sustainability

MDPI

Article

Environmental Claim Reconfiguration Around the EU Anti-Greenwashing Transposition Deadline: Evidence from a 36-Wave EU–UK Digital Retail Panel

Munise Hayrun Sağlam [1,*] and Oğuzhan Aygören [2]

[1] Department of Business Administration, Yıldız Technical University, Istanbul 34220, Turkey
[2] Institute for Business Innovation, Haas School of Business, University of California, Berkeley, CA 94720-1930, USA; oguzhan@berkeley.edu
* Correspondence: mhsaglam@yildiz.edu.tr

Abstract

Environmental claims shape online product evaluation despite limited verifiability. This study examines whether claim architecture changed around the 27 March 2026 EU transposition deadline and whether adjustment varied by certification and baseline seller type. A prospective 36-wave quasi-experimental panel followed 357 EU Ecolabel-certified products and 357 matched comparison products across Amazon marketplaces in France, Germany, Italy, and the United Kingdom, yielding 96,784 validated product-market-wave observations. Multilingual content analysis measured generic, qualified, and evidence-bearing claims as separate, potentially co-occurring indicators. Triple-difference and event-study models with product-market and capture-wave fixed effects estimated deadline-associated EU–UK changes. Relative to the United Kingdom, generic claims declined by 8.7 percentage points among comparison products and 3.1 points among certified products; qualified and evidence-bearing claims increased in each stratum. Verified EU Ecolabel display rose by 5.1 points among certified products, and generic-claim withdrawal was stronger for marketplace-retail offers. The findings distinguish withdrawal from signal upgrading and extend signalling analysis to sender revision around a transposition deadline. The smaller decline among certified products is consistent with greater evidentiary capacity, while seller-type contrasts indicate heterogeneity by baseline organisational setting. The estimates capture anticipatory marketplace adjustment rather than enactment or enforcement effects.

Keywords: environmental claims; greenwashing regulation; signalling theory; EU Ecolabel; digital retail; triple-difference design; marketplace governance

Check for updates

Academic Editor: Bernardino Benito López

Received: 14 August 2026
Revised: 28 August 2026
Accepted: 1 September 2026
Published: 3 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Environmental claims are now routine in digital retail, yet their informational value remains uncertain. Terms such as “green”, “eco-friendly” and “sustainable” can shape product evaluation at the point of choice, although consumers cannot readily verify the attributes to which they refer [1]. Evidence reported by the European Commission indicates that 53.3% of examined environmental claims were vague, misleading or unfounded, and 40% were unsubstantiated. Half of the green labels assessed offered weak or non-existent verification [2]. On e-commerce pages, claims, logos and technical details compete for attention within a single selling interface. General environmental wording can consequently influence choice before consumers reach the evidence needed to evaluate it.

This verification problem depends partly on how environmental information is framed. Abstract environmental compensation claims generate stronger perceptions of greenwashing than concrete claims, and perceived greenwashing weakens purchase intentions [3]. In fashion e-commerce, specific sustainability labels can reduce ambiguity and strengthen trust relative to general labels, although their effects vary across retailer categories [4]. Visual sustainability labels can improve consumer perceptions and behaviour, but their effectiveness depends on design, purchase context and consumers' ability to interpret the signal [5]. Label proliferation introduces a further source of uncertainty. Experimental evidence from Germany and the United States shows that meta-labels facilitate sustainable choices only under particular market and consumer conditions [6]. The credibility of environmental communication therefore depends on claim precision, accessible evidence and the institutional source of verification.

The evidentiary weaknesses identified in this literature have become a matter of consumer protection law through Directive (EU) 2024/825. It prohibits generic environmental claims when a trader cannot demonstrate recognised excellent environmental performance relevant to the claim. It also restricts sustainability labels unless they are based on a certification scheme or established by a public authority. Member States were required to transpose these provisions by 27 March 2026 and to apply the transposing provisions from 27 September 2026 [7]. The analysis treats 27 March as a common, predetermined EU transposition deadline rather than the date of simultaneous national enactment or harmonised enforcement. Changes around the deadline are therefore interpreted as anticipatory product-page adjustment.

Regulatory preparation may begin before formal application and need not produce wholesale silence. Longitudinal evidence from large hotel groups shows that report quality, materiality orientation and transparency can improve even when reporting volume changes little [8]. Evidence from European accommodation markets further indicates that certification decisions under regulatory uncertainty vary with organisational resources and strategic capacity [9]. Firms managing digital product content can remove a weak claim, narrow its scope, provide supporting evidence or increase the visibility of recognised certification. EU Ecolabel certification places certified products in a distinct evidential position by providing a recognised basis for demonstrating environmental performance. Generic language should therefore recede after the transposition deadline, whereas qualified and evidence-bearing communication should become more prevalent. Certified products may display a weaker withdrawal response because their environmental positioning rests on a defensible external signal.

The capacity to implement these revisions also depends on how product-page content is governed. Amazon operates as a retailer within the marketplace that it provides to independent sellers. Marketplace-retail offers and third-party offers consequently operate under different modes of commercial control [10]. Centralised marketplace-retail governance may support more consistent claim revision, whereas third-party responses can vary with seller capacity and compliance resources. Adjustment around the transposition deadline is therefore expected to differ by baseline seller type. Seller classification is treated as an indicator of the offer's organisational setting rather than proof that the displayed claim was authored by the active seller.

These predicted responses remain largely untested on live product pages, despite their policy relevance. A systematic methodological review found that only five of 127 empirical greenwashing studies examined real rather than hypothetical cases [11]. Consumer experiments, periodic disclosure studies and legal analyses therefore provide limited insight into how environmental communication changes on live product pages as regulation develops. Consumer studies usually estimate responses to experimentally varied claims or labels [12].

Disclosure research examines periodic organisational reports, while legal analysis establishes the obligations created by new rules [13–15]. Label research has also devoted limited attention to the interaction between certification and external market conditions [5]. It remains unclear whether anti-greenwashing regulation is followed by claim withdrawal, greater substantiation or a redistribution of visibility towards certified products. Extensive withdrawal may reduce misleading wording yet also suppress defensible environmental information. Measuring claim presence alone cannot distinguish these responses.

To address this field-evidence gap, the study examines whether environmental-claim architecture changed around the 27 March 2026 EU transposition deadline, whether certified products exhibited a weaker withdrawal response, and whether adjustment varied by baseline seller type. It uses a 36-wave panel of Amazon product pages in France, Germany, Italy and the United Kingdom. The panel follows 357 EU Ecolabel-certified products and 357 matched comparison products from 16 February to 18 June 2026, comprising 12 pre-deadline and 24 post-deadline waves. After validation, the dataset contains 96,784 product-market-wave snapshots. France, Germany and Italy constitute the EU-exposed group, while the United Kingdom provides a non-EU comparator that was not subject to the Directive's transposition deadline. A triple-difference design compares changes across time, market exposure and EU Ecolabel status. Seller-type heterogeneity is estimated separately using baseline classifications. The analysis distinguishes generic, qualified and evidence-bearing claims, verified-label display, claim count and prominence. By conceptualising deadline-associated adjustment as a change in claim architecture, the study moves beyond the binary question of whether firms communicate environmental attributes. It provides repeated field evidence on how environmental information changed around a common EU transposition deadline.

2. Theoretical Background and Hypothesis Development

2.1. Environmental Claims as Signals Under Information Asymmetry

Environmental performance is rarely observable at the digital point of sale. Consumers can inspect a product's price, appearance and ratings, but they cannot directly assess the environmental effects of its materials, production or disposal. Many environmental attributes therefore remain credence qualities whose validity depends on information supplied by sellers or independent institutions [16]. Environmental labels can reduce this asymmetry by converting otherwise opaque attributes into visible cues, although their value depends on whether consumers understand and trust the information conveyed [17].

Signalling theory explains how such cues operate when one party knows more about product quality than another [18]. The seller or brand possesses information about the basis of an environmental claim, while consumers observe only the message displayed on the product page. A signal becomes informative when it is visible and more difficult for a low-quality sender to imitate [19]. Its cost may arise before communication through measurement, documentation and verification. Further costs may arise if an inaccurate signal is challenged by consumers, platforms or regulators. Contemporary signalling research identifies this differential cost as the central mechanism separating credible signals from inexpensive assertions [20].

Environmental claims differ markedly along this dimension. A generic claim such as "green" or "eco-friendly" makes a broad promise without identifying the relevant attribute, comparison or evidential basis. It is inexpensive to produce, flexible in meaning and difficult to evaluate against a precise standard. These features make it accessible to sellers whose environmental performance may not support a narrower statement. Its persuasive reach can therefore exceed its diagnostic value [1].

A qualified claim accepts a more limited and testable commitment. It identifies the product component, environmental attribute or condition to which the statement applies. Phrases referring to recycled packaging, reduced material use or a defined production practice constrain how the claim can be interpreted. This narrower scope increases the effort required to formulate the message and exposes inconsistencies more readily. Recent experimental evidence indicates that precise green claims generate different consumer responses from vague claims because they provide clearer grounds for evaluation [21].

An evidence-bearing claim goes further by connecting the environmental statement to observable support, such as a quantitative measure, named standard, accessible documentation or verifiable assessment. The claim is more costly to maintain because the supporting information must exist, remain traceable and correspond to the product being described. Assurance research shows that credibility depends on the fit between the communicated signal and the depth of its supporting basis; visible assurance alone does not make a weakly grounded statement reliable [22]. Textual evidence from corporate climate communication also suggests that operationally specific language is more closely aligned with subsequent performance than reactive or symbolic rhetoric [23].

Generic, qualified and evidence-bearing claims consequently form an ordered claim architecture. They differ in scope, falsifiability and the resources required to support them. This distinction avoids treating environmental communication as a binary choice between claiming and remaining silent. A product page may preserve its environmental positioning while moving from an easily imitated signal towards one that is narrower or better supported. Regulation can alter this architecture by changing the relative cost of each signal type.

2.2. Regulatory Pressure and Claim Architecture

Directive (EU) 2024/825 raises the expected legal and reputational cost of environmental signals that convey broad persuasive meaning without a defensible basis [7]. The increase should be strongest for generic claims because their unrestricted scope creates the largest gap between consumer inference and substantiation. Firms' incentives to greenwash are also shaped by market and organisational pressures [24].

Sellers have several ways to reduce this exposure. They may remove environmental language, restrict the scope of a claim or connect it to supporting evidence. Formal signalling models show that stronger penalties for inaccurate environmental claims can discourage weak signals, although firms may choose silence when credible communication remains too costly [25]. Regulatory pressure should therefore affect the composition of communication rather than produce a uniform decline in every form of environmental expression.

The 27 March 2026 transposition deadline provides a salient point around which firms could reassess existing content. Regulatory preparation can begin before formal application because firms need time to review disclosures, develop internal controls and anticipate stakeholder scrutiny. Evidence from the EU's earlier non-financial reporting directive shows that affected firms adjusted their CSR activities and reporting practices before the mandate entered into force [26]. Mandatory sustainability disclosure can influence behaviour through preparation costs, stakeholder attention, peer comparison and the expected consequences of future non-compliance [27]. The March deadline may therefore operate as a coordination point for anticipatory product-page revision, even though the substantive provisions apply from September.

Generic claims offer the clearest target for early revision. Their low specificity makes them difficult to defend once the evidential threshold becomes more prominent. Product pages in France, Germany and Italy were exposed to the EU transposition process,

whereas the United Kingdom experienced no equivalent regulatory milestone during the observation window. This difference leads to the first hypothesis:

H1. *Following the 27 March 2026 transposition deadline, the probability of generic environmental claims will decline more in EU-exposed markets than in the United Kingdom.*

Withdrawal is not the only plausible response. Environmental positioning retains commercial value, giving sellers an incentive to preserve communication where a narrower statement can be defended. Replacing a generic claim with a qualified claim reduces the breadth of the promise while retaining a visible environmental attribute [28]. The resulting message is more precise and less likely to imply benefits beyond the stated product feature. This form of revision represents signal refinement rather than communication avoidance.

H2a. *Following the 27 March 2026 transposition deadline, the probability of qualified environmental claims will increase more in EU-exposed markets than in the United Kingdom.*

Some sellers may strengthen the claim further by displaying evidence that consumers, platforms or enforcement bodies can inspect. Evidence-bearing communication imposes greater documentation and maintenance costs, but it offers a more defensible basis for retaining environmental visibility. Its relative attractiveness should rise when unsupported wording becomes riskier.

H2b. *Following the 27 March 2026 transposition deadline, the probability of evidence-bearing environmental claims will increase more in EU-exposed markets than in the United Kingdom.*

These hypotheses distinguish withdrawal from signal upgrading. A reduction in generic claims without growth in more defensible forms would indicate communication retreat. A concurrent increase in qualified or evidence-bearing claims would instead show that regulatory pressure restructured the evidential content of environmental communication. The hypotheses concern adjustment around the transposition deadline and do not interpret that date as the beginning of harmonised enforcement.

2.3. Certification as an Evidentiary Resource

Regulatory pressure does not affect every product equally because sellers enter the transition with different evidentiary resources. Certification is a costly signal whose credibility derives from external assessment and defined eligibility criteria. It can reduce information asymmetry by linking an environmental assertion to a verification process outside the seller's immediate control. Recent evidence confirms that organisations use certification to strengthen reputation and communicate otherwise difficult-to-observe attributes [29]. Formal analysis likewise shows that sufficiently precise environmental certification can separate higher-quality products from conventional alternatives, although weak certification systems may still permit pooling or manipulation [30].

The EU Ecolabel provides a particularly relevant evidentiary basis. Its criteria are product-category specific, science-based and independently verified. The label is intended for products with high environmental performance within their category [31]. Certification therefore supplies a recognised symbol, a defined claim scope and supporting documentation. For products certified before the regulatory milestone, much of the investment required to establish environmental performance has already been incurred.

This prior investment should lower the marginal cost of revising a product page because the certification process has already required firms to document compliance, obtain test results and complete verification [32]. A certified seller can then reuse these

materials by replacing an unsupported expression with language tied to the certification criteria, displaying the verified label more clearly or providing evidence generated during certification. A matched comparison product without the focal certification cannot draw on the same established basis. Its seller must produce new evidence, narrow the claim to an independently defensible attribute or remove the environmental language.

Certification does not validate every statement attached to a product. Broad claims may remain problematic when they exceed the certified attribute or product scope. Certified products may therefore experience substantial revision after the transposition deadline. Their adjustment should, however, rely more heavily on qualification and substantiation, reducing the need for complete communication withdrawal. This leads to the following hypothesis:

H3. *Following the 27 March 2026 transposition deadline, the probability of generic environmental claims will decline less for EU Ecolabel-certified products than for matched comparison products in EU-exposed markets, relative to the United Kingdom.*

2.4. Seller Governance and Capacity for Coordinated Revision

Claim adjustment also depends on how offers are governed within the marketplace. Amazon combines a marketplace-retail channel, commonly described as first-party retail, with a third-party channel in which independent sellers transact through the platform. These arrangements assign different operational roles to the platform and participating firms [33]. Amazon's position as marketplace operator and first-party seller makes this distinction visible within the same digital environment [34].

Responding to a regulatory milestone requires sellers to identify exposed claims, evaluate their defensibility, approve revisions and implement changes across product pages. Marketplace-retail operations can coordinate these activities through centralised compliance systems and common content rules. Control over independent sellers is more contingent because it depends on how readily seller behaviour can be programmed, observed and evaluated [35]. This governance difference should support more consistent revision across a centrally managed product portfolio. The expected response is therefore faster contraction of generic language and more systematic replacement with qualified or evidence-bearing content.

Third-party offers are governed by sellers that vary in size, legal knowledge, resources and dependence on the marketplace. Platform-dependent businesses operate within rules and technical systems that they do not fully control, creating persistent asymmetries between the platform and participating entrepreneurs [36]. Some third-party sellers may respond quickly, while others may interpret the regulatory signal later or revise claims less consistently. The predicted difference concerns coordination capacity rather than deliberate non-compliance.

Seller classification does not establish who originally authored the environmental wording because several sellers can be associated with a shared product page. It identifies the organisational setting of the active offer and the governance structure through which page-level adjustment is likely to occur. The proposed mechanism therefore concerns differences in coordinated response across seller settings.

H4. *Following the 27 March 2026 transposition deadline, the probability of generic environmental claims will decline more for marketplace-retail offers than for third-party offers in EU-exposed markets, relative to the United Kingdom.*

3. Methodology

3.1. Research Design and Regulatory Setting

This study used a prospective quasi-experimental panel to estimate whether environmental-claim architecture shifted in EU marketplaces relative to the United Kingdom around the 27 March 2026 transposition deadline. The observational unit was a product-country-wave record. Identical products were followed on Amazon's national marketplaces in France, Germany, Italy and the United Kingdom from 16 February to 18 June 2026. Product pages were captured every Monday and Thursday, producing 12 pre-deadline waves and 24 post-deadline waves. France, Germany and Italy constituted the EU-exposed markets, while the United Kingdom served as the non-EU comparator. The target estimands were the post-deadline EU–UK contrast among matched comparison products and the additional certification differential.

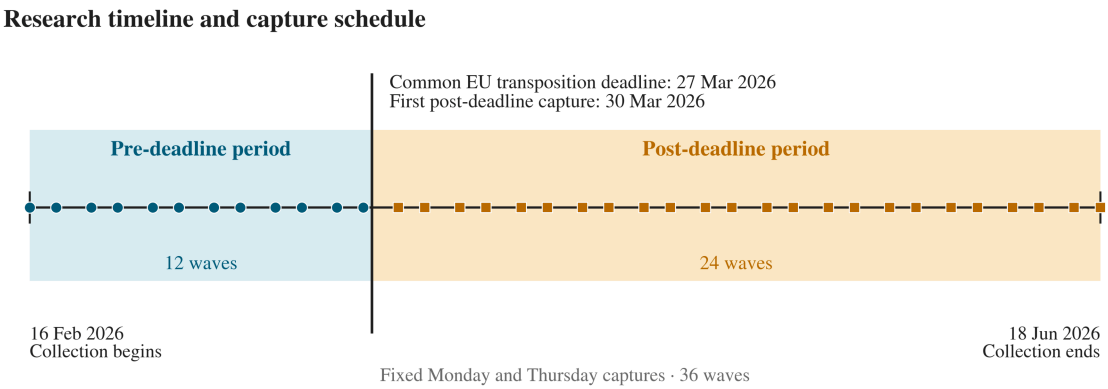
Member States were required to transpose Directive (EU) 2024/825 by 27 March 2026, while the national provisions apply from 27 September 2026 [7]. National transposition did not occur simultaneously. Germany's implementing statute was promulgated on 19 February 2026, while Italy's legislative decree entered into force on 24 March 2026. France had not notified complete transposition by 28 May 2026 and remained in the legislative process during the observation window [37–39]. The cutoff therefore represents a shared EU transposition deadline that may have coordinated anticipatory revision, rather than a common date of national enactment or enforcement.

France, Germany, and Italy were selected because they were among the largest EU Ecolabel markets in 2026 and provided sufficient cross-market product overlap [31]. Restricting the study to parallel national domains of one marketplace held page architecture, fulfilment cues, and core merchandising modules relatively constant. The United Kingdom was not treated as an unregulated market. Its Competition and Markets Authority had maintained the Green Claims Code since 2021 and issued supply-chain guidance in January 2026 [40]. No comparable UK rule entered application during the observation window. The United Kingdom therefore provides a contemporaneous comparison under a stable, pre-existing claims regime. Comparator validity rests on parallel counterfactual changes rather than regulatory equivalence. The design therefore estimates post-deadline EU–UK differences around the common EU milestone; it does not identify the enforcement effect of a uniform national intervention. Weekly event-time indicators were defined relative to 27 March 2026 to assess the timing of anticipatory revision and post-deadline adjustment. Figure 1 summarises the fixed capture schedule and the triple-difference comparison. This structure separates the certification differential in deadline-associated adjustment from the higher baseline claim intensity observed among certified products.

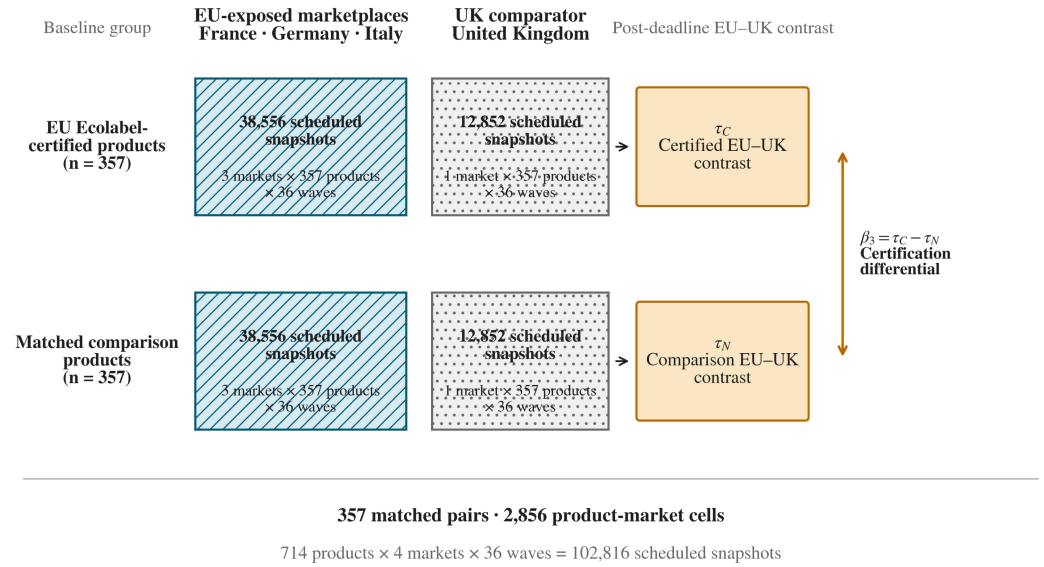
3.2. Data Sources and Panel Construction

The dataset combined official certification records with non-reactive digital trace data. EU Ecolabel status was obtained from the European Commission's Product Catalogue through its open-data interface. The catalogue provides licence identifiers, product groups, licence holders, Global Trade Item Numbers (GTINs), and declared countries of availability; it is designed to support verification by consumers and online retailers [41]. Certification strata were defined using the EU Ecolabel registry snapshot dated 9 February 2026. Comparison-product eligibility required the absence of a registry-linked EU Ecolabel record on that date; it did not imply the absence of alternative certifications, which were not comprehensively registry-verified at baseline or during the panel and therefore could not support a consistent exclusion sensitivity analysis. Baseline certification assignment was held fixed throughout the confirmatory analysis to avoid conditioning group membership on a post-baseline event. Any subsequent change in registry status was retained as an

audit flag and did not trigger reclassification or rematching. Verified EU Ecolabel display remained a time-varying page-level outcome coded at every capture wave.



(**a**)



(**b**)

Figure 1. Research Design and Identification Strategy. (**a**) Research timeline and capture schedule, comprising 12 pre-deadline and 24 post-deadline waves around the 27 March 2026 transposition deadline. (**b**) Triple-difference comparison structure across the EU-exposed marketplaces and the United Kingdom within EU Ecolabel-certified and matched comparison products. τN and τC denote the post-deadline EU–UK contrasts for comparison and certified products, respectively; β3 = τC − τN is the certification differential. Blue and beige backgrounds denote the pre- and post-deadline periods, while horizontal and vertical arrows denote the EU–UK contrasts and certification differential, respectively.

A certified product entered the sampling frame when its GTIN, brand, model or variant, pack size, and active offer could be reconciled across the four marketplaces at baseline. Of the 512 eligible certified products, sequential screening excluded 155. The exclusions comprised 61 products with no comparison product within the exact subtype, 42 whose minimum-distance match exceeded the calliper, 27 with unresolved comparison-product identity, 15 whose proposed comparison candidate had a registry-linked EU

Ecolabel record, and 10 for which no unique optimal match remained after the restrictions. The remaining 357 certified products were paired one-to-one with comparison products that had no corresponding EU Ecolabel record. Comparison products were drawn from the same marketplace category nodes. Optimal one-to-one matching without replacement was exact on product category and subtype, then minimised distance on brand segment, pack size, pre-deadline median price, star rating, log review count, and seller type. A 0.20 pooled-standard-deviation calliper was imposed, and the prespecified balance criterion was an absolute standardised mean difference below 0.10 [42]. Matching achieved exact balance on category and subtype. The median absolute post-match SMD was 0.041, and the maximum was 0.074; every post-match SMD remained below the prespecified 0.10 threshold. The final cohort comprised 149 cleaning-product pairs, 113 personal-care pairs, and 95 textile pairs, giving 714 products. Table 1 reports the panel construction and analytical yield; Supplementary Tables S1 and S2 provide the screening and matching diagnostics.

$$N_{\text{scheduled}} = N_{\text{products}} \times N_{\text{markets}} \times N_{\text{waves}} = 714 \times 4 \times 36 = 102{,}816$$

Each wave used fixed, signed-out browser profiles with native language, currency, and a constant central-city delivery location. Marketplace order and product order were randomised within each wave. The collector archived the source HTML, rendered document object model, full-page image, canonical and resolved URLs, UTC timestamp, response status, GTIN or marketplace identifier, price, seller type, rating, review count, and every claim-bearing page module. Raw files received SHA-256 hashes. Data collection used a frozen, version-controlled computational environment, and structured records were written to versioned Parquet files. Full software versions and configuration details are documented in Supplementary S2 and the replication package.

Table 1. Panel construction and analytical yield.

Construction Stage	Retained Units	Decision Rule
Eligible certified products	512 products	Exact cross-market identity and active baseline offer
Matched analytical cohort	357 pairs; 714 products	One-to-one optimal matching within product subtype
Product-market cells	2856 cells	Each product observed in four national marketplaces
Scheduled panel	102,816 records	714 products × 4 markets × 36 waves
Validated text panel	96,784 records (94.1%)	Usable rendered page with verified product identity

Note. Missing page content was never coded as claim absence. Unavailable pages remained in the full market-availability panel.

The full availability panel retained all 102,816 scheduled records. Claim-text analyses used 96,784 validated snapshots, comprising 32,292 pre-deadline and 64,492 post-deadline observations. The remaining 6032 records represented 4281 unavailable or delisted pages, 1079 redirects to a different GTIN or variant, and 672 incomplete renders after two timed retries. These states were logged separately without imputing textual values. All sampled products remained in the cohort, averaging 33.9 valid waves per product-market cell. A balanced-panel specification was used only for robustness analysis. Identity checks, retry logs, immutable raw captures and explicit missing-state codes addressed measurement and platform-selection errors in digital trace research [43]. No review text, usernames, purchase histories, account information or other person-level data were collected. Supplementary Table S6 reports the missing-state composition.

3.3. Measurement and Claim Classification

Measurement followed the regulatory distinction between generic and specified environmental communication. A claim unit was the shortest visible textual or graphical

span that conveyed one environmental proposition. Text from titles, bullet points, descriptions, enhanced-content modules, technical fields, image alt text, and OCR-detected image text was segmented at sentence or list-item boundaries. Repeated modules with identical content hashes were retained at the snapshot level for panel estimation but coded once, preventing duplicated text from dominating model training. Language-specific normalisation preserved numbers, units, negation, licence identifiers, and comparative baselines.

The primary outcome, generic claim, equalled one when a snapshot contained a broad environmental assertion, such as green, eco-friendly, ecological, sustainable, or climate-friendly, without a visible qualifying attribute in the same claim-bearing module. French, German, and Italian equivalents were coded under the same rule. A qualified claim identified the environmental attribute and at least one material boundary: quantified content, comparator, product component, life-cycle stage, condition of use or disposal, or verification source. An evidence-bearing claim required a traceable licence, standard, test, dataset, or supporting source accessible from the page. EU Ecolabel display was coded separately and treated as verified only when the logo or licence number matched the official registry. These measures describe marketplace communication; they do not assign legal compliance. Table 2 summarises the outcome definitions, and Supplementary Table S3 provides multilingual examples. The complete claim-classification codebook is provided in Supplementary S1.

Table 2. Outcome definitions and coding scales.

Outcome	Operational Definition	Scale
Generic claim (primary)	Broad environmental assertion without a qualifying attribute in the same module	Binary
Qualified claim	Environmental attribute plus a quantity, comparator, scope, condition, or verification source	Binary
Evidence-bearing claim	Traceable licence, standard, test, dataset, or accessible supporting source	Binary
Verified label display	EU Ecolabel logo or licence number reconciled to the official product registry	Binary
Claim count	Distinct environmental propositions after within-snapshot deduplication	Count
Prominence	Highest location: absent; lower page; first viewport or bullets; title or hero image	0–3

Note. The generic, qualified, and evidence-bearing indicators are coded independently and may co-occur within a single snapshot. Their frequencies and percentages should therefore not be interpreted as mutually exclusive or summed to 100%.

Concrete examples clarify the coding boundary. Qualified claims included “Bottle made with 75% recycled plastic”, “Flacon composé de 75% de plastique recyclé”, “Flasche aus 75% recyceltem Kunststoff” and “Flacone realizzato con il 75% di plastica riciclata”. These statements identify an environmental attribute and a quantitative boundary but do not direct the reader to supporting documentation. Evidence-bearing claims included “Test report available through the product link”, “Rapport d’essai disponible via le lien produit”, “Prüfbericht über den Produktlink abrufbar” and “Rapporto di prova disponibile tramite il link del prodotto”. These statements provide a traceable route to an accessible evidential source.

At the snapshot level, the three binary outcomes were not mutually exclusive. A page could contain separate generic and qualified claims, and an evidence-bearing claim could also satisfy the qualification rule.

Baseline seller type was defined at the product-market level from the modal winning-offer seller across the 12 pre-deadline waves. A cell was classified as marketplace retail when the marketplace itself was the modal seller and as third-party when the winning offer belonged to an independent seller. Marketplace fulfilment of a third-party offer did not alter this classification. Seller type was fixed before the transposition deadline to avoid conditioning on post-deadline changes in offer ownership.

A high-recall multilingual lexicon generated candidate units. A locally fine-tuned XLM-RoBERTa-large model then assigned the generic, qualified, and evidence labels. Label images were identified through OCR, perceptual hashing, and registry matching. The

gold-standard corpus contained 9600 units, with 2400 from each market, stratified by product category, certification, claim type, and period. Four bilingual coders applied a shared codebook. Coders received only the claim units; product identity, EU Ecolabel status, and capture-period metadata were withheld, although textual content could reveal certification cues. Each unit received two independent decisions, and a senior coder adjudicated disagreements. Before splitting, exact and near-identical units across products were grouped by normalised content hashes and character n-gram similarity, including repeated brand and template language. Grouped product-level splitting allocated 6720 units to training, 1440 to validation, and 1440 to a locked test set; no product or text-similarity cluster crossed partitions [44].

Krippendorff's alpha ranged from 0.81 to 0.92 across the binary labels; weighted kappa for prominence was 0.86. On the locked test set, precision, recall, and F1 were 0.92, 0.90, and 0.91 for generic claims; 0.89, 0.86, and 0.87 for qualified claims; 0.88, 0.83, and 0.85 for evidence-bearing claims; and 0.96, 0.94, and 0.95 for verified-label display. Macro-F1 was 0.90. Decision thresholds were selected on validation data and fixed before panel prediction. Per-language performance and complete-page audits were retained because classifier validity can shift across domains and languages [45]. Transfer learning reduced reliance on sparse language-specific examples while preserving one common coding ontology [46]. Reliability and classifier validation are reported in Supplementary Tables S4 and S5.

3.4. Econometric Specification

This study uses a longitudinal product-market-wave panel rather than a respondent survey. The hypotheses concern changes in observed product-page outcomes around a predetermined regulatory milestone. H1, H2a, and H2b are evaluated from post-deadline EU–UK contrasts estimated through fixed-effects triple-difference models. H3 is evaluated from the certification differential, while H4 is tested through seller-type contrasts from the saturated extension. Product-market fixed effects absorb stable heterogeneity, while capture-wave fixed effects absorb platform-wide and date-specific shocks. Because the variables are coded page characteristics rather than latent constructs, psychometric scales and a measurement model are not applicable. Table 2 defines the coding scale for each outcome, and Supplementary Tables S4 and S5 report the corresponding reliability and validation evidence.

The tables follow the analytical sequence of the study. Tables 1 and 2 document panel construction and outcome measurement. Tables 3 and 4 establish pre-deadline comparability and describe the initial claim profile. Table 5 reports the confirmatory triple-difference estimates, Table 6 presents event-time dynamics, Table 7 evaluates robustness, and Table 8 reports exploratory heterogeneity. Each table therefore addresses a separate requirement of identification, estimation, or validation.

Table 3. Matched-sample balance before the transposition deadline.

Covariate	Certified (n = 357)	Comparison (n = 357)	Absolute SMD
Category and subtype	Exact	Exact	0.000
Premium brand segment (%)	38.7	37.3	0.029
Category-standardised log pack size, mean	0.03	−0.02	0.049
Pre-deadline median price (euro equivalent)	27.84	27.19	0.052
Star rating (1–5), mean	4.42	4.39	0.041
Review count, median (IQR)	1284 (418–4091)	1176 (397–3866)	0.074
Marketplace retailer (%)	46.8	45.4	0.028

Note. SMD = standardised mean difference. Price and review-count balance used log transforms; displayed price is in euro equivalents. Values are product-level pre-deadline summaries.

Table 4. Pre-deadline claim profile by certification and market.

Outcome	Certified EU	Certified UK	Comparison EU	Comparison UK
Validated snapshots, *n*	12,078	4050	12,123	4041
Generic claim (%)	43.8	43.1	31.6	31.0
Qualified claim (%)	37.4	36.9	19.8	20.1
Evidence-bearing claim (%)	29.6	29.1	10.9	11.2
Verified label display (%)	61.7	60.9	0.0	0.0
Claim count, mean	2.73	2.69	1.56	1.58
Prominence (0–3), mean	1.62	1.59	1.08	1.07

Note. EU pools France, Germany, and Italy. Percentages and means use validated snapshots. Comparison products had no registry-linked EU Ecolabel record at baseline. Certification strata remained fixed from the baseline registry snapshot, while verified-label display was coded separately at every capture wave.

Table 5. Fixed-effects estimates of the post-deadline EU–UK contrasts.

Outcome	Matched Comparison, τN [95% CI]	Certified Products, τC [95% CI]	Certification Differential, β3 [95% CI]
Generic claim	−8.7 [−10.6, −6.8] ***	−3.1 [−4.8, −1.4] ***	5.6 [3.4, 7.8] ***
Qualified claim	5.9 [4.1, 7.7] ***	2.8 [1.1, 4.5] **	−3.1 [−5.2, −1.0] **
Evidence-bearing claim	4.4 [3.0, 5.8] ***	2.0 [0.6, 3.4] **	−2.4 [−4.1, −0.7] **
Verified label display	0.4 [0.0, 0.8]	5.1 [3.5, 6.7] ***	4.7 [3.0, 6.4] ***
Claim count, IRR	0.91 [0.89, 0.94] ***	0.97 [0.95, 0.99] *	1.07 [1.03, 1.10] ***
Prominence	−0.18 [−0.23, −0.13] ***	−0.06 [−0.10, −0.02] **	0.12 [0.06, 0.18] ***

Note. Binary outcomes are percentage-point changes, prominence is a scale-point change and claim count is reported as an incidence-rate ratio. Standard errors are clustered by matched pair and capture wave. Secondary outcomes report Benjamini–Hochberg-adjusted q values. * p or $q < 0.05$. ** p or $q < 0.01$. *** p or $q < 0.001$.

Table 6. Pre-deadline and selected post-deadline event-time estimates for generic claims.

Event Period	Matched Comparison [95% CI]	Certified Products [95% CI]	Certification Differential [95% CI]
Week − 6	−0.1 [−1.9, 1.6]	0.1 [−1.6, 1.7]	0.2 [−2.3, 2.6]
Week − 5	0.4 [−1.3, 2.1]	−0.2 [−1.8, 1.4]	−0.6 [−2.9, 1.6]
Week − 4	−0.6 [−2.2, 1.0]	0.0 [−1.5, 1.5]	0.6 [−1.5, 2.8]
Week − 3	0.7 [−0.9, 2.3]	0.2 [−1.2, 1.6]	−0.5 [−2.5, 1.5]
Week − 2	−0.3 [−1.8, 1.2]	0.4 [−1.0, 1.8]	0.7 [−1.2, 2.5]
Week − 1, reference	**0.0**	**0.0**	**0.0**
Week 0	−2.4 [−3.8, −1.0]	−0.7 [−1.8, 0.4]	1.7 [0.0, 3.4]
Week 1	−5.8 [−7.5, −4.1]	−1.8 [−3.1, −0.5]	4.0 [2.0, 6.0]
Week 2	−8.1 [−10.0, −6.2]	−2.7 [−4.2, −1.2]	5.4 [3.1, 7.7]
Week 3	−9.4 [−11.4, −7.4]	−3.4 [−5.0, −1.8]	6.0 [3.6, 8.4]
Weeks 4–7	−10.1 [−11.8, −8.4]	−3.8 [−5.2, −2.4]	6.3 [4.2, 8.4]
Weeks 8–11	−9.6 [−11.4, −7.8]	−3.4 [−4.9, −1.9]	6.3 [4.1, 8.5]

Note. Estimates are percentage-point changes relative to the week ending 26 March 2026. Weeks 4–7 and 8–11 report linear combinations of weekly coefficients. Confidence intervals use standard errors clustered by matched pair and capture wave. Complete weekly post-deadline estimates are reported in Supplementary Table S8. The bolded Week − 1 row denotes the omitted reference period.

Table 7. Descriptive product-market transitions from the reference-week generic-only state to the late post-deadline claim architecture.

Panel A. Late Post-Deadline Claim Architecture Among Reference-Week Generic-Only Cells				
Late post-deadline claim state	**Matched Comparison EU, *n* = 255**	**Matched Comparison UK, *n* = 85**	**Certified EU, *n* = 220**	**Certified UK, *n* = 73**
Remained generic only	97 (38.0%)	53 (62.4%)	83 (37.7%)	31 (42.5%)
Mixed generic with qualified or evidence-bearing content	48 (18.8%)	19 (22.4%)	84 (38.2%)	30 (41.1%)
Qualified without generic	34 (13.3%)	4 (4.7%)	15 (6.8%)	3 (4.1%)
Evidence-bearing without generic	46 (18.0%)	5 (5.9%)	25 (11.4%)	6 (8.2%)
No coded environmental claim	30 (11.8%)	4 (4.7%)	13 (5.9%)	3 (4.1%)

Table 7. *Cont.*

Panel B. Destination states among product-market cells that lost generic claims				
Destination after generic-claim loss	**Matched Comparison EU, *n* = 110**	**Matched Comparison UK, *n* = 13**	**Certified EU, *n* = 53**	**Certified UK, *n* = 12**
Transitioned to qualified or evidence-bearing communication	80 (72.7%)	9 (69.2%)	40 (75.5%)	9 (75.0%)
Transitioned to no coded environmental claim	30 (27.3%)	4 (30.8%)	13 (24.5%)	3 (25.0%)

Note. Values are *n* (column %). EU pools France, Germany, and Italy. Panel A percentages use reference-week generic-only cells as the denominator. Panel B percentages are conditional on generic-claim loss. Eligibility criteria and state-assignment rules are reported in Section 3.4.

Table 8. Robustness estimates for the generic-claim outcome.

Specification	Observations	Matched Comparison, τN [95% CI]	Certified Products, τC [95% CI]	Certification Differential, β3 [95% CI]
Main specification	96,784	−8.7 [−10.6, −6.8]	−3.1 [−4.8, −1.4]	5.6 [3.4, 7.8]
Balanced panel	60,156	−8.5 [−10.7, −6.3]	−3.0 [−4.9, −1.1]	5.5 [3.0, 8.0]
Inverse-probability weighting	96,784	−8.9 [−10.9, −6.9]	−3.2 [−5.0, −1.4]	5.7 [3.4, 8.0]
Transition captures excluded	86,027	−9.0 [−11.1, −6.9]	−3.3 [−5.1, −1.5]	5.7 [3.3, 8.1]
Market-specific linear trends	96,784	−8.1 [−10.4, −5.8]	−2.8 [−4.8, −0.8]	5.3 [2.7, 7.9]
Any-page qualification rule	96,784	−7.8 [−9.6, −6.0]	−2.6 [−4.2, −1.0]	5.2 [3.1, 7.3]
Misclassification-corrected estimates	96,784	−9.1 [−11.3, −6.9]	−3.4 [−5.4, −1.4]	5.7 [3.1, 8.3]
Market-by-wave fixed effects	96,784	NI	NI	5.4 [3.0, 7.8]

Note. NI = not identified because market-by-wave fixed effects absorb market-level contrasts; β3 remains identified within markets. Confidence intervals use two-way clustered standard errors, except for the misclassification correction.

Let i index products, c markets and t capture waves. T_c identifies France, Germany and Italy, C_i identifies baseline EU Ecolabel certification, and P_t identifies observations from the first post-deadline capture on 30 March 2026. The confirmatory triple-difference model followed the formulation developed by Olden and Møen [47]:

$$Y_{ict} = \alpha_{ic} + \lambda_t + \beta_1 T_c P_t + \beta_2 C_i P_t + \beta_3 T_c C_i P_t + \varepsilon_{ict}$$

Product-market fixed effects absorb stable product, seller, translation and market differences. Capture-wave fixed effects absorb common platform changes, seasonality and date shocks. β1 identifies the post-deadline EU–UK contrast among matched comparison products. β2 captures the additional post-deadline change associated with certification in the UK comparator. β3 identifies how that certification difference changes in EU markets. The two post-deadline EU–UK contrasts were therefore:

$$\tau_N = \beta_1; \tau_C = \beta_1 + \beta_3$$

The generic-claim indicator was the primary dependent variable. Binary outcomes were estimated with a fixed-effects linear probability model and reported as percentage-point changes. Claim count was estimated using Poisson pseudo-maximum likelihood with the same fixed effects. Prominence was analysed on its 0–3 scale; a binary first-viewport measure served as a sensitivity outcome. The confirmatory model omitted price, rating, review count, seller type and availability controls because these variables could change within the post-deadline window and therefore operate as post-cutoff covariates. Covariate-adjusted models were reported only as sensitivity checks. Conventional inference used two-way CRV1 standard errors clustered by matched pair and capture wave, accommodating serial dependence within pairs and shocks shared on a collection date [48]. The finite-

cluster correction was applied, with the reference degrees of freedom determined by the smaller clustering dimension ($G = 36$, $df = 35$). Small-cluster sensitivity was assessed using 9999 null-imposed wild restricted cluster-bootstrap replications with Webb weights. Bootstrap weights were assigned at the capture-wave level while the variance estimator retained two-way clustering. Confirmatory, seller-type, and exploratory subgroup-equality tests were examined (Table S10). The primary outcome used a two-sided 0.05 threshold and 95% confidence intervals. Benjamini–Hochberg correction at q = 0.05 controlled the secondary outcome family.

H4 was tested through a saturated seller-type extension of the primary model. Let R_{ic} equal one for baseline marketplace-retail cells. The model added $R_{ic}\ x\ P_t$, $R_{ic}\ x\ T_c\ x\ P_t$, $R_{ic}\ x\ C_i\ x\ P_t$, and $R_{ic}\ x\ T_c\ x\ C_i\ x\ P_t$ interactions. Linear combinations recovered the post-deadline EU-UK contrasts for marketplace-retail and third-party cells within each certification stratum. Wald tests evaluated the corresponding seller-type contrasts using the same two-way clustered standard errors as the primary specification.

EU Ecolabel status and baseline seller type are pre-cutoff grouping variables rather than randomly assigned treatments. Accordingly, β3 measures differential change across certification strata within the matched cohort, while the seller-type interactions measure heterogeneity by the offer's baseline organisational setting. These estimands do not identify a causal effect of EU Ecolabel certification or attribute product-page revisions to the active seller.

A weekly event study replaced the post indicator and its interactions with event-week indicators, using the week ending 26 March 2026 as the reference. The model included EU-market-by-week, certification-by-week and EU-market-by-certification-by-week terms, while retaining product-market and capture-wave fixed effects. Pre-deadline coefficients assessed differential trends; post-deadline coefficients traced the timing and persistence of claim revision. Because national transposition dates differed, the coefficients are indexed to the common EU deadline and do not estimate an average staggered-adoption effect. H1 requires that, absent deadline-associated exposure, matched comparison products would have changed similarly in EU and UK marketplaces. H3 requires the counterfactual EU–UK change to evolve similarly across EU Ecolabel and comparison strata. These assumptions concern changes, not levels, and require stable product identity and no contemporaneous EU-specific cutoff shock [49].

To distinguish aggregate changes from within-page transitions, we conducted a descriptive analysis at the product-market level. The reference state used the last valid snapshot from the week ending 26 March 2026. The late post-deadline state was defined as the modal claim architecture during weeks 8–11, with ties resolved by the latest valid observation. Eligibility required a valid reference-week snapshot and at least three valid late-period observations. Claim indicators were recoded into five mutually exclusive states: no coded environmental claim, generic only, mixed generic and qualified or evidence-bearing, qualified without generic, and evidence-bearing without generic. Hierarchical assignment classified a page as mixed whenever generic language appeared with qualified or evidence-bearing content. Evidence-bearing status took precedence over qualified status when generic language was absent. The analysis was descriptive and was not used for confirmatory hypothesis testing.

3.5. Validation, Robustness, and Reproducibility

End-to-end validity was assessed on 1280 complete snapshots stratified by market, period, certification and predicted class. Coders reviewed rendered pages without certification or period metadata. Audit accuracy was 92.7%, with no market-period stratum below 89.4%. Errors centred on qualifications appearing only in lower-page modules. Estimates

were therefore repeated using same-module, first-viewport and any-page qualification rules. A likelihood-based correction used locked test-set confusion matrices to assess classifier-induced bias [50].

Design checks covered trends, timing and sample composition. Pre-deadline event coefficients were assessed through joint Wald tests and ±2.5 percentage-point equivalence bounds. Honest difference-in-differences intervals allowed trend violations ranging from 0.5 to 2.0 times the largest pre-deadline deviation [51]. Placebo cutoffs used pre-deadline data from 2, 9 and 16 March. Further specifications excluded captures from 23 March to 2 April, added market-specific linear trends, removed each EU-exposed market in turn and restricted the sample to 1671 balanced product-market cells, yielding 60,156 records. Inverse-probability weights modelled text availability from baseline covariates and lagged page status and were trimmed at the first and ninety-ninth percentiles. Additional tests used saturated market-by-wave fixed effects, within-pair label permutations and cells with stable seller classification in at least 10 of 12 pre-deadline waves.

The protocol, outcome hierarchy, matching rules, codebook, thresholds and equations were time-stamped before collection. Classification used PyTorch 2.7.0, Transformers 5.14.1 and scikit-learn 1.9.0; fixed-effects estimation used pyfixest 0.60.0. Separate locked environments, deterministic seeds, versioned configurations, source commits, and SHA-256 manifests preserved computational identity [52]. A version-controlled replication archive contains code, configurations, the codebook, annotation identifiers, derived measures, manifests, and a synthetic fixture; these materials are available upon reasonable request. Platform terms prohibit redistribution of raw marketplace files and annotated text; hashes, provenance records, schemas, and reconstruction scripts are retained for verification. No remote generative model or mutable API was used for classification. The replication-package manifest is provided in Supplementary S2.

4. Results

This section presents the empirical analysis in six stages. It first reports the analytical sample and pre-deadline balance, followed by the main fixed-effects estimates and event-time dynamics. It then evaluates the stability of the estimates through robustness and sensitivity analyses before examining heterogeneity across markets, product categories, and baseline seller types.

4.1. Sample Profile and Pre-Deadline Balance

The analytical cohort contained 357 EU Ecolabel-certified products and 357 matched comparison products. Category composition was identical by design: 149 cleaning-product pairs (41.7%), 113 personal-care pairs (31.7%), and 95 textile pairs (26.6%). Observation across four marketplaces yielded 2856 product-market cells. The 12-wave pre-deadline window contributed 32,292 validated snapshots from 34,272 scheduled captures, a retention rate of 94.2%. Retention was 94.2% in the EU markets and 94.4% in the UK; no certification-by-market cell differed by more than 0.6 percentage points. Baseline classification identified 1316 marketplace-retail product-market cells (46.1%) and 1540 third-party cells (53.9%). These groups contributed 44,619 and 52,165 validated snapshots, respectively. Marketplace-retail shares were 46.8% among certified products and 45.4% among comparison products, consistent with the matched-sample balance reported in Table 3.

Balance improved materially after matching. The median absolute SMD was 0.041 and the maximum was 0.074, below the prespecified 0.10 threshold. Exact category and subtype matching preserved the intended product-domain composition; no pair was removed after outcome extraction. Table 4 reports the corresponding pre-deadline claim profile.

Certified products carried more environmental communication before the transposition deadline. In EU markets, generic claims were 12.2 percentage points more prevalent among certified products; the corresponding gaps were 17.6 points for qualified claims and 18.7 points for evidence-bearing claims. These level differences were substantively expected and do not violate the design. Product-market fixed effects absorb baseline claim intensity, while identification rests on counterfactual changes after the transposition deadline.

Pre-deadline EU–UK gaps were small within each certification stratum: no binary outcome exceeded 0.8 percentage points, claim count 0.04 per page, or prominence 0.03 units. Treatment-by-week leads assess the H1 counterfactual among comparison products; treatment-by-certification-by-week leads assess the additional H3 differential. For generic claims, joint tests did not reject these lead sets ($p = 0.61$ and $p = 0.47$); secondary-outcome p values ranged from 0.38 to 0.74. The widest 90% confidence interval for a generic-claim lead, [−1.9, 2.2] percentage points, lay within the prespecified ±2.5-point equivalence region. These diagnostics support, but cannot establish, the assumptions.

The inferential value of the six-week pre-period lies in its high-frequency observation of the same product-market cells immediately before the predetermined cutoff. Twelve capture waves yielded 32,292 validated snapshots, with balanced retention and minimal baseline EU–UK differences. This local evidence is suited to detecting differential movements near the cutoff, but it cannot establish longer-term parallelism. Placebo dates, market-specific trends, and relative-magnitude sensitivity bounds therefore complement the lead-coefficient tests. Figure 2 and Table 6 report the complete set of pre-period weekly estimates.

4.2. Main Post-Deadline Estimates

The confirmatory model identified a marked decline in generic environmental claims after the transposition deadline. Among matched comparison products, EU pages showed an 8.7-percentage-point reduction relative to the concurrent UK change (95% CI [−10.6, −6.8], $p < 0.001$). The corresponding reduction among EU Ecolabel-certified products was 3.1 points (95% CI [−4.8, −1.4], $p < 0.001$). The 5.6-point certification differential was significant (95% CI [3.4, 7.8], $p < 0.001$), indicating a smaller post-deadline decline in the certified stratum of the matched cohort. The negative comparison-product contrast supports H1, while the positive certification differential supports H3. Table 5 reports the full outcome estimates; Supplementary Table S7 provides the coefficient decomposition.

Secondary outcomes indicated a shift in claim composition. Qualified claims increased by 5.9 percentage points among matched comparison products and 2.8 points among certified products. Evidence-bearing claims increased by 4.4 and 2.0 points, respectively. Claim volume declined by 9% among comparison products and 3% among certified products, while prominence fell by 0.18 and 0.06 scale points. Verified EU Ecolabel display increased by 5.1 points among certified products; the 0.4-point change among comparison products was not statistically significant after multiplicity adjustment. The positive EU–UK contrasts for qualified and evidence-bearing claims support H2a and H2b.

The outcome pattern is consistent with a reconfiguration of claim composition rather than a uniform contraction. Comparison products showed the largest decline in generic claims and the largest gains in qualified and evidence-bearing claims. Certified products showed a smaller decline in generic claims and greater visibility of registry-verified labels.

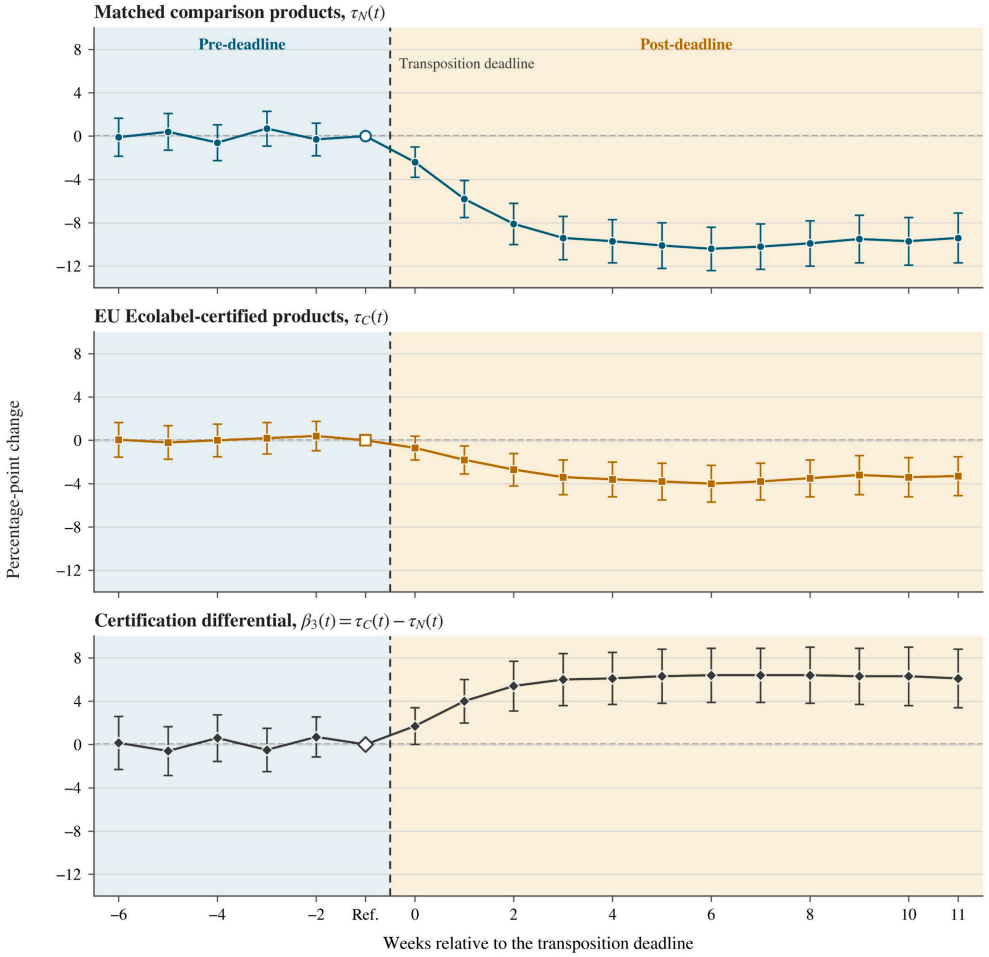


Figure 2. Event-Time Estimates for Generic Environmental Claims. Weekly estimates are relative to the week ending 26 March 2026; whiskers show 95% confidence intervals. The panels report τN(t), τC(t) and β3(t) = τC(t) − τN(t). The dashed line marks the 27 March 2026 transposition deadline. Standard errors are clustered by matched pair and capture wave. Circles, squares, and diamonds denote τN(t), τC(t), and β3(t), respectively; open markers denote the omitted reference week (Week − 1), set to zero.

4.3. Event-Time Dynamics and Adjustment Path

Event-time estimates showed that claim adjustment began immediately but was not completed at the first post-deadline capture. Among matched comparison products, the generic-claim effect was −2.4 percentage points in week 0 and −5.8 points in week 1. It reached −9.4 points by week 3 and remained close to −10 points during weeks 4–7. The response among certified products was smaller. The week 0 estimate was −0.7 points and statistically insignificant, followed by reductions of 1.8 points in week 1 and 3.4 points in week 3. As shown in Figure 2, the certification differential widened during the initial adjustment period and remained positive after the weekly effects stabilised.

The pattern indicates a persistent difference in adjustment intensity rather than a response confined to the first post-deadline captures. Table 6 reports selected coefficients, and Supplementary Table S8 provides the complete weekly estimates.

The generic-claim effects stabilised after the third post-deadline week. Equality tests across weeks 4–11 were not rejected for matched comparison products ($p = 0.38$) or certified products ($p = 0.44$). The absence of late-period reversal indicates that the average effects reported in Section 4.2 were not driven by a short-lived response around the transposition deadline. Figure 2 clarifies that adjustment was staged rather than instantaneous. Generic language contracted before qualified and evidence-bearing content reached their later-

period levels. The persistence of the post-deadline coefficients after week 3 makes a short-lived marketplace disturbance an unlikely explanation and is more consistent with sequential product-page revision.

Changes in claim substantiation developed more gradually. Among comparison products, qualified claims increased by 1.8 percentage points in week 0, 4.9 points by week 2, and an average of 7.3 points during weeks 8–11. Evidence-bearing claims increased by 1.1, 3.6, and 5.6 points across the same intervals. Verified-label display among certified products rose by 1.3 points in week 0 and 3.9 points by week 2, reaching an average increase of 6.5 points during the final four weeks. Figure 3 presents these trajectories.

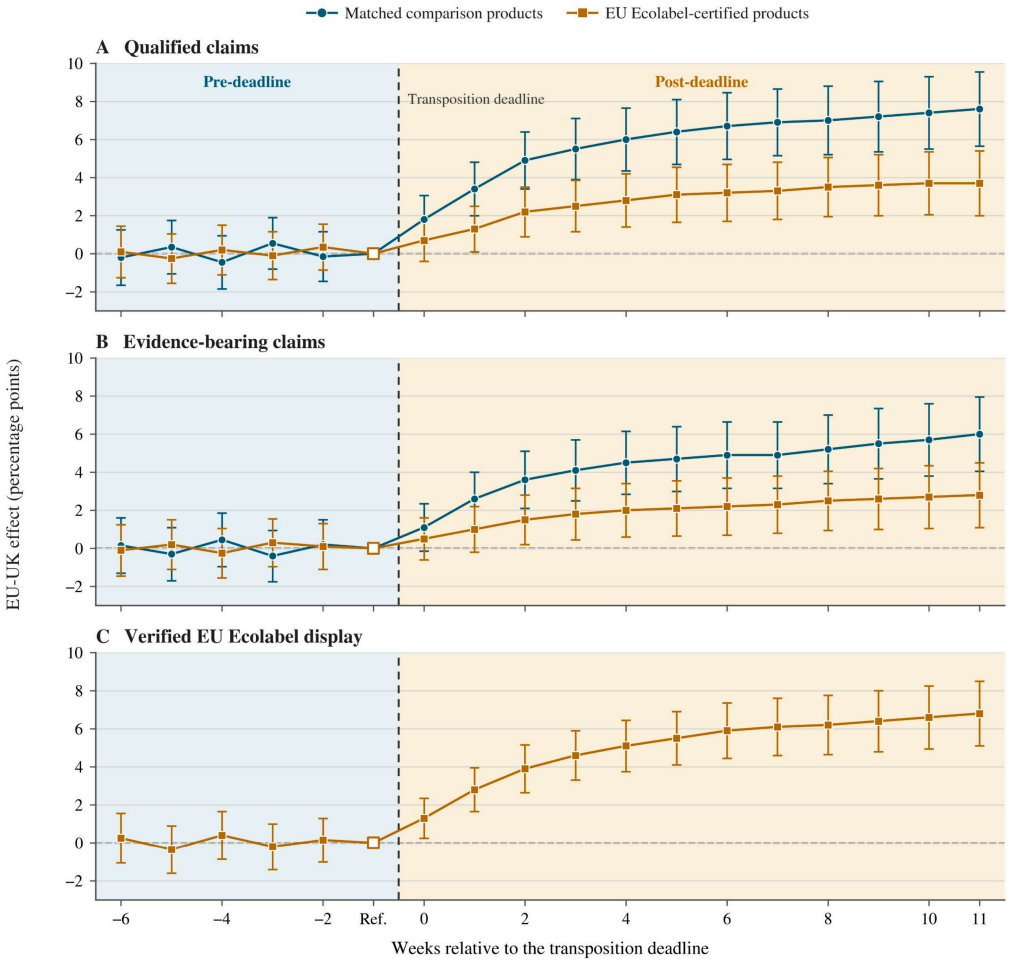


Figure 3. Event-Time Reconfiguration of Environmental Claim Architecture. (**A**) Qualified claims; (**B**) evidence-bearing claims; (**C**) verified EU Ecolabel display. Weekly EU–UK estimates are relative to the week ending 26 March 2026; whiskers show 95% confidence intervals. Panel (**C**) reports certified products only; the comparison series was omitted because verified-label display remained negligible and showed no significant post-deadline change. The dashed line marks the 27 March 2026 transposition deadline. Standard errors are clustered by matched pair and capture wave. Circles and squares denote estimates for matched comparison products and EU Ecolabel-certified products, respectively; open markers denote the omitted reference week (Week − 1), set to zero.

The event-time results reveal distinct adjustment speeds. Generic-claim effects stabilised by week 3, whereas qualified, evidence-bearing and label-display effects continued to increase. The event-time estimates therefore indicate faster withdrawal than substantiation, although they do not trace direct within-page replacement.

4.4. Descriptive Product-Market Transition Analysis

To examine the page-level transitions underlying the aggregate outcome pattern, we conducted a descriptive secondary analysis at the product-market level. The analysis was designed to clarify the interpretation of claim reconfiguration and was not used for confirmatory hypothesis testing. Table 7 reports the resulting transition matrix.

Among matched-comparison EU cells classified as generic only during the reference week, 43.1% no longer displayed generic language in the late post-deadline period. Of the cells that lost generic claims, 72.7% subsequently displayed qualified or evidence-bearing content, while 27.3% moved to a no-claim state. The corresponding generic-loss rate was 15.3% among UK comparison cells. Among certified EU cells, 24.1% lost generic claims, and 75.5% of these cells subsequently displayed qualified or evidence-bearing content. The corresponding rates among certified UK cells were 16.4% and 75.0%. The descriptive transition matrix indicates that aggregate reconfiguration comprised signal upgrading and communication withdrawal, although the pathways were not uniform across product pages.

4.5. Robustness and Sensitivity Analyses

The generic-claim estimates remained substantively stable across the prespecified robustness analyses. Restricting the analysis to the 1671 product-market cells observed in every wave produced effects close to the main estimates. Inverse-probability weighting yielded slightly larger reductions, indicating that differential page availability did not explain the findings. Table 8 reports the main robustness estimates; extended diagnostics appear in Supplementary Table S9.

Removing the four captures surrounding the transposition deadline increased the estimated reduction among comparison products from 8.7 to 9.0 percentage points. The result therefore did not depend on pages collected during the immediate transition period. Adding market-specific linear trends reduced the coefficients modestly but preserved their signs and statistical significance. The first-viewport qualification rule produced estimates of −8.3 points for comparison products, −2.9 points for certified products, and 5.4 points for the certification differential. Under the more permissive any-page rule, the corresponding estimates were −7.8, −2.6, and 5.2 points.

Likelihood-based correction for classifier errors increased the estimated generic-claim reductions by 0.3 to 0.4 percentage points. The limited change is consistent with the high locked-test and complete-page validation performance. The market-by-wave specification produced a certification differential of 5.4 points, indicating that the estimated difference persisted after absorbing market-wide shocks coinciding with the transposition deadline.

None of the three placebo dates generated a generic-claim discontinuity larger than 1.1 percentage points in absolute value. All placebo confidence intervals included zero, with p values ranging from 0.29 to 0.81. Single-market exclusions produced comparison-product effects between −9.1 and −8.2 points and certification differentials between 5.2 and 6.0 points. In 9999 within-pair label permutations, only five simulated differentials were at least as large as the observed estimate, yielding a randomisation p value of 0.0006.

Sensitivity to non-parallel post-deadline trends was assessed using the prespecified relative-magnitude bounds. The confidence interval for the comparison-product effect remained below zero when the permitted trend violation reached twice the largest pre-deadline deviation (95% CI [−15.4, −0.9]). The certification differential remained positive through a 1.5-times violation bound (95% CI [0.6, 10.5]) but included zero at the two-times bound (95% CI [−1.0, 12.0]). Evidence for generic-claim withdrawal is therefore robust to sizeable departures from parallel trends; the exact magnitude of the certification differential requires a stronger counterfactual-trend assumption.

Secondary results were similarly stable, although the smaller certified-product effects were less precise in demanding specifications. In the market-trend model, the certified-product claim-count estimate weakened to an incidence-rate ratio of 0.98 (95% CI [0.96, 1.01], $q = 0.083$). The increases in qualified claims, evidence-bearing claims, and verified-label display retained their expected directions and remained significant after multiplicity adjustment.

4.6. Heterogeneity Across Markets and Product Categories

Exploratory interaction models assessed whether the pooled estimates concealed divergent responses across EU-exposed markets or product categories. Generic claims declined in France, Germany, and Italy. Among matched comparison products, the estimated reductions ranged from 7.8 percentage points in Germany to 9.4 points in France. Certified-product reductions ranged from 2.7 to 3.5 points. The certification differential remained positive in every EU-exposed market. Table 9 reports the subgroup estimates.

Table 9. Heterogeneity in post-deadline generic-claim contrasts.

Subgroup	Matched Comparison, τN [95% CI]	Certified Products, τC [95% CI]	Certification Differential, β3 [95% CI]
Market			
France	−9.4 [−12.0, −6.8]	−3.5 [−5.7, −1.3]	5.9 [3.0, 8.8]
Germany	−7.8 [−10.2, −5.4]	−2.7 [−4.8, −0.6]	5.1 [2.4, 7.8]
Italy	−9.1 [−11.7, −6.5]	−3.2 [−5.4, −1.0]	5.9 [2.9, 8.9]
Product category			
Cleaning products	−10.1 [−12.5, −7.7]	−3.6 [−5.7, −1.5]	6.5 [3.6, 9.4]
Personal-care products	−8.6 [−11.2, −6.0]	−3.0 [−5.3, −0.7]	5.6 [2.4, 8.8]
Textiles	−7.2 [−10.2, −4.2]	−2.5 [−5.1, 0.1]	4.7 [1.0, 8.4]

Note. Estimates are percentage-point changes. Market effects are relative to the United Kingdom; category effects pool the EU marketplaces. Standard errors are clustered by matched pair and capture wave.

Joint Wald tests did not reject equality of the comparison-product effects across France, Germany, and Italy ($p = 0.29$). Cross-market heterogeneity was similarly unsupported for certified products ($p = 0.73$) and the certification differential ($p = 0.88$). The pooled estimate was therefore not driven by a single national marketplace. Because national transposition timelines differed, this cross-market consistency argues against a result driven by one country's enactment, but it does not make 27 March a common national legal intervention.

Category estimates showed greater numerical variation. The generic-claim reduction among comparison products was largest for cleaning products at 10.1 percentage points and smallest for textiles at 7.2 percentage points. The category interaction was nominally significant ($p = 0.047$) but did not remain significant after correction for the heterogeneity-test family ($q = 0.094$). Differences among certified products were smaller ($p = 0.68$). The positive certification differential ranged from 4.7 to 6.5 points, with no statistically reliable category interaction ($p = 0.41$). Secondary outcomes followed a similar pattern. Among comparison products, qualified claims increased by 7.0 percentage points for cleaning products, 5.7 points for personal-care products, and 4.4 points for textiles. Evidence-bearing claims increased by 5.2, 4.3, and 3.4 points, respectively. These category interactions did not survive multiplicity adjustment. The increase in verified-label display among certified products ranged from 4.5 points for textiles to 5.8 points for cleaning products, with $q = 0.46$ for the interaction test.

4.7. Heterogeneity by Baseline Seller Type

Figure 4 reports estimates by EU marketplace, product category and baseline seller type. Market estimates use the United Kingdom as the comparator, while category estimates

pool France, Germany and Italy. Seller type was defined by the modal winning-offer seller during the 12 pre-deadline waves. All subgroup models retained product-market and capture-wave fixed effects with two-way clustered standard errors.

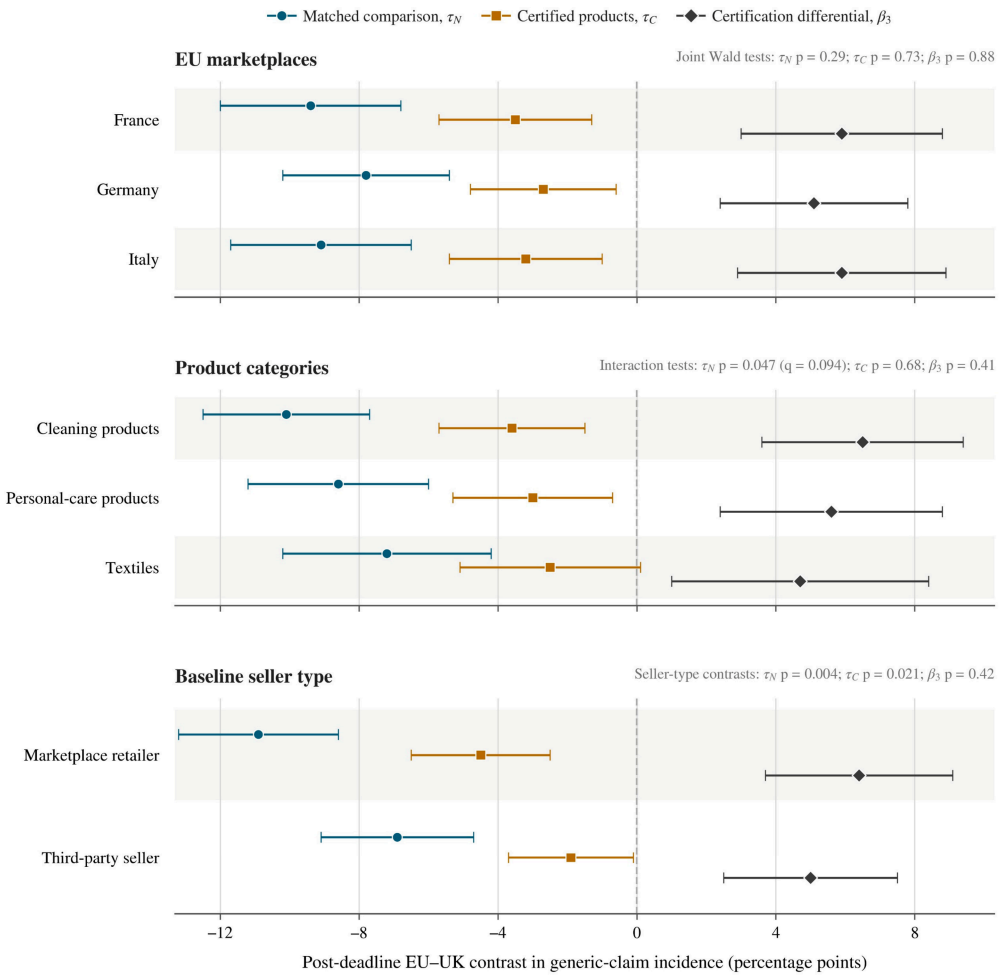


Figure 4. Heterogeneity in Post-Deadline Generic-Claim Contrasts. Points report subgroup-specific percentage-point estimates, and whiskers indicate 95% confidence intervals. Blue circles denote the EU–UK contrast for matched comparison products (τN), amber squares denote the corresponding contrast for certified products (τC), and grey diamonds denote the certification differential (β3). Market estimates use the United Kingdom as the comparator; category and seller-type estimates pool the EU-exposed marketplaces. The vertical dashed line marks zero. Displayed *p* values report the relevant joint Wald, interaction, or seller-type contrast tests.

Generic claims declined in every subgroup. All comparison-product effects and certification differentials excluded zero; among certified products, only the textile estimate crossed zero. Interaction tests identified seller type as the only reliable source of variation in effect magnitude.

Among comparison products, the post-deadline contrast was −10.9 percentage points for marketplace-retail offers (95% CI [−13.2, −8.6]) and −6.9 points for third-party offers (95% CI [−9.1, −4.7]). The −4.0-point difference was significant (95% CI [−6.7, −1.3], $p = 0.004$). Among certified products, the corresponding estimates were −4.5 points (95% CI [−6.5, −2.5]) and −1.9 points (95% CI [−3.7, −0.1]). Their −2.6-point difference was also significant (95% CI [−4.8, −0.4], $p = 0.021$). The average seller-type contrast across certification strata was −3.2 points (95% CI [−4.9, −1.5], $p < 0.001$), supporting H4.

The certification differential remained positive for marketplace-retail offers at 6.4 points (95% CI [3.7, 9.1]) and third-party offers at 5.0 points (95% CI [2.5, 7.5]). These

estimates did not differ significantly (difference = 1.4 points, 95% CI [−2.0, 4.8], $p = 0.42$). Claim-withdrawal intensity therefore differed by baseline seller type, while the direction of the certification differential remained consistent. Restricting the analysis to cells with stable pre-deadline seller classification produced a similar contrast of −3.0 points (95% CI [−4.8, −1.2], $p = 0.001$).

Wild-cluster-bootstrap inference preserved the conclusions for H1, H3, and the seller-type contrasts (Table S10). The seller-type difference in the EU Ecolabel differential remained non-significant (WCB $p = 0.447$). Among the exploratory tests, category equality for matched-comparison effects weakened to WCB $p = 0.072$; all other equality conclusions were unchanged.

5. Discussion

The findings reveal a structured shift in environmental-claim architecture around the 27 March 2026 transposition deadline. Generic claims contracted most sharply among matched comparison products, whereas qualified claims, evidence-bearing claims, and verified-label display became more prominent. The smaller decline among EU Ecolabel-certified products is consistent with greater access to existing evidentiary resources. The seller-type contrast identifies variation associated with the offer's baseline organisational setting. These patterns describe anticipatory adjustment around a common EU transposition deadline, not the effect of simultaneous national enactment or enforcement.

5.1. From Generic Claim Withdrawal to Signal Upgrading

The 8.7-point decline in generic claims exceeded the 5.9-point increase in qualified claims and the 4.4-point increase in evidence-bearing claims. Because these binary outcomes can co-occur, the coefficients do not represent a one-for-one replacement rate. Claim prevalence alone is insufficient: total claim volume fell as qualified and evidence-bearing content became more prevalent. Greenwashing assessment frameworks likewise emphasise claim scope, accessibility, verifiability and supporting evidence rather than the number of environmental statements [53].

This contraction is consequential because vague claims can influence product evaluations without being readily recognised as misleading. Experimental evidence shows that consumers identify demonstrably false environmental statements more easily than vague language, which can continue to shape brand attitudes through less deliberative routes [12,54]. Regulation aimed at generic expressions therefore addresses a form of communication that may be persuasive precisely because its meaning is difficult to test. The observed decline is consistent with anticipatory revision around the common transposition deadline, although the design cannot isolate the underlying mechanism.

The increases in qualified and evidence-bearing claims support H2a and H2b. Among matched comparison products, qualified claims increased by 5.9 percentage points and evidence-bearing claims by 4.4 points. The corresponding increases among EU Ecolabel-certified products were 2.8 and 2.0 points. This difference should not be interpreted as weaker adaptation among certified products. Their pre-deadline levels of qualified and evidence-bearing communication were already substantially higher. Comparison products faced a larger evidentiary deficit and had more generic language to replace. Their larger change is consistent with greater scope for revision among claims furthest from the new evidentiary standard.

Event-time estimates add a temporal distinction. Generic-claim effects stabilised after the third post-deadline week, while qualified and evidence-bearing effects continued to grow. This timing is consistent with faster removal of exposed language than development of narrower wording and supporting material. It does not establish that the same pages

underwent those transitions. Internal coordination, unavailable indicators, and uncertainty over defensible claims are documented barriers to substantiated digital sustainability communication [55].

The descriptive transition analysis clarifies the page-level composition of this pattern. Most comparison EU cells that lost generic wording subsequently displayed qualified or evidence-bearing content, yet 27.3% moved to a no-claim state. The certified EU stratum showed a lower generic-loss rate, although the proportion moving towards qualified or evidence-bearing communication after withdrawal was similar. The descriptive pattern comprised distinct pathways of signal upgrading and communication withdrawal rather than uniform substitution within every product page.

The decline should not automatically be described as greenhushing. That concept requires evidence that genuine environmental performance is being under-communicated. The present study observes communication but does not measure the underlying environmental performance of each product. Recent research defines environmental silence through the gap between performance and disclosure, a distinction that cannot be established from claim text alone [56]. "Claim withdrawal" is therefore the more defensible interpretation. Greenhushing remains one possible explanation for products whose defensible attributes were no longer communicated, but it is not directly identified here.

5.2. Certification Status and Baseline Seller Setting as Sources of Heterogeneity

The certification differential supports H3. Generic claims declined by 3.1 percentage points among EU Ecolabel-certified products, compared with 8.7 points among matched comparison products. Verified-label display also increased by 5.1 points among certified products. The certified stratum therefore exhibited less withdrawal and greater visibility of a registry-verified signal after the deadline.

This result refines the role assigned to certification. The EU Ecolabel did not protect every environmental expression attached to a certified product. Certified pages still removed generic language, demonstrating that certification cannot serve as unrestricted support for claims extending beyond its criteria or scope. This pattern is consistent with greater pre-existing evidentiary capacity in the EU Ecolabel stratum. These products entered the period with a recognised licence, established criteria and an existing technical dossier.

The increase in label display is particularly informative. Earlier research found that many certified organisations communicate their certification weakly or not at all, despite having acquired a potentially credible signal [57]. The post-deadline increase in verified-label display is consistent with greater reliance on this previously underused signalling resource as the regulatory setting changed. The result concerns increased visibility of an existing EU Ecolabel signal, rather than a causal effect of certification.

Baseline seller setting marked a second source of heterogeneity. Generic-claim reductions were larger for marketplace-retail offers than for third-party offers, at 10.9 and 6.9 percentage points respectively, supporting H4. Hybrid platforms occupy a central position from which they can set rules, coordinate processes and influence ecosystem participants [58]. Marketplace-retail operations can apply common review procedures across product portfolios. Third-party sellers differ in legal knowledge, technical capacity and access to substantiating evidence. Their average response is therefore likely to be less synchronised.

This finding does not establish that the active seller authored the displayed claim. Amazon product pages may be shared across several offers and content contributors. Seller type identifies the governance setting surrounding the offer, not textual authorship. The result remains theoretically informative because deadline-associated adjustment differed across the organisational settings through which marketplace content was maintained.

Market and category analyses locate the boundary of heterogeneity. Effect direction was stable, but magnitude did not differ reliably after multiplicity correction. Response intensity differed by baseline seller type. This pattern is consistent with a more prominent role for coordination capacity than for national marketplace or product-category variation.

5.3. Theoretical Contributions

The study shifts signalling analysis from audience responses to sender revision. Generic, qualified and evidence-bearing claims form a signal set whose composition changes when institutional costs change. A disclosure-versus-silence binary would miss the decline in broad claims, rise in substantiated content and greater label visibility. Claim architecture therefore makes regulatory signal revision observable.

The certification pattern is consistent with accumulated evidentiary capacity. Certified products still revised generic wording, but an existing licence, defined criteria and supporting documentation may have expanded the set of defensible responses. The estimated differential indicates how pre-existing certification resources can coincide with adjustment under greater scrutiny without identifying a causal certification effect.

The baseline seller setting identifies an organisational boundary condition for signalling theory. The stronger marketplace-retail contrast is consistent with signal revision depending on coordination capacity and evidentiary cost. External consumer-protection rules are implemented through platform content systems, seller arrangements and review procedures.

These contributions remain confined to communication. Mandatory environmental disclosure can alter operational performance when measurement and benchmarking create pressure for real change [59]. The present estimates do not demonstrate such an effect. They reveal how environmental information was reorganised at the digital point of sale.

5.4. Policy Implications

The policy implications concern monitoring around the transposition process; the study does not estimate enforcement effects. Regulators should assess claim quality alongside prevalence. Lower claim volume may reflect removal of weak signals or reduced visibility of defensible environmental information. Monitoring should distinguish generic withdrawal, qualification, evidentiary support and verified-label display. Claim counts alone cannot establish improved communication.

Prominence also requires attention. Online consumers often evaluate green claims using limited cues because supporting resources are absent or difficult to locate [60]. Substantiation placed far below the initial claim may do little to correct the impression created in the first viewport. Enforcement guidance should specify where qualification and evidence must appear, rather than focusing solely on whether supporting information exists somewhere on the page.

5.5. Recommendations for Platforms and Firms

Digital marketing governance increasingly depends on the alignment of technological systems, legal innovation, and sustainability objectives [61]. Marketplaces could facilitate review through structured environmental-claim fields, evidence links and automated licence verification. Generic terms could trigger a request for the relevant attribute, claim boundary and source. Registry connections could validate public labels before display. Automated screening should support human review because claim meaning depends on scope, product context and visual presentation.

The seller-type findings indicate a need for targeted support. Third-party sellers may require multilingual templates, category-specific examples and clear procedures for attaching evidence. Evidence from European SMEs indicates that information gaps and

access to support influence the adoption of resource-efficiency practices and sustainable innovation [62]. A blanket removal policy would be easier to administer but could encourage unnecessary silence. Communication research shows that greenwashing and environmental under-communication create distinct trust problems [63]. Marketplace governance should therefore enable defensible communication rather than reward simple deletion.

For firms, environmental claims should be managed as auditable information assets. Each claim needs an identified scope, evidence owner, verification date and page location. Certification should be connected to the precise attribute it supports. This approach may reduce legal risk and preserve useful sustainability information.

6. Limitations and Future Research Directions

The estimates identify differential adjustment around the common 27 March 2026 EU deadline, not the effect of an identical national intervention in France, Germany and Italy. National transposition followed different legislative paths, and France had not completed transposition during the observation window. Member States were required to apply the transposing provisions from 27 September 2026, after data collection ended. The results should therefore be interpreted as anticipatory marketplace adjustment rather than an enactment or enforcement effect. Extending the panel beyond formal application would show whether the observed revisions persisted, intensified or reversed once the legal requirements became operative.

Any causal interpretation would require parallel counterfactual changes across the EU and UK marketplaces and across certification strata. The comparison stratum may include products carrying alternative environmental labels; β3 therefore identifies heterogeneity by baseline EU Ecolabel status, not certification generally. The United Kingdom provided an informative comparator because it was not subject to the EU transposition deadline and no comparable new rule entered application during the study period, although it remained subject to the Green Claims Code and related guidance [40]. Sensitivity analyses preserved the negative comparison-product estimate under sizeable deviations from parallel trends. The certification differential was more sensitive to the most demanding bound. Its positive sign persisted across specifications, but its magnitude depends more heavily on the assumed counterfactual trend. H1 requires parallel counterfactual EU–UK changes among comparison products; H3 additionally requires parallel EU–UK change differentials across EU Ecolabel and comparison strata. The six-week pre-period offers local evidence, not proof of longer-term parallelism. The UK remained subject to the Green Claims Code, but no comparable new rule entered application [40]. The comparison-product estimate remained negative under sizeable trend violations. The certification differential retained its sign but was more sensitive to the most demanding bound. These diagnostics cannot separate the EU deadline from every contemporaneous EU-specific development. Capture-wave fixed effects account for Amazon-wide changes common to the four sampled marketplaces. EU-specific platform initiatives or contemporaneous EU-level developments remain part of the estimated deadline-associated contrast. Distinguishing a Directive-specific response would require access to platform policy records or seller revision histories, additional non-EU comparison markets, or observation around national enactment and the 27 September 2026 application date.

The outcomes capture visible environmental communication, not environmental truth, legal compliance or deceptive intent. A generic claim is not necessarily false, and an evidence-bearing claim may still provide an incomplete account of environmental impact. Although locked-test performance was strong and correction for classification error produced little change in the estimates, automated coding may miss qualifications embedded in complex visual, dynamic or linked material. The analysis also cannot establish whether

underlying product performance changed. Future research could link product-page content to life-cycle assessments, laboratory tests, registry histories and enforcement decisions to distinguish substantive environmental improvement from strategic reformulation.

The baseline seller classification captures the governance structure of the winning offer before the deadline. It does not identify who authored or revised content shared at the product-page level. Seller-side contribution histories or controlled platform access would permit cleaner attribution of content changes. Research covering less centralised marketplaces could also test whether the observed seller-type differences extend beyond Amazon.

Generalisability is limited to sampled Amazon listings in France, Germany, Italy and the United Kingdom and to cleaning products, personal-care products and textiles. The post-deadline window is also too short to identify durable adaptation. Longer panels could examine additional categories, platforms and seller sizes while linking claim revisions to search visibility, consumer trust, conversion and price premiums. Such evidence would establish whether more verifiable environmental communication improves consumer decisions or merely redistributes attention across products.

7. Conclusions

The 36-wave EU–UK panel documents a marked reconfiguration of environmental claims around the EU anti-greenwashing transposition deadline. Generic claims declined, while qualified claims, evidence-bearing content, and verified-label display became more prevalent. The smaller decline in the certified stratum is consistent with greater access to defensible evidentiary resources. The stronger contrast among marketplace-retail offers identifies the offer's baseline organisational setting as a source of heterogeneity in adjustment intensity.

These findings refine the expected communication pattern around an EU transposition deadline. The observed changes concern the composition and evidential form of environmental information rather than simple claim withdrawal. Variation across EU Ecolabel strata and baseline seller settings is consistent with differences in available evidentiary resources and organisational content systems.

The estimates capture anticipatory adjustment around the common EU transposition deadline and should not be interpreted as national enactment or enforcement effects. Within this boundary, the study documents systematic variation in environmental communication across EU exposure, certification strata, and baseline seller settings at the digital point of sale.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su18179033/s1, Table S1: Detailed sample screening and product-identity reconciliation; Table S2: Complete matching diagnostics before and after matching; Table S3: Multilingual claim-classification examples and boundary cases; Table S4: Intercoder reliability by outcome and marketplace language; Table S5: Classifier and complete-page validation results; Table S6: Page availability and missing-state composition; Table S7: Full coefficient decomposition for the fixed-effects models; Table S8: Complete weekly event-time estimates; Table S9: Extended robustness and heterogeneity diagnostics; Table S10: Wild-cluster bootstrap sensitivity analysis for small-cluster inference; Supplementary S1: Claim-classification codebook; Supplementary S2: Replication-package manifest.

Author Contributions: Conceptualization, M.H.S. and O.A.; methodology, M.H.S. and O.A.; software, M.H.S. and O.A.; validation, M.H.S. and O.A.; formal analysis, M.H.S. and O.A.; investigation, M.H.S. and O.A.; re-sources, M.H.S. and O.A.; data curation, M.H.S. and O.A.; writing—original draft preparation, M.H.S. and O.A.; writing—review and editing, M.H.S. and O.A.; visualisation, M.H.S. and O.A.; supervision, M.H.S. and O.A.; project administration, M.H.S. and O.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The derived data and replication materials supporting the findings are available from the corresponding author upon reasonable request, subject to applicable restrictions. The replication package is maintained in a version-controlled project archive and includes analysis and classifier code, locked configurations, the codebook, annotation identifiers, derived measures, computational-environment files, manifests, reconstruction scripts, and a synthetic test fixture. Raw marketplace HTML, screenshots, OCR-derived text, and annotated claim text cannot be redistributed because of platform terms. EU Ecolabel registry data are publicly accessible through the European Commission's EU Ecolabel Product Catalogue at https://environmental-data.ec.europa.eu/ecolabel/index.html (accessed on 3 August 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

API	Application programming interface
CI	Confidence interval
CSR	Corporate social responsibility
EU	European Union
F1	Harmonic mean of precision and recall
GTIN	Global Trade Item Number
HTML	Hypertext Markup Language
IQR	Interquartile range
IRR	Incidence rate ratio
OCR	Optical character recognition
SHA-256	Secure Hash Algorithm 256-bit
SMD	Standardised mean difference
UK	United Kingdom
UTC	Coordinated Universal Time
XLM-RoBERTa	Cross-lingual Language Model–RoBERTa

References

1. Polonsky, M.J.; Rotman, J.; Weber, V.; Kumar, P. How meaningless and substantive green claims jointly determine product environmental perceptions. *Int. J. Advert.* **2025**, *44*, 396–427. [CrossRef]
2. European Commission. Consumer Protection: Enabling Sustainable Choices and Ending Greenwashing. Available online: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_1692 (accessed on 1 August 2026).
3. Neureiter, A.; Stubenvoll, M.; Matthes, J. Is it greenwashing? Environmental compensation claims in advertising, perceived greenwashing, political consumerism, and brand outcomes. *J. Advert.* **2024**, *53*, 511–529. [CrossRef]
4. Cho, M.; Ko, E.; Borenstein, B.E. The interaction effect of fashion retailer categories on sustainable labels: The role of perceived benefits, ambiguity, trust, and purchase intention. *Int. J. Advert.* **2024**, *43*, 1016–1038. [CrossRef]
5. Majer, J.M.; Henscher, H.A.; Reuber, P.; Fischer-Kreer, D.; Fischer, D. The effects of visual sustainability labels on consumer perception and behavior: A systematic review of the empirical literature. *Sustain. Prod. Consum.* **2022**, *33*, 1–14. [CrossRef]
6. Torma, G.; Thøgersen, J. Can a meta sustainability label facilitate more sustainable consumer choices? *Bus. Strategy Environ.* **2024**, *33*, 283–306. [CrossRef]
7. European Parliament; Council of the European Union. Directive (EU) 2024/825 of 28 February 2024 as regards empowering consumers for the green transition through better protection against unfair practices and through better information. *Off. J. Eur. Union* **2024**, *L 2024/825*. Available online: https://eur-lex.europa.eu/eli/dir/2024/825/oj/eng (accessed on 13 August 2026).
8. Guix, M.; Nájera Sánchez, J.J.; Bonilla Priego, M.J.; Font, X. The changing institutional logics behind sustainability reports from the largest hotel groups in the world in 2014, 2018 and 2021. *Tour. Manag.* **2025**, *106*, 105031. [CrossRef]

9. Font, X.; Gallego, I.; Mellinas Cánovas, J.P. Strategic responses to sustainability certification under regulatory uncertainty: A real-options perspective. *J. Sustain. Tour.* **2026**, 1–29. [CrossRef]
10. Li, J.; Chen, Z.; Kwon, O.; Wang, J. Modes of control in international digital commerce: Evidence from Amazon.com. *J. Int. Bus. Stud.* **2025**, *56*, 440–450. [CrossRef]
11. Bernini, F.; Giuliani, M.; La Rosa, F. Measuring greenwashing: A systematic methodological literature review. *Bus. Ethics Environ. Responsib.* **2024**, *33*, 649–667. [CrossRef]
12. Fella, S.; Bausa, E. Green or greenwashed? Examining consumers' ability to identify greenwashing. *J. Environ. Psychol.* **2024**, *95*, 102281. [CrossRef]
13. Arkoh, P.; Costantini, A.; Scarpa, F. Determinants of sustainability reporting: A systematic literature review. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 1578–1597. [CrossRef]
14. Hahn, R.; Kühnen, M. Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research. *J. Clean. Prod.* **2013**, *59*, 5–21. [CrossRef]
15. MacLennan, J.; De Catelle, W. The revision of EU greenwashing laws: A new framework of analysis. *Bus. Hum. Rights J.* **2025**, *10*, 456–462. [CrossRef]
16. Darby, M.R.; Karni, E. Free competition and the optimal amount of fraud. *J. Law Econ.* **1973**, *16*, 67–88. [CrossRef]
17. Singh, P.; Sahadev, S.; Wei, X.; Henninger, C.E. Modelling the antecedents of consumers' willingness to pay for eco-labelled food products. *Int. J. Consum. Stud.* **2023**, *47*, 1256–1272. [CrossRef]
18. Spence, M. Job market signaling. *Q. J. Econ.* **1973**, *87*, 355–374. [CrossRef]
19. Kirmani, A.; Rao, A.R. No pain, no gain: A critical review of the literature on signaling unobservable product quality. *J. Mark.* **2000**, *64*, 66–79. [CrossRef]
20. Connelly, B.L.; Certo, S.T.; Reutzel, C.R.; DesJardine, M.R.; Zhou, Y.S. Signaling theory: State of the theory and its future. *J. Manag.* **2025**, *51*, 24–61. [CrossRef]
21. Tan, S.Z.; Yang, L.; Loo, W.H. Green specificity: Igniting curiosity and arousing emotional ambivalence. *J. Retail. Consum. Serv.* **2026**, *88*, 104414. [CrossRef]
22. Baier, C.; Göttsche, M.; Hellmann, A.; Schiemann, F. Too good to be true: Influencing credibility perceptions with signaling reference explicitness and assurance depth. *J. Bus. Ethics* **2022**, *178*, 695–714. [CrossRef]
23. Coen, D.; Herman, K.; Pegram, T. Are corporate climate efforts genuine? An empirical analysis of the climate "talk–walk" hypothesis. *Bus. Strategy Environ.* **2022**, *31*, 3040–3059. [CrossRef]
24. Delmas, M.A.; Burbano, V.C. The drivers of greenwashing. *Calif. Manag. Rev.* **2011**, *54*, 64–87. [CrossRef]
25. Hilton, J. Greenhush and greenwash: A signalling game analysis of strategic environmental disclosure. *Environ. Dev. Econ.* **2025**, 1–22. [CrossRef]
26. Fiechter, P.; Hitz, J.-M.; Lehmann, N. Real effects of a widespread CSR reporting mandate: Evidence from the European Union's CSR Directive. *J. Account. Res.* **2022**, *60*, 1499–1549. [CrossRef]
27. Christensen, H.B.; Hail, L.; Leuz, C. Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Rev. Account. Stud.* **2021**, *26*, 1176–1248. [CrossRef]
28. Fröberg, E.; Kolesova, S.; Rosengren, S. Does the label fit the channel? How "bricks" and "clicks" influence demand for environmental and social sustainability labels. *J. Serv. Res.* **2024**, *27*, 124–140. [CrossRef]
29. Valero-Gil, J.; Daddi, T.; Scarpellini, S.; Marrucci, L. Determinants and benefits of over-certification: A signaling theory perspective. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 5984–5999. [CrossRef]
30. Wittreich, T. Environmental certification, consumer greenness, and greenwashing. *Manag. Decis. Econ.* **2025**, *46*, 2289–2313. [CrossRef]
31. European Commission. EU Ecolabel Facts and Figures. Available online: https://environment.ec.europa.eu/topics/circular-economy-topics/eu-ecolabel/businesses/ecolabel-facts-and-figures_en (accessed on 1 August 2026).
32. Iraldo, F.; Barberio, M. Drivers, barriers and benefits of the EU Ecolabel in European companies' perception. *Sustainability* **2017**, *9*, 751. [CrossRef]
33. Bei, Z.; Gielens, K. The one-party versus third-party platform conundrum: How can brands thrive? *J. Mark.* **2023**, *87*, 253–274. [CrossRef]
34. Sikdar, S.; Kadiyali, V.; Hooker, G. Does Amazon's dual role weaken marketplace competition? *J. Mark.* 2026, *advance online publication*. [CrossRef]
35. Danani, S.; Behl, A.; Gupta, N. Factors influencing platform owners' seller control choices in E-marketplaces. *Inf. Organ.* **2025**, *35*, 100555. [CrossRef]
36. Cutolo, D.; Kenney, M. Platform-dependent entrepreneurs: Power asymmetries, risks, and strategies in the platform economy. *Acad. Manag. Perspect.* **2021**, *35*, 584–605. [CrossRef]

37. European Commission. La Commission Demande à la France de Transposer la Directive sur les Règles Donnant aux Consommateurs les Moyens D'agir en Faveur de la Transition Écologique. Available online: https://france.representation.ec.europa.eu/informations-et-evenements/informations/la-commission-demande-la-france-de-transposer-la-directive-sur-les-regles-donnant-aux-consommateurs-2026-05-28_fr (accessed on 1 August 2026).
38. Federal Republic of Germany. Drittes Gesetz zur Änderung des Gesetzes gegen den unlauteren Wettbewerb vom 12. Februar 2026 (BGBl. 2026 I Nr. 43). *Bundesgesetzblatt*, 19 February 2026. Available online: https://www.recht.bund.de/bgbl/1/2026/43/VO.html (accessed on 13 August 2026).
39. Italian Republic. Decreto legislativo 20 febbraio 2026, n. 30: Attuazione della direttiva (UE) 2024/825. *Gazzetta Ufficiale, Serie Generale*, 9 March 2026. Available online: https://www.gazzettaufficiale.it/eli/id/2026/03/09/26G00047/sg (accessed on 13 August 2026).
40. Competition and Markets Authority. Making Green Claims: Getting It Right, Across the Supply Chain. Available online: https://www.gov.uk/government/publications/making-green-claims-getting-it-right-across-the-supply-chain (accessed on 13 August 2026).
41. European Commission. One Label. More than 100,000 Greener Products. One Catalogue. Available online: https://environment.ec.europa.eu/news/one-label-more-100000-greener-products-one-catalogue-2025-09-17_en (accessed on 13 August 2026).
42. Greifer, N.; Stuart, E.A. Matching methods for confounder adjustment: An addition to the epidemiologist's toolbox. *Epidemiol. Rev.* **2021**, *43*, 118–129. [CrossRef] [PubMed]
43. Sen, I.; Floeck, F.; Weller, K.; Weiss, B.; Wagner, C. A total error framework for digital traces of human behavior on online platforms. *Public Opin. Q.* **2021**, *85*, 399–422. [CrossRef]
44. Barberá, P.; Boydstun, A.E.; Linn, S.; McMahon, R.; Nagler, J. Automated text classification of news articles: A practical guide. *Political Anal.* **2021**, *29*, 19–42. [CrossRef]
45. van Atteveldt, W.; van der Velden, M.A.C.G.; Boukes, M. The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Commun. Methods Meas.* **2021**, *15*, 121–140. [CrossRef]
46. Kroon, A.; Welbers, K.; Trilling, D.; van Atteveldt, W. Advancing automated content analysis for a new era of media effects research: The key role of transfer learning. *Commun. Methods Meas.* **2024**, *18*, 142–162. [CrossRef]
47. Olden, A.; Møen, J. The triple difference estimator. *Econom. J.* **2022**, *25*, 531–553. [CrossRef]
48. MacKinnon, J.G.; Nielsen, M.Ø.; Webb, M.D. Cluster-robust inference: A guide to empirical practice. *J. Econom.* **2023**, *232*, 272–299. [CrossRef]
49. Roth, J.; Sant'Anna, P.H.C.; Bilinski, A.; Poe, J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *J. Econom.* **2023**, *235*, 2218–2244. [CrossRef]
50. TeBlunthuis, N.; Hase, V.; Chan, C.-H. Misclassification in automated content analysis causes bias in regression: Can we fix it? Yes we can! *Commun. Methods Meas.* **2024**, *18*, 278–299. [CrossRef]
51. Rambachan, A.; Roth, J. A more credible approach to parallel trends. *Rev. Econ. Stud.* **2023**, *90*, 2555–2591. [CrossRef]
52. Kapoor, S.; Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **2023**, *4*, 100804. [CrossRef] [PubMed]
53. Nemes, N.; Scanlan, S.J.; Smith, P.; Smith, T.; Aronczyk, M.; Hill, S.; Lewis, S.L.; Montgomery, A.W.; Tubiello, F.N.; Stabinsky, D. An integrated framework to assess greenwashing. *Sustainability* **2022**, *14*, 4431. [CrossRef]
54. Schmuck, D.; Matthes, J.; Naderer, B. Misleading consumers with green advertising? An affect–reason–involvement account of greenwashing effects in environmental advertising. *J. Advert.* **2018**, *47*, 127–145. [CrossRef]
55. Guercini, S.; Monteverde, G.; Runfola, A.; Milanesi, M. To tell or not to tell? On the under communicating of sustainability in digital environments. *Bus. Strategy Environ.* **2025**, *34*, 8118–8133. [CrossRef]
56. Huang, Y.; Francoeur, C.; Zhang, S.; Brammer, S. Conditional silence: Organizational status and under-communication of environmental performance. *Br. J. Manag.* **2025**, *36*, 1219–1239. [CrossRef]
57. Heras-Saizarbitoria, I.; Boiral, O.; Allur, E.; García, M. Communicating environmental management certification: Signaling without signals? *Bus. Strategy Environ.* **2020**, *29*, 422–431. [CrossRef]
58. Rietveld, J.; Schilling, M.A. Platform competition: A systematic and interdisciplinary review of the literature. *J. Manag.* **2021**, *47*, 1528–1563. [CrossRef]
59. Tomar, S. Greenhouse gas disclosure and emissions benchmarking. *J. Account. Res.* **2023**, *61*, 451–492. [CrossRef]
60. Xiao, M.; Myers, P. Pride and prejudice and country-of-origin ecological images: The influence of COO ecological image on consumer evaluation of product greenness and green claim credibility. *Environ. Commun.* **2022**, *16*, 473–489. [CrossRef]
61. Deng, Z.; Sumbal; Li, L.; Khaskheli, M.B. Inclusive and sustainable digital technological transformation, legal innovation in marketing and business performance of China and the European Union. *Sustainability* **2026**, *18*, 2523. [CrossRef]

62. Majid, S.; Zhang, X.; Khaskheli, M.B.; Hong, F.; King, P.J.H.; Shamsi, I.H. Eco-efficiency, environmental and sustainable innovation in recycling energy and their effect on business performance: Evidence from European SMEs. *Sustainability* **2023**, *15*, 9465. [CrossRef]
63. Christis, J.; Wang, Y. Communicating environmental CSR towards consumers: The impact of message content, message style and praise tactics. *Sustainability* **2021**, *13*, 3981. [CrossRef]

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