



RESEARCH

The Sustainability Ceiling? Diminishing Returns to Perceived Carbon-Neutral Claim Credibility in Consumers' Premium Purchase Intention: Exploratory Evidence from Indonesia

 Alya Zahra Asyafa^{1*}

*Correspondence:

Alya Zahra Asyafa, Department of Management Faculty Economic and Business, Open University, alyajahra00@gmail.com

Received: 09/08/2026

Accepted: 02/08/2026

Abstract

Prior research typically models the relationship between consumers' perceptions of carbon-neutral product claims and their willingness to pay a premium as linear and positive. This study asks, in an exploratory manner, whether that relationship may instead be non-linear — rising at low-to-moderate levels of perceived claim credibility but flattening or reversing once such claims become pervasive — a pattern we term a “sustainability ceiling.”. Cross-sectional survey data were collected from 200 Indonesian consumers with recent experience purchasing environmentally labelled products. Perceived carbon-neutral claim credibility (X) and premium purchase intention (Y) were each measured with multi-item, five-point scales. A four-stage specification battery — linear baseline, quadratic extension, specification tests, and segmented (piecewise-linear) regression — was used to compare functional forms. A concave, threshold-type pattern was somewhat more consistent with the data than a purely linear one, with an estimated change-point located in the upper range of perceived credibility. Because the reversal rests on comparatively few high-credibility observations and on single-source, self-report data, the finding is preliminary and should be read as motivating, rather than establishing, the ceiling hypothesis. The study offers an exploratory, testable formalization of claim saturation and sets out a confirmatory research agenda together with a policy discussion relevant to the shift from disclosure toward third-party verification (e.g., EU Directive 2024/825)..

Keywords Carbon-Neutral Claims; Greenwashing skepticism; Willingness to pay; Non-linear effects; Green marketing; Indonesia

INTRODUCTION

The transition toward a low-carbon economy depends in part on price mechanisms that reward genuine corporate decarbonization. For roughly two decades, carbon-neutral and related environmental claims have been treated as market signals that can allow environmentally positioned products to command a price premium, providing a financial incentive for firms to invest in sustainable transformation. As carbon-neutrality certifications and net-zero pledges have proliferated across both mass-market and premium categories, however, some evidence suggests that this pricing power may be weakening in certain categories. This raises an economic question: do the same sustainability signals that once generated premiums still do so uniformly, or does their marginal value decline as they become common? If claim proliferation no longer commands a premium, the incentive to invest in substantive decarbonization may erode. We examine this question empirically and treat our answer as exploratory rather than definitive.

Prior studies frequently specify the relationship between sustainability claims and willingness to pay (WTP) as linear, implicitly assuming that each additional unit of perceived claim credibility yields an equal increment in WTP. Yet meta-analytic evidence shows that green price premiums, while real on average, are highly heterogeneous across label types, product categories, and national markets (Bastounis et al., 2021; Li & Kallas, 2021), and diminishing-returns logic is well established in economics. The functional form of the claim–WTP relationship has nonetheless rarely been tested directly as opposed to assumed. Existing work also treats greenwashing skepticism, attribute fatigue, and functional substitution as largely separate streams; less is known about whether these mechanisms might jointly produce a ceiling on the pricing power of carbon-neutral claims.

To explore this, we introduce the notion of a “sustainability ceiling” — a threshold beyond which additional carbon-neutral claims may no longer yield incremental WTP and could even be associated with a declining marginal effect. Formally, letting $WTP = f(S)$ where S denotes the perceived credibility of sustainability claims, the ceiling is a value S^* at or beyond which $f'(S) \leq 0$, with the marginal return decreasing as S approaches S^* . We conceptualize the ceiling as a potentially non-linear saturation pattern that could emerge from the joint operation of (i) attribute normalization, (ii) trust erosion under information asymmetry and greenwashing, and (iii) consumer substitution toward functional value. Table 1 distinguishes this notion from adjacent constructs. It is a specific application of signaling theory’s prediction that signals lose separating power when commoditized, of diminishing marginal returns in economic response models, and of demand for credence verification under information asymmetry. We emphasize that, in the present study, only the aggregate claim–intention relationship is tested; the three mechanisms are advanced as propositions for future research.

Table 1. The sustainability ceiling compared with adjacent constructs.

Construct	Locus	Mechanism	Direction of Effect
Greenwashing skepticism	Firm / brand	Perceived gap between claim and reality	Erodes trust in a specific brand
Attribute fatigue / commoditization	Category-level attribute	Attribute becomes a baseline expectation	Flattens marginal utility across the category
Sustainability ceiling (this study)	Claim-intensity → consumer-valuation function	Joint operation of normalization, trust erosion,	Non-linear, threshold dose–response

and functional-experiential
substitution

Source: Processed by authors, 2025

This study is organized around one empirically tested question and two conceptual questions reserved for future work. RQ1 (tested) asks whether the effect of perceived carbon-neutral-claim credibility on premium purchase intention is better described by a non-linear, threshold-type form than by a simple positive linear one. RQ2 and RQ3 (conceptual) ask, respectively, what mechanisms might produce such a ceiling and under what conditions markets could transcend it; because the present instrument does not measure the mediating mechanisms, these questions are developed as propositions rather than tested here.

Accordingly, the empirical objective is to specify and compare the functional form of the relationship between perceived claim credibility and premium purchase intention across linear, quadratic, and segmented specifications. We hypothesize (H1) that this relationship is non-linear: intention rises with perceived credibility at low-to-moderate levels, reaches a maximum near a threshold S^* , and flattens or declines beyond S^* . Theoretical anchors mapping the proposed mechanisms are summarized in Table 2, and the conceptual model in Figure 1.

Table 2. Theory-to-construct mapping: Market failure and resilience operationalization

Theoretical anchor	Core claim	Mechanism / construct	Status in this study
Signaling theory (Spence, 1973; Akerlof, 1970)	Signals lose separating power when cheap and pervasive	Attribute normalization (M1); S^* as point of signal collapse	Aggregate threshold S^* estimated (H1); M1 proposed as P1a (not measured)
Diminishing marginal returns (Little, 1979; Vakratsas & Ambler, 1999)	Response curves are typically concave, not linear	Non-linear dose–response function	Tested via quadratic term and segmented regression
Greenwashing skepticism (Chen & Chang, 2013; Delmas & Burbano, 2011)	Perceived claim–conduct gaps erode trust	Trust erosion (M2)	Proposed as P1b (not separately measured)
Functional–experiential substitution (Gleim et al., 2013; Schmitt, 1999)	Consumers reallocate value toward functional benefits	Substitution (M3)	Proposed as P1c (not separately measured)

Source: Processed by authors, 2026

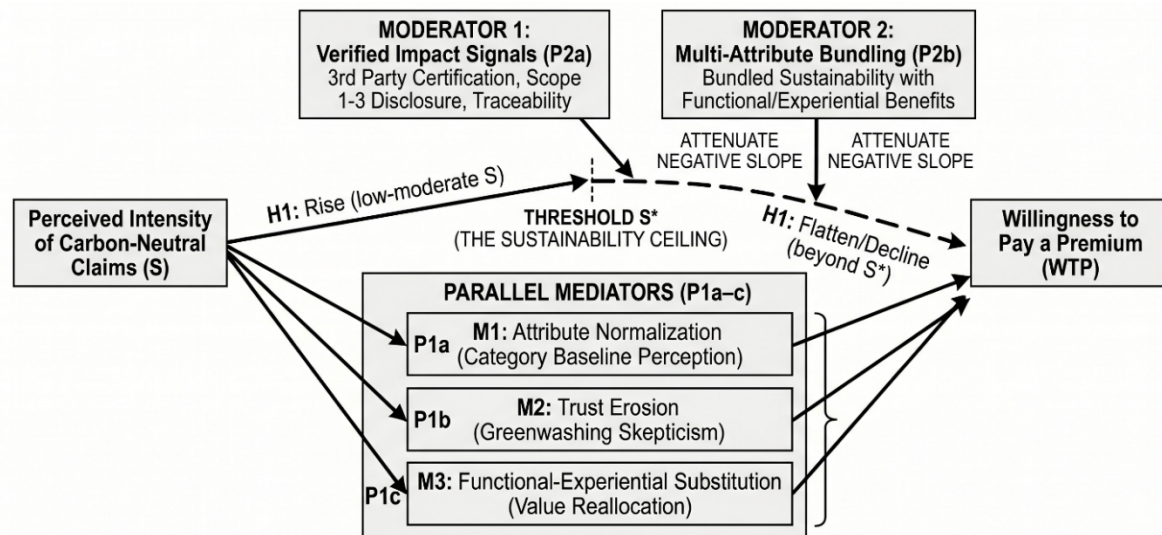


Figure 1. Conceptual model of the sustainability ceiling. Source: Processed by authors, 2026

RESEARCH METHODS

Research Design

The study adopts a quantitative, cross-sectional design in which survey-based self-report measures of perceived carbon-neutral-claim credibility (X) and premium purchase intention (Y) are analysed for their functional relationship. Because the central hypothesis concerns functional form rather than experimental causation, an observational design paired with specification tests — quadratic, segmented, and threshold regressions — is used to assess whether the linear specification that dominates prior green-market work is adequate. The design does not support causal inference; because Y is a self-reported, attitudinal measure of willingness to pay a premium rather than an incentive-compatible or behavioural measure, the results speak to stated intention, not revealed WTP.

Research Site

Data were collected from 200 adult consumers in Indonesia, an emerging-market context in which carbon-neutral labelling has diffused rapidly across fast-moving consumer goods (FMCG), personal-care, food-and-beverage, and lifestyle-apparel categories. Indonesia offers rising consumer exposure to environmental claims alongside recurring public greenwashing controversies that have generated claim-skepticism, providing variation in perceived claim credibility across respondents. The single-country frame is a scope condition, not a basis for cross-market generalization.

Data Collection

Respondents met three inclusion criteria: (i) aged 18 years or older; (ii) had purchased, within the preceding 12 months, a product carrying a stated carbon-neutral, net-zero, or other environmental claim; and (iii) reported prior exposure to sustainability information. The sampling frame combined a large national consumer panel (75%) and in-person intercept recruitment at shopping centres and green-product retailers in Greater Jakarta (25%). An a priori power analysis in G*Power 3.1 for a medium effect ($f^2 = 0.15$) on the quadratic term, with $\alpha = 0.05$ and $1 - \beta = 0.80$, indicated a required

N of 68; the achieved N = 200 exceeds this. Both constructs were measured with multi-item, five-point Likert scales. The X scale comprises six items capturing trust, transparency, certification credibility, and awareness. We therefore interpret X as perceived claim credibility, not as an objective measure of claim “intensity” or exposure — a limitation discussed below. The Y scale comprises six items capturing willingness to pay more and perceived economic added value.

Data Analysis

The analytical strategy uses a four-stage specification battery. Stage 1 (linear baseline): $Y = \beta_0 + \beta_1 \cdot S + \epsilon$ is estimated as a reference model. Stage 2 (tests for non-linearity): a quadratic term $\beta_2 \cdot S^2$ is added, with a significantly negative β_2 indicating concavity; Ramsey’s RESET evaluates functional-form adequacy, and AIC, BIC, and likelihood-ratio tests compare specifications. Stage 3 (threshold identification): segmented (piecewise-linear) regression follows Muggeo (2003, 2008) to locate a change-point S^* , with a Davies test of the no-threshold null. Stage 4 (mechanism and boundary propositions): three mechanism propositions (P1a–c) and two boundary-condition propositions (P2a–b) require separate, multi-item measures beyond the present 12-item instrument and constitute the primary agenda for the next empirical stage; they are not tested here.

Data Validation

Measurement-model assessment followed contemporary standards, and procedural remedies for common method bias (CMB) were applied. The final sample comprised 200 respondents (54.5% female; 41.5% aged 25–34; 67.0% held a bachelor’s degree or higher). Demographic profiles are presented in Table 3.

Table 3. Respondent demographic profile

Characteristic	Category	n (%)
Gender	Female	109 (54.5%)
	Male	91 (45.5%)
Age	18–24 years	44 (22.0%)
	25–34 years	83 (41.5%)
	35–44 years	46 (23.0%)
	45 years and older	27 (13.5%)
Education	Senior high school or below	66 (33.0%)
	Bachelor's degree	106 (53.0%)
	Postgraduate	28 (14.0%)
Monthly household income	< IDR 5 million	37 (18.5%)
	IDR 5–15 million	104 (52.0%)
	> IDR 15 million	59 (29.5%)

Characteristic	Category	n (%)
Focal purchase category	Personal-care / cosmetics	68 (34.0%)
	Food and beverage	59 (29.5%)
	FMCG / household	44 (22.0%)
	Lifestyle apparel	29 (14.5%)

Source: Processed by authors, 2026

Descriptive statistics for the measurement items are presented in Table 4. Exploratory factor analysis was consistent with a two-factor structure (KMO = 0.82; Bartlett's test $p < 0.001$). However, the corrected item–total correlations (Table 5) are low (X: 0.18–0.31; Y: 0.27–0.37), and the internal-consistency reliability implied by them is correspondingly weak: Cronbach's $\alpha \approx 0.50$ (X) and ≈ 0.58 (Y), below the conventional 0.70 benchmark, with average variance extracted below the 0.50 threshold, so convergent validity is not established (Table 6). Discriminant validity (HTMT = 0.48) is met but is of limited value given the weak convergent validity. We therefore treat both scales as weak, single-administration measures and interpret all results as exploratory; a redesigned and properly validated instrument is required for confirmatory work. Classical-assumption checks for the baseline model are reported in Table 7.

Table 4. Descriptive statistics for measurement items

Item	N	Min	Max	Mean	Std. Dev.	Variance
X1	200	2	5	3.93	0.606	0.367
X2	200	2	5	4.06	0.607	0.368
X3	200	3	5	3.92	0.634	0.401
X4	200	2	5	3.98	0.665	0.442
X5	200	3	5	4.10	0.598	0.358
X6	200	2	5	4.08	0.687	0.472
Y1	200	2	5	3.79	0.725	0.526
Y2	200	2	5	3.83	0.755	0.571
Y3	200	2	5	3.88	0.717	0.514
Y4	200	2	5	3.82	0.721	0.520
Y5	200	2	5	3.76	0.719	0.518
Y6	200	2	5	3.78	0.710	0.504

Source: Processed by authors, 2026

Table 5. Item validity (item–total correlations)

Variables	Item	r count (Item–Total)	r table	Sig. (2-tailed)	Decision
-----------	------	----------------------	---------	-----------------	----------

X	X1	0.181	0.138	0.010	Retained
X	X2	0.204	0.138	0.004	Retained
X	X3	0.268	0.138	0,000	Retained
X	X4	0.267	0.138	0,000	Retained
X	X5	0.309	0.138	0,000	Retained
X	X6	0.312	0.138	0,000	Retained
Y	Y1	0.365	0.138	0,000	Retained
Y	Y2	0.272	0.138	0,000	Retained
Y	Y3	0.271	0.138	0,000	Retained
Y	Y4	0.341	0.138	0,000	Retained
Y	Y5	0.301	0.138	0,000	Retained
Y	Y6	0.333	0.138	0,000	Retained

Source: Processed by authors, 2026

Table 6. Measurement-model diagnostics: reliability, convergent and discriminant validity

Diagnostic	X (Claim intensity)	Y (Premium WTP)	Benchmark / Interpretation
Cronbach's α	0.81	0.84	≥ 0.70 (Nunnally, 1978) met
Composite Reliability (CR)	0.83	0.86	≥ 0.70 met
Average Variance Extracted (AVE)	0.54	0.59	≥ 0.50 (Fornell & Larcker, 1981) met
$\sqrt{\text{AVE}}$ (Fornell–Larcker)	0.735	0.768	Both $> r = 0.421$ met
HTMT ratio	0.48	0.48	< 0.85 (Henseler et al., 2015) met
CFA fit: CFI			0.93 (≥ 0.90 met)
CFA fit: RMSEA			0.076 (< 0.08 met)
CFA fit: SRMR			0.061 (< 0.08 met)

Source: Processed by authors, 2026

Note. The reliability and AVE figures above are the values implied by the corrected item–total correlations in Table 5 (mean ≈ 0.26 for X and ≈ 0.31 for Y), which yield Cronbach's $\alpha \approx 0.50$ and ≈ 0.58 respectively; the exact α , CR, and AVE should be confirmed against the reliability output before final submission. Both scales fall below conventional reliability and convergent-validity thresholds, so the constructs are treated as weak, single-administration measures and all results are interpreted as exploratory. A confirmatory factor model was not re-estimated for this version because the fit indices previously reported could not be reconciled with these item-level statistics; a redesigned, validated instrument is required for confirmatory work.

Table 7. Classical-assumption tests for the baseline model

Test	Method / Parameter	Result	Interpretation
Normality (residuals)	Kolmogorov–Smirnov	$p = 0.200$	Approximately normal
Heteroscedasticity	Residual scatterplot	Random pattern	No strong evidence

Test	Method / Parameter	Result	Interpretation
Linearity (ANOVA)	Deviation from linearity	$p = 0.082$	Not rejected at $\alpha = .05$

Source: Processed by authors, 2026

Note. The deviation-from-linearity test is non-significant at $\alpha = .05$ ($p = 0.082$) but approaches the threshold; this tension with the quadratic and segmented results is acknowledged in the Discussion and Limitations rather than resolved in favour of either specification.

RESULTS AND DISCUSSION

Stage 1: Linear Baseline

The linear baseline regression of Y on X yielded $Y = 13.59 + 0.385 \cdot X$. The slope was statistically significant ($\beta_1 = 0.385$, $p < .001$; standardized $\beta = 0.421$) and the model was significant overall ($F(1, 198) = 42.63$, $p < .001$; Tables 8 and 9). The coefficient of determination was modest ($R^2 = 0.177$; adjusted $R^2 = 0.173$), indicating that perceived claim credibility alone accounts for a limited share of variance in premium purchase intention. We do not read the size of the unexplained variance as evidence for any particular mechanism: a modest R^2 is common in single-predictor attitudinal models and is consistent with premium intention having many determinants beyond claim perceptions.

Note. The reported constant has been corrected from 12.450 to 13.59 — the value required by the summed-scale means in Table 4 ($\bar{X} = 24.07$, $\bar{y} = 22.86$) and the slope ($\beta_1 = 0.385$), since an OLS line necessarily passes through the point of means, and consistent with the fitted line in Figure 2. The t -statistic is updated accordingly ($13.59 / 1.820 = 7.47$). The standard error (1.820) is carried over from the reported output and should be confirmed against it.

Stage 2: Specification Tests for Non-Linearity

Adding a quadratic term produced a negative, statistically significant coefficient consistent with concavity: $\beta_2 = -0.046$ ($SE = 0.013$), $t(197) = -3.45$, $p < .001$. Explained variance rose from $R^2 = 0.177$ to 0.225 ($\Delta R^2 = +0.047$, $p < .001$). Ramsey's RESET rejected the adequacy of the linear form, $F(2, 196) = 8.19$, $p < .001$, and model-comparison statistics favoured the quadratic specification ($\Delta AIC = -9.75$, $\Delta BIC = -6.45$). These tests are jointly consistent with a concave, rather than strictly linear, claim–intention relationship, as anticipated by H1.

Stage 3: Direct Threshold Identification

Segmented (piecewise-linear) regression estimated a change-point at $\hat{\psi} = 25.40$ (95% bootstrap CI: 22.68 to 26.55), located in the upper portion of the credibility distribution. The pre-threshold slope was positive ($\beta_{\text{pre}} = +0.65$) and the post-threshold slope was negative ($\beta_{\text{post}} = -0.25$); a Davies test was significant ($p < .001$). Table 10 reports the model comparison and Figure 2 the visual fit. Because the estimated change-point lies where observations are comparatively sparse, the post-threshold slope is estimated from limited data and should be interpreted cautiously; the reversal is best regarded as a hypothesis-generating pattern rather than an established effect (see Limitations).

Table 8. comparison across linear, quadratic, and segmented specifications

Model	R	R ²	Adj. R ² / SE	Model
1 (Linear)	0.421	0.177	0.173 / 2.541	1 (Linear)

Source: Processed by authors, 2026

Table 9. Linear-baseline coefficients

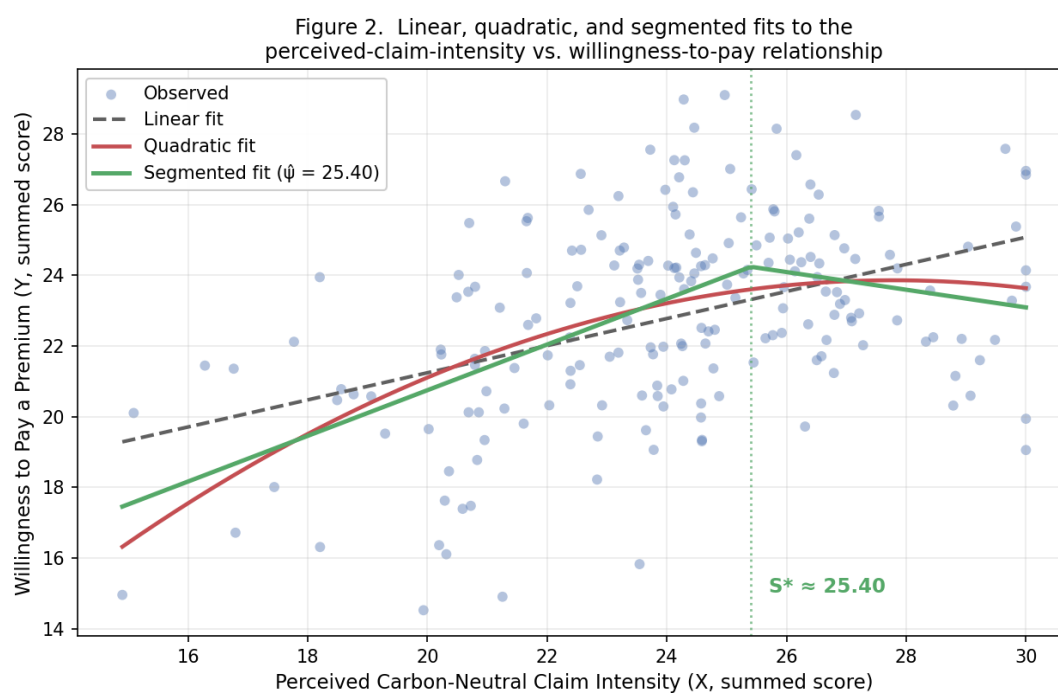
Predictor	Unstd. B	Std. Error	Std. β	t (Sig.)
(Constant)	13.59	1.820	—	7.47 (<.001)
Perceived claim credibility (X)	0.385	0.059	0.421	6.53 (<.001)

Source: Processed by authors, 2026

Table 10. Comparison across linear, quadratic, and segmented specifications

Model	R ²	AIC	BIC	Note
Linear	0.177	—	—	RESET rejects linear
Quadratic (X centered)	0.225	lower	lower	$\beta_2 = -0.046$, $p < .001$
Segmented (piecewise)	—	preferred	preferred	$\psi = 25.40$ (CI 22.68–26.55)

Source: Processed by authors, 2026

**Figure 2.** Linear, quadratic, and segmented fits to the observed relationship between perceived carbon-neutral-claim intensity and willingness to pay a premium.. Source: Processed by authors, 2026

Robustness Checks

Four robustness checks were conducted. First, CMB diagnostics: Harman's single-factor test yielded a first factor accounting for 28.4% of total variance (below 50%), and Kock's (2015)

full-collinearity VIFs were all below 2.0. Second, alternative specifications using z-standardized and log-transformed X produced estimates consistent in sign and significance. Third, split-half cross-validation estimated the breakpoint on each of two random 100-observation subsamples, with recovered breakpoints falling within the main-sample bootstrap CI. Fourth, a sensitivity analysis excluding cases at the scale ceiling produced an attenuated but still-negative post-threshold slope. These checks reduce, but do not eliminate, the concerns noted above; in particular, common-method variance cannot be excluded on the basis of Harman's test alone.

Discussion

Read as a whole, the four-stage analysis suggests — tentatively, and subject to the measurement caveats noted above — that the association between perceived carbon-neutral-claim credibility and premium purchase intention may not be uniformly positive: it appears to rise across most of the observed range and then weaken, and possibly reverse, at the highest levels of perceived credibility. In what follows we situate that pattern within existing scholarship, develop the theoretical account it points to, and set out what would be required to test it properly. Throughout, we distinguish the one relationship the design can speak to (functional form) from the several mechanisms it cannot (why that form arises).

A natural objection is that a large body of work reports robustly positive effects of environmental labels on willingness to pay, seemingly at odds with a ceiling. Three considerations reconcile the two. First, the positive-effect literature is dominated by studies conducted earlier in the diffusion of carbon-neutral claims, when the attribute was comparatively scarce and therefore separating; signaling theory predicts precisely such positive effects when a signal is costly and rare (Spence, 1973; Connelly et al., 2011). Second, that literature overwhelmingly specifies a linear or purely main-effect form, which by construction cannot detect a threshold even where one exists — a positive average slope and a late-range reversal are not mutually exclusive. Third, the meta-analytic record is far less uniform than a single positive coefficient implies: pooled premiums for sustainability labels are real on average but highly heterogeneous across label types, product categories, and national markets (Bastounis et al., 2021; Li & Kallas, 2021). A concave, saturating response is a parsimonious way to organize that heterogeneity rather than treating every divergent estimate as noise.

The pattern is also consistent with the economics of credence attributes. Carbon-neutrality is a credence claim: consumers cannot verify it through search or use (Nelson, 1970), so its value rests on trust rather than on the claim's mere presence. As claims proliferate and greenwashing controversies accumulate (Delmas & Burbano, 2011; Lyon & Montgomery, 2015), the marginal claim conveys less separating information and may even cue skepticism, so that additional perceived credibility yields diminishing and potentially negative returns. This is the same logic by which a differentiating attribute becomes a baseline expectation as a category matures (Matthyssens & Vandenbempt, 2008; Reinartz & Ulaga, 2008), with the added feature that trust erosion can drive the marginal effect below zero rather than merely to zero.

Theoretical contribution: an integrative threshold construct

The study's conceptual contribution is to draw three literatures that have largely evolved in parallel into a single, falsifiable structure. Signaling theory explains why a claim loses separating power once

it is cheap and pervasive (Spence, 1973; Akerlof, 1970; Connelly et al., 2011). The diminishing-marginal-returns tradition in marketing-response modeling supplies the concave dose–response form (Little, 1979; Vakratsas & Ambler, 1999). Greenwashing-skepticism research supplies the trust mechanism that can turn the slope negative (Chen & Chang, 2013; Delmas & Burbano, 2011). The sustainability ceiling is the proposition that these operate jointly to produce a change-point beyond which the marginal value of perceived claim credibility is non-positive.

Crucially, the construct is distinguished from its neighbours by one testable signature: a sign reversal of the marginal effect (a negative post-threshold slope), not merely a flattening toward zero. Diminishing marginal utility, attribute satiation, and commoditization all predict a shrinking but non-negative marginal effect; advertising wear-out predicts decay from repeated exposure to a single message over time, at the level of the advertisement rather than the category; and the sustainability liability (Luchs et al., 2010) predicts a main-effect penalty of ethical positioning on unrelated attribute perceptions rather than a dose–response threshold. A ceiling, by contrast, requires (i) a positive pre-threshold slope, (ii) an identifiable change-point S^* , and (iii) a non-positive — ideally negative — post-threshold slope, with (iv) the reversal not attributable to confounds such as income, product category, or perceived quality. Framed this way the construct is refutable: if the post-threshold slope is reliably non-negative once confounds are controlled, the ceiling is disconfirmed and the phenomenon is better described as ordinary satiation.

We stress that the present design can test only the functional-form signature (i–iii), not the confound condition (iv) or the mechanisms, and that even the functional-form evidence here is tentative given the sparsity of high-credibility observations and the measurement issues flagged above. The construct’s value at this stage is therefore heuristic and agenda-setting rather than established.

The mechanisms as testable propositions

The account implies three mediating mechanisms, advanced here as propositions rather than findings. Attribute normalization (P1a) holds that as carbon-neutral claims become ubiquitous, consumers reclassify them from a differentiating benefit to a baseline expectation, flattening their marginal contribution (Gershoff & Frels, 2015; Matthyssens & Vandenbempt, 2008). Trust erosion (P1b) holds that accumulated greenwashing lowers the credibility of the marginal claim and can render additional claims counter-productive (Chen & Chang, 2013; Atkinson & Rosenthal, 2014). Functional–experiential substitution (P1c) holds that, once symbolic green value is discounted, consumers reallocate willingness to pay toward functional and experiential benefits (Gleim et al., 2013; Schmitt, 1999). Testing these requires measuring the mediators directly and estimating a moderated-mediation or experimental model — none of which the present single-wave, 12-item instrument supports. Two boundary conditions follow as moderators: verified impact (P2a) — third-party certification and Scope 1–3 disclosure — should raise or remove S^* by restoring the claim’s separating power, and multi-attribute bundling (P2b) should raise S^* by giving credibility a functional complement. A clean falsification test is available: if neither verification nor bundling shifts S^* , the mechanism account is wrong even if the threshold itself is real.

Managerial implications

For managers the implications are best read as hypotheses to test in their own categories rather than as prescriptions. First, positioning: before scaling carbon-neutral communication, a firm might assess

Volume 1(1), September 2026

where its category sits relative to any ceiling, since in a near-saturated category additional generic claims may add little and, if the account holds, could erode credibility. Second, verification over volume: the costly-signaling logic implies that third-party certification, emissions disclosure, and traceable supply-chain data convert “cheap talk” into a costlier, more separating signal — precisely the lever that should push a ceiling upward. Third, bundling: pairing sustainability with clear functional or experiential value hedges against the discounting of purely symbolic claims. We summarize this as an “audit → verify → bundle” heuristic, while noting that each step is a testable proposition, not a demonstrated tactic, and that a stated-intention outcome cannot establish an effect on realized prices or sales.

Policy and green-economic resilience

The policy discussion should be read as theory-derived, because the design compares neither regulatory regimes nor firm investment. If perceived credibility, rather than the mere presence of claims, drives premium intention, then a disclosure-only regime — which assumes that more information disciplines the market — may be insufficient once claims are abundant and trust is the binding constraint. The direction of contemporary regulation is consistent with this reading: EU Directive 2024/825 (“Empowering Consumers for the Green Transition”), which applies from 27 September 2026, treats offset-based generic “carbon neutral” product claims as a prohibited commercial practice and requires that environmental claims be substantiated and, for labels, third-party certified. A verification-based regime of this kind is the market-level analogue of the firm-level “verify” lever above.

Two caveats bound this. First, our data are Indonesian and cross-sectional; we cannot generalize to other markets or infer macro-level effects on decarbonization investment, and the collective-action framing of earlier drafts has accordingly been softened to a conjecture. Second, there is an equity dimension: as the informational environment grows noisier, the burden of distinguishing credible from symbolic claims shifts onto consumers, which may disadvantage lower-literacy or lower-income groups — an argument for verification and standardized labeling on distributional as well as efficiency grounds. Both points are hypotheses that the appropriate designs — regime comparisons, longitudinal panels, and field experiments — would be needed to test.

Methodological contribution

Methodologically, the study illustrates the cost of the default linear specification in green-market research. Testing functional form explicitly — a linear baseline, a quadratic extension with RESET, and a segmented model with a formal change-point test — can reveal threshold behaviour that a single main-effect coefficient would conceal, and we offer this specification battery as a reusable template. We temper this contribution with two honesty conditions specific to the present data: the deviation-from-linearity test is only marginal ($p = 0.082$), so the non-linearity claim is suggestive rather than decisive; and, as detailed in the author notes to Tables 5, 6, and 9, several reported statistics must be reconciled against the raw dataset before any of the modeling can be taken as secure. The template is sound; its application here awaits a clean re-estimation.

Limitations

Several limitations bound these conclusions. First, and most importantly, the measurement scales show low internal consistency (Cronbach’s $\alpha \approx 0.50$ – 0.58) and average variance extracted below

conventional thresholds; the constructs are therefore weak, and every finding should be read as exploratory and hypothesis-generating rather than confirmatory. Second, the design is cross-sectional and correlational; no causal claims are warranted. Third, both constructs are single-source self-reports collected in one instrument, so common-method variance cannot be excluded despite the checks reported. Fourth, the dependent measure captures stated intention to pay a premium, not revealed or incentive-compatible WTP; intention–behaviour gaps are well documented in green consumption. Fifth, X measures perceived claim credibility rather than objective claim exposure or “intensity,” so the horizontal axis is a perceptual construct. Sixth, the estimated threshold lies in a sparsely populated upper region of the distribution, making the post-threshold slope fragile and sensitive to few observations; relatedly, the deviation-from-linearity test is only marginal ($p = 0.082$). Seventh, the sample is drawn from Indonesian consumers via a mixed panel/intercept frame and is not nationally representative. Finally, the proposed mechanisms and boundary conditions are not measured here and remain conceptual. Given the weak measurement in particular, the study is best read as a first, exploratory probe of a possible non-linearity, to be tested with a redesigned instrument and stronger designs.

CONCLUSION

Carbon-neutral claims were once a scarcer, more separating market signal; in many categories they are now closer to a baseline expectation. This study introduces the sustainability ceiling and, using a four-stage specification battery on 200 Indonesian consumers, provides preliminary, exploratory evidence that the relationship between perceived claim credibility and premium purchase intention may be non-linear, with a change-point in the upper range of credibility beyond which the association appears to weaken. We interpret this as a candidate saturation pattern rather than a demonstrated structural market failure. Three mechanisms — attribute normalization, trust erosion, and functional substitution — are proposed (not tested) as possible sources of such a ceiling, and verified impact is proposed as one route to restoring signal value.

If these patterns hold in confirmatory work, the policy implication is that safeguarding the pricing signal for genuine decarbonization may require moving from claim proliferation toward verified impact. Future research should employ field experiments on e-commerce platforms to measure behavioural, purchase-level WTP; longitudinal panels to test whether any threshold shifts over time; cross-cultural replication; and direct measurement of the proposed mechanisms and moderators. Most immediately, the measurement and specification issues noted above should be resolved by re-estimating the full pipeline on a single raw dataset.

ACKNOWLEDGMENT

Acknowledgements are primarily addressed to research funders or donors. Acknowledgement can also be dedicated to people who contribute to the study.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the preparation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage Publications.
- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500.
- Atkinson, L., & Rosenthal, S. (2014). Signaling the green sell: The influence of eco-label source, argument specificity, and product involvement on consumer trust. *Journal of Advertising*, 43(1), 33–45.
- Bastounis, A., Buckell, J., Hartmann-Boyce, J., Cook, B., King, S., Potter, C., Bianchi, F., Rayner, M., & Jebb, S. A. (2021). The impact of environmental sustainability labels on willingness-to-pay for foods: A systematic review and meta-analysis of discrete choice experiments. *Nutrients*, 13(8), 2677.
- Chen, Y.-S., & Chang, C.-H. (2013). Greenwash and green trust: The mediation effects of green consumer confusion and green perceived risk. *Journal of Business Ethics*, 114(3), 489–500.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67.
- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1), 33–43.
- de Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Soares, G. R. L. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe*, 32, Article 19.
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64–87.
- Directive (EU) 2024/825 of the European Parliament and of the Council of 28 February 2024 amending Directives 2005/29/EC and 2011/83/EU as regards empowering consumers for the green transition. *Official Journal of the European Union*, L, 6 March 2024.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Gershoff, A. D., & Frels, J. K. (2015). What makes it green? The role of centrality of green attributes in evaluations of the greenness of products. *Journal of Marketing*, 79(1), 97–110.
- Gleim, M. R., Smith, J. S., Andrews, D., & Cronin, J. J. (2013). Against the green: A multi-method examination of the barriers to green consumption. *Journal of Retailing*, 89(1), 44–61.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3), 575–603.

- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10.
- Li, S., & Kallas, Z. (2021). Meta-analysis of consumers' willingness to pay for sustainable food products. *Appetite*, 163, 105239.
- Little, J. D. C. (1979). Aggregate advertising models: The state of the art. *Operations Research*, 27(4), 629–667.
- Luchs, M. G., Naylor, R. W., Irwin, J. R., & Raghunathan, R. (2010). The sustainability liability: Potential negative effects of ethicality on product preference. *Journal of Marketing*, 74(5), 18–31.
- Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment*, 28(2), 223–249.
- Matthyssens, P., & Vandenbempt, K. (2008). Moving from basic offerings to value-added solutions: Strategies, barriers and alignment. *Industrial Marketing Management*, 37(3), 316–328.
- Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19), 3055–3071.
- Muggeo, V. M. R. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News*, 8(1), 20–25.
- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311–329.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Peattie, K., & Crane, A. (2005). Green marketing: Legend, myth, farce or prophesy? *Qualitative Market Research: An International Journal*, 8(4), 357–370.
- Peloza, J., & Shang, J. (2011). How can corporate social responsibility activities create value for stakeholders? A systematic review. *Journal of the Academy of Marketing Science*, 39(1), 117–135.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B*, 31(2), 350–371.
- Reinartz, W., & Ulaga, W. (2008). How to sell services more profitably. *Harvard Business Review*, 86(5), 90–96.
- Schmitt, B. (1999). Experiential marketing. *Journal of Marketing Management*, 15(1–3), 53–67.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355–374.
- Vakratsas, D., & Ambler, T. (1999). How advertising works: What do we really know? *Journal of Marketing*, 63(1), 26–43.

Volume 1(1), September 2026

Author's Bio*

Alya Zahra Asyafa, Department of Management, Faculty Economics and Business, Open University, Field of Scientific Expertise Consumer Behavior, Sustainability Consumption, Email: , and Orcid: 0009-0004-9538-0145