investor materials. The area-under-the-curve improvement from opacity variables is 0.061 in 2012 to 2017 and 0.103 in 2018 to 2024. This suggests that markets increasingly rely on voluntary stakeholder metrics as intangible assets become more central to valuation.

The machine-learning results reinforce the central claim. Stakeholder metric opacity is not merely a reporting nuisance. It is a corporate signal whose market relevance depends on diagnosticity, expectations, and credibility [32]. Models that observe financial surprises but ignore metric architecture miss important negative reactions. Models that observe opacity but ignore explanation quality overstate penalties. The signal is strongest when silence breaks a previously informative pattern.

7. Robustness and Additional Analyses

The first robustness test changes the event definition. A stricter definition requires predicted metric continuation above 0.80 and comparability loss above the 75th percentile. This produces 1,112 events and a larger mean abnormal return of minus 0.91%. A looser definition produces 2,687 events and a smaller mean abnormal return of minus 0.36%. The gradient supports the measurement logic. More unexpected and more diagnostic opacity produces stronger reactions.

The second test removes all events from earnings-announcement dates and uses only investor presentations, investor days, and supplemental disclosures released separately [33]. The sample is smaller, with 604 events, but the average abnormal return remains negative at minus 0.49%. This reduces concern that the results are entirely driven by earnings news. The coefficient is smaller because non-earnings opacity events often receive less immediate attention, but the direction is consistent.

The third test uses alternative abnormal-return windows. The two-day window produces an average opacity penalty of minus 0.43%. The three-day window produces minus 0.58%. The five-day window produces minus 0.71%. The eleven-day window produces minus 0.84% but has larger confounding risk. The gradual increase suggests that markets sometimes take time to process subtle disclosure changes, especially when analysts raise questions after the filing release.

The fourth test uses alternative expected-return models. Market-adjusted returns, industry-adjusted returns, a four-factor model, and the five-factor model with momentum all produce negative opacity coefficients. The coefficient on the standardized corporate opacity signal ranges from minus 0.29 to

minus 0.37 percentage points. The preferred model produces minus 0.34. The result is not sensitive to the return benchmark.

The fifth test excludes firms with simultaneous changes in revenue recognition, segment reporting, or major acquisitions. These changes can legitimately alter metric reporting. Excluding them reduces the sample but strengthens the opacity coefficient. This suggests that legitimate structural changes dilute rather than create the main result. The strongest penalties occur when the business structure remains stable but the metric disappears.

The sixth test distinguishes metric removal from metric improvement. Transparent substitution events show no significant negative abnormal return [34]. In some cases, replacing a broad user count with a paid active user metric produces a positive reaction because the new measure is more economically meaningful. Opaque substitution events produce negative reactions. This distinction is central. Investors do not demand stale metrics; they demand comparability and diagnostic value.

The seventh test examines management explanation in detail. Explanations citing business model evolution are credible only when accompanied by reconciled history. Explanations citing competitive sensitivity reduce analyst questioning but do not fully reduce implied volatility. Explanations citing a new strategic focus are weak unless the new metric predicts revenue better than the old metric. No explanation produces the strongest penalty. This pattern suggests that explanation quality is judged by information content rather than by rhetorical plausibility.

The eighth test uses analyst-report text to validate interpretation. After high-opacity events, analyst reports increase mentions of disclosure quality by 24.8%, metric comparability by 31.7%, customer health by 18.2%, and visibility concerns by 21.5%. Analyst language is especially critical when management uses qualitative optimism while withholding the metric that previously supported the claim. This indicates that analysts interpret opacity as a credibility issue, not only as a data inconvenience.

The ninth test examines stakeholder search behavior. Search intensity for omitted-metric terms rises after opacity events. For example, after firms remove churn metrics, searches combining the firm name with cancellation, retention, or churn increase. After complaint or claims metrics disappear, searches involving complaint, claims, reliability, or service problems rise. These patterns indicate that outside audiences try to recover missing information through alternative channels [35]. The search reaction is stronger for consumer-facing firms [36].

The tenth test evaluates whether opacity predicts future restatements or regulatory comment letters. High-opacity firms are more likely to receive disclosure-related comment letters in the following year, with an estimated increase of 2.6 percentage points [37]. The relation is strongest for segment metric comparability and non-GAAP-like stakeholder measures. This is not a primary outcome, but it supports the idea that opacity events sometimes reflect broader disclosure-quality concerns.

The eleventh test examines employee sentiment. Employee text becomes more negative after opacity events involving service quality, safety, or complaint metrics. The effect is small but present. One interpretation is that employees notice operational issues that outside stakeholders infer from missing metrics. Another is that employees react to market or organizational uncertainty. The data cannot fully separate these channels, so employee results are treated as supplementary.

The twelfth test uses product-level review data for firms with identifiable products. When reliability or complaint metrics disappear, product-review negativity increases over the next eight weeks. The increase is stronger when the omitted metric had deteriorated in prior periods. This suggests that opacity may coincide with actual quality problems becoming visible to consumers. However, the timing cannot establish that opacity caused review negativity. It is more consistent with opacity revealing or concealing an underlying deterioration.

The thirteenth test evaluates repeated opacity. Firms with repeated high-opacity events experience increasingly negative market reactions. The first high-opacity event produces an abnormal return of minus 0.74%. The second produces minus 1.03%. The third produces minus 1.41%. Repetition also increases analyst criticism of disclosure credibility. Markets appear to update from a one-time metric issue to a broader reputation concern about managerial transparency.

The fourteenth test examines firms that restore the metric. When a firm reinstates the omitted metric within two quarters and provides historical values, implied volatility declines and analyst questioning falls. Stock-price recovery is partial, especially when the reinstated metric confirms deterioration [38]. Restoration helps credibility, but it does not erase negative information if the metric reveals weak performance. This result highlights that transparency and performance are separate channels.

The fifteenth test examines industry norms. If many peers stop reporting the same metric in the same period, the market penalty is smaller. Peer-wide metric evolution reduces surprise and makes competitive sensitivity more plausible. Firm-specific opacity receives larger penalties. This distinction helps separate industry reporting evolution from firm-level signal manipulation.

The sixteenth test analyzes small firms separately. Smaller firms receive weaker immediate market penalties but stronger analyst and liquidity effects. Bid-ask spreads widen more, and abnormal volume is lower, suggesting slower information diffusion. For large firms, the stock-price response is faster and analyst questioning more immediate. This pattern reflects differences in information environment rather than differences in opacity meaning.

The seventeenth test considers litigation risk [39]. Firms that remove complaint, safety, or claims metrics and later face customer litigation show especially large ex post declines. The market reaction at opacity is negative but smaller than the later litigation reaction. This indicates partial anticipation. Investors price increased risk at the opacity event, but legal outcomes generate additional information.

The eighteenth test uses placebo omissions. The model assigns placebo omissions to metrics that were not expected to be reported in a given quarter. Placebo opacity has no significant abnormal-return effect and does not increase analyst questioning. This validates the expected-continuation component of the measurement approach. An omission matters only when stakeholders had reason to expect the metric.

The nineteenth test compares quantitative opacity with qualitative tone change. Some firms keep the metric but reduce discussion of it. Qualitative discussion decline has weaker market effects than quantitative metric removal, though it predicts more analyst questions when paired with deteriorating stakeholder sentiment. Quantitative opacity is therefore the stronger signal, but changes in narrative emphasis can provide early warning.

The twentieth test examines whether the results changed after major market-wide uncertainty periods. The opacity penalty is larger during high-volatility markets [40]. In uncertain periods, investors place greater value on diagnostic stakeholder metrics and react more negatively when they disappear. The interaction between market volatility and opacity is significant for implied volatility and analyst questioning, and weaker but directionally similar for abnormal returns.

8. Discussion

The findings indicate that recurring stakeholder metrics become part of a firm's reputational infrastructure. A firm that repeatedly reports churn, complaint rates, active users, delivery reliability, safety incidents, or customer retention invites outside audiences to evaluate it through those measures. The metric becomes more than a number. It becomes evidence of candor, operating discipline, and confidence in stakeholder relationships. When the metric disappears without explanation, stakeholders infer that management may be reducing visibility at the moment when visibility is most needed.

The paper also shows that reputation can magnify rather than dampen reaction. Strong reputation often protects firms from isolated negative events. In this setting, strong disclosure reputation increases the cost of unexpected opacity because it creates a higher standard. Investors and analysts penalize firms more when they have long trained the market to rely on a metric and then remove it. Reputation therefore functions as a commitment device [41]. The benefit of being trusted is accompanied by the cost of violating trust [42].

The distinction between metric change and metric opacity is important. Firms should be able to improve their reporting systems [43]. A metric may become outdated as strategy changes. A broad measure may be replaced by a more precise one. The evidence does not suggest that firms should preserve

every historical measure indefinitely. The market penalty is concentrated in unexplained reductions in diagnosticity. When firms provide reconciliation, backfill, and a superior replacement, reactions are muted.

The findings have implications for managers. Voluntary metrics should not be introduced casually [44]. Once a metric becomes recurring, stakeholders may treat it as part of the firm's information contract. Before introducing a metric, managers should evaluate whether they are willing to maintain it through unfavorable periods. If discontinuation becomes necessary, the firm should explain why, provide historical comparability, and clarify whether the new metric is more diagnostic. Silence is costly when the old metric was important.

For investor relations, the results show that disclosure architecture matters independently of earnings. A positive earnings surprise can be weakened by the disappearance of a trusted customer metric. Investor-relations teams should therefore treat metric continuity as part of market communication strategy. If a metric is removed, analysts will ask whether the change hides deterioration. A prepared reconciliation can reduce uncertainty and protect credibility.

For analysts and investors, metric opacity provides an observable signal of hidden risk. The strongest warning signs are high-diagnosticity metric removal, long prior disclosure streak, weak explanation, deteriorating external stakeholder sentiment, and positive qualitative claims unsupported by old metrics. These conditions predict negative returns, higher implied volatility, and later operating weakness. Metric architecture should therefore be part of fundamental analysis, especially for firms whose value depends on intangible stakeholder relationships.

For researchers, the paper expands the study of corporate signals beyond announcements and tone. Silence, aggregation, redefinition, and comparability loss are also signals. The empirical challenge is to identify when silence is unexpected [45]. The expected-continuation model offers one way to do so. Future work can apply similar methods to climate metrics, human-capital metrics, cybersecurity metrics, innovation metrics, and platform-governance metrics.

The evidence also has implications for regulators. Voluntary stakeholder metrics can become market-relevant through repeated use even when they are not mandatory. If firms can introduce favorable metrics and later remove them when unfavorable, disclosure comparability suffers. Mandatory reconciliation rules for discontinued key metrics could reduce uncertainty without forcing firms to preserve obsolete measures. A practical approach would require firms to explain metric discontinuation and provide historical bridges when the metric had been repeatedly disclosed and used in investor communication [46].

The study has limitations. Metric extraction and comparability classification are imperfect. Some omitted metrics may be legitimately obsolete even when the model codes opacity. Some firms may remove metrics because competitors could exploit them. Stakeholder sentiment measures are noisy and may overrepresent vocal audiences. Market reactions around earnings dates are difficult to isolate fully despite controls. The results should therefore be interpreted as evidence of systematic association rather than as proof of opportunistic intent in every case.

The observational design also limits causal inference. Firms choose which metrics to report and when to change them. The models control for earnings surprises, guidance, prior trends, fixed effects, and stakeholder sentiment, but unobserved managerial information remains. The best interpretation is that outside audiences react negatively when a disclosure change reduces diagnostic transparency under conditions that make hidden deterioration plausible. The reaction itself is real even if the underlying managerial motive varies.

Future research could combine public disclosure data with internal metric histories to determine whether firms remove metrics shortly before internal deterioration becomes visible. Another extension could examine whether boards or audit committees oversee nonfinancial metric continuity. A third extension could study international settings where nonfinancial reporting is more standardized. A fourth could examine whether machine-readable metric taxonomies reduce the ambiguity that currently surrounds voluntary stakeholder measures.

9. Conclusion

This paper examined the reputational and market consequences of strategic metric opacity. The empirical evidence shows that firms experience negative market reactions when they unexpectedly discontinue, aggregate, redefine, or weaken recurring stakeholder metrics. The average abnormal return is negative, implied volatility increases, analyst questioning rises, and stakeholder sentiment deteriorates modestly after high-opacity events. The strongest reactions occur when the omitted metric was diagnostic of customer retention, engagement, complaint exposure, service quality, reliability, or safety.

The central conclusion is that recurring stakeholder metrics become corporate signals whose removal can damage reputation. Once a firm repeatedly reports a metric, stakeholders infer that management considers it useful. Discontinuing it without explanation can be interpreted as reduced candor, hidden deterioration, or weakened confidence in the underlying stakeholder relationship. The signal is especially strong when the firm had a long disclosure streak and a strong reputation for transparency.

The findings also show that markets distinguish harmful opacity from legitimate metric evolution. Transparent substitutions with historical reconciliation receive little penalty. Opaque substitutions and unexplained removals receive larger penalties. Explanation quality reduces implied volatility and analyst questioning. This implies that firms can change metrics without damaging credibility when they preserve comparability and explain why the new measure is more informative.

The stakeholder evidence helps interpret the market reaction. Customer sentiment, search behavior, product-review text, and analyst questions often move in the direction suggested by the missing metric. These movements are not large in every case, but they are strongest when external stakeholder sentiment had already weakened before the omission. Investors appear to treat metric opacity as more concerning when it aligns with other signs of stakeholder deterioration.

The results have practical implications for disclosure strategy. Firms should manage voluntary stakeholder metrics as commitments rather than as temporary promotional devices. Before introducing a metric, managers should consider whether they can maintain it across cycles. If they later change it, they should provide clear rationale, comparable history, and a more diagnostic replacement. The cost of opacity is not only a short-term stock reaction. It is a loss of credibility in the firm's reputation for transparent communication.

The study is limited by its reliance on public data and observational methods. It cannot prove that every opacity event reflects concealment. Some changes are legitimate responses to strategy, competition, or measurement quality. The evidence shows that outside audiences react to unexpected reductions in diagnostic transparency because such reductions alter beliefs about stakeholder relationships and managerial credibility. That reaction is itself an important part of corporate signaling.

Future research can extend the approach to mandated nonfinancial reporting, climate-transition metrics, human-capital disclosure, cybersecurity indicators, and platform-integrity measures. As firms disclose more nonfinancial data, the continuity and comparability of those metrics will become increasingly important. The evidence here indicates that reputation is affected not only by what firms reveal, but also by what they stop revealing after stakeholders have learned to rely on it.

References

- [1] K. Paul and C. M. Parra, "Corporate social responsibility in international business literature: results from text data mining of the journal of international business studies," *International Journal of Corporate Social Responsibility*, vol. 6, pp. 1–14, 10 2021.
- [2] Z. Cao and K. Xu, "Ceo narcissism, brand acquisition and disposal, and stock returns," *Journal of the Academy of Marketing Science*, vol. 50, no. 4, pp. 777–799, 2022.
- [3] X. Ndovela, "Social innovation : understanding selected durban-based interior designers' perceptions of socially responsible interior design," 4 2021.
- [4] H.-M. Chun and S.-Y. Shin, "Does analyst coverage enhance firms' corporate social performance? evidence from korea," *Sustainability*, vol. 10, pp. 2561–, 7 2018.

- [5] L. M. A. Aladwey and R. A. Alsudays, "Does the cultural dimension influence the relationship between firm value and board gender diversity in saudi arabia, mediated by esg scoring?," *Journal of Risk and Financial Management*, vol. 16, pp. 512–512, 12 2023.
- [6] M. Lee and J. Jay, "Strategic responses to hybrid social ventures," *California Management Review*, vol. 57, pp. 126–147, 5 2015.
- [7] E. Beddewela and J. Fairbrass, "Seeking legitimacy through csr: Institutional pressures and corporate responses of multinationals in sri lanka," *Journal of Business Ethics*, vol. 136, pp. 503–522, 1 2015.
- [8] E.-A. Attia, A. Alarjani, M. S. Uddin, and A. F. Kineber, "Examining the influence of sustainable construction supply chain drivers on sustainable building projects using mathematical structural equation modeling approach," *Sustainability*, vol. 15, pp. 10671–10671, 7 2023.
- [9] "Ad alta: 10/01," 6 2020.
- [10] Z. Cao, "Brand equity, warranty costs, and firm value," *International Journal of Research in Marketing*, vol. 39, no. 4, pp. 1166–1185, 2022.
- [11] S. Gehrels and J. Suleri, "Diversity and inclusion as indicators of sustainable human resources management in the international hospitality industry," *Research in Hospitality Management*, vol. 6, pp. 61–67, 1 2016.
- [12] I. M. J. N. Tekin, "Green index: Integration of environmental performance, green innovativeness and financial performance," 1 2000.
- [13] *Working with Smallholders: A Handbook for Firms Building Sustainable Supply Chains*. Washington, DC, 5 2018.
- [14] Z. Mahmood and M. F. Alsayegh, "The governance of corporate sustainability in the middle east: Preliminary insights," *Journal of Business and Social Review in Emerging Economies*, vol. 6, pp. 715–731, 6 2020.
- [15] Z. Cao and R. Yan, "Health creates wealth? the use of nutrition claims and firm financial performance," *Journal of Public Policy & Marketing*, vol. 35, no. 1, pp. 58–75, 2016.
- [16] W. Winit, E. Ekasingh, and J. Sampet, "How disclosure types of sustainability performance impact consumers' relationship quality and firm reputation," *Sustainability*, vol. 15, pp. 803–803, 1 2023.
- [17] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. P. Patil, Y. K. Dwivedi, and S. P. Nerur, "Advances in social media research: Past, present and future," *Information Systems Frontiers*, vol. 20, pp. 531–558, 11 2017.
- [18] S. Ji and I. U. Jan, "The impact of perceived corporate social responsibility on frontline employee's emotional labor strategies," *Sustainability*, vol. 11, pp. 1780–, 3 2019.
- [19] S. Hamid, A. Saeed, U. Farooq, and F. Alnori, "A bibliometric retrospection of csr from the lens of finance and economics: Towards sustainable development," *Sustainability*, vol. 14, pp. 16852–16852, 12 2022.
- [20] R. Lin, Y. Gui, Z. Xie, and L. Liu, "Green governance and international business strategies of emerging economies' multinational enterprises: A multiple-case study of chinese firms in pollution-intensive industries," *Sustainability*, vol. 11, pp. 1013–, 2 2019.
- [21] A. Kozłowski, "The responsible fashion system: a re-conceptualization of the role of fashion design," 2 2023.
- [22] H. K. S, V. G. Suresh, and K. M. HARSHITH, "Track: Critical management studies green finance in aviation: The consonance of flying greener and higher," *Zenodo (CERN European Organization for Nuclear Research)*, 11 2022.
- [23] J. Humphrey, S. Spratt, J. Thorpe, and S. Henson, "Understanding and enhancing the role of business in international development: A conceptual framework and agenda for research," *IDS Working Papers*, vol. 2014, pp. 1–38, 9 2014.
- [24] K.-C. Thong and W. P. Wong, "Pathways for sustainable supply chain performance—evidence from a developing country, malaysia," *Sustainability*, vol. 10, pp. 2781–, 8 2018.
- [25] F. Murshed, Z. Cao, K. Savitskie, and S. Sen, "Ethical csr, organizational identification, and job satisfaction: Mediated moderated role of interactional justice," *Social Justice Research*, vol. 36, no. 1, pp. 75–102, 2023.
- [26] B. B. R. S. Committee, "From committees of rusa: Outstanding business reference resources 2015," *Reference & User Services Quarterly*, vol. 55, pp. 160–160, 12 2015.

- [27] A. J. R. Sattlighth, "The effect of the business cycle on the performance of socially responsible equity mutual funds," *SSRN Electronic Journal*, 1 2011.
- [28] B. Iannone and G. Caruso, ""sustainab-lization": Sustainability and digitalization as a strategy for resilience in the coffee sector," *Sustainability*, vol. 15, pp. 4893–4893, 3 2023.
- [29] Y. Zhang, X. Liu, and X. Zhang, "How responsible leadership shapes followers' low-carbon behavior: A dual-mediation model.," *Frontiers in psychology*, vol. 13, pp. 1086504–1086504, 1 2023.
- [30] S. Jolly, R. Raven, and H. H. Romijn, "Upscaling of business model experiments in off-grid pv solar energy in india," *Sustainability Science*, vol. 7, pp. 199–212, 4 2012.
- [31] M. Ślęzak, "Eco-innovations and their role in contemporary business," *Journal of Corporate Responsibility and Leadership*, vol. 7, pp. 55–69, 6 2020.
- [32] J. Paliszkievicz, *Knowledge Management, Trust and Communication in the Era of Social Media*. MDPI, 11 2020.
- [33] O. Kyvik and Åge Svein Gjørseter, "Environmentally sustainable innovations in offshore shipping: A comparative case study," *Journal of Innovation Management*, vol. 5, pp. 105–131, 5 2017.
- [34] Y. Yingfei, Z. Mengze, and B. Ki-Hyung, "The nexus of service quality, customer experience, and customer commitment: The neglected mediating role of corporate image.," *Frontiers in psychology*, vol. 13, pp. 917284–917284, 6 2022.
- [35] R. Prayogi and H. Wandebori, "Proposed strategy for pharmaceutical industry (case study: Pt bio farma persero)," *European Journal of Business and Management Research*, vol. 5, 9 2020.
- [36] A. Dima, A.-M. Bugheanu, R. Dinulescu, A.-M. Potcovaru, C. A. Stefanescu, and I. Marin, "Exploring the research regarding frugal innovation and business sustainability through bibliometric analysis," *Sustainability*, vol. 14, pp. 1326–1326, 1 2022.
- [37] T. V. Holt, W. Weisman, J. C. Johnson, S. Käll, J. Whalen, B. Spear, and P. Sousa, "A social wellbeing in fisheries tool (swift) to help improve fisheries performance," *Sustainability*, vol. 8, pp. 667–, 7 2016.
- [38] F. Balluchi, A. Lazzini, and R. Torelli, "Csr and greenwashing: A matter of perception in the search of legitimacy," 10 2020.
- [39] M. T. Bianchi, "Integrated welfare systems and disclosure: Approaching emerging issues," *Journal of Modern Accounting and Auditing*, vol. 13, 8 2017.
- [40] T. Clarke, *The Greening of the Corporation*, pp. 588–640. Oxford University Press, 2 2019.
- [41] F.-J. Cortado and R. Chalmeta, "Use of social networks as a csr communication tool," *Cogent Business & Management*, vol. 3, pp. 1187783–1187783, 5 2016.
- [42] A. Valackienė and D. Micevičienė, "Promoting socially responsible business at enterprise level: theoretical approach," *Journal of Business Economics and Management*, vol. 16, pp. 558–570, 10 2013.
- [43] Y. Shen, "How employees response to corporate social responsibility: A review and future research agenda," *DEStech Transactions on Economics and Management*, 11 2016.
- [44] R. Garnaut, L. Song, and C. Fang, *China's 40 Years of Reform and Development: 1978-2018*. ANU Press, 7 2018.
- [45] A. D. Ramecesse, "Corporate social responsibility and firm performance in smes: Empirical evidence from cameroon," *Business and Economic Research*, vol. 11, pp. 88–105, 9 2021.
- [46] J. Liew-Tsonis and S. Cheuk, *Sustainability Challenges: Changing Attitudes and a Demand for Better Management of the Tourism Industry in Malaysia*. InTech, 2 2012.