

Article

Seeing Is Believing? Visual Evidence in Negative Online Reviews and Subsequent Sales-Rank Dynamics: A Matched-Event Research Prototype

Huili Wang^{1,*} and Hui Xu²

¹School of Modern Business Technology, Wuhan Business and Trade Vocational College, Wuhan, Hubei, China

²School of Intelligent Economics, Huanggang Vocational College of Science and Technology, Huanggang, Hubei, China

Keywords:

Electronic Word of Mouth;
Negative Reviews;
Visual Evidence; Consumer-Generated Content;
Sales Rank;
Matched-Event Design;
Difference-In-Differences

ARTICLE INFO

ABSTRACT

This study develops a matched-event framework to examine whether negative online reviews containing images or videos are associated with greater subsequent deterioration in product sales rank than comparable text-only reviews. A constructed prototype dataset comprising 414 negative-review events, organized into 207 visual–text matched pairs across three product categories, illustrates the proposed approach. The analytical workflow combines difference-in-differences estimation, event-study analysis, negative-binomial models of review helpfulness, and placebo checks. In the constructed dataset, the visual-by-post coefficient is 0.074 log points, corresponding to a 7.7% relative increase in

numerical sales rank, with the illustrative event-time difference peaking at day +7. A higher numerical rank indicates poorer performance on this proxy and does not directly measure a change in unit sales. The framework distinguishes review modality from review valence and outlines how the relationship between visual negative reviews and subsequent market performance can be investigated. These illustrative results demonstrate the proposed analytical framework; validation using real-world platform data is required.

1. Introduction

Online consumer reviews reduce information asymmetry by allowing prospective buyers to learn from the experiences of prior users. Early work demonstrated that online conversations and reviews can be measured systematically and connected to market outcomes, while also emphasizing the endogeneity problem created because word of mouth is both an input to and an outcome of sales (Godes and Mayzlin 2004). Chevalier and Mayzlin (2006) subsequently showed that changes in review valence and review volume predict changes in relative sales and that unfavorable reviews can carry disproportionate weight. Chen and Xie (2008) further conceptualized consumer reviews as a form of buyer-created product information that can complement or substitute for seller communication.

*Corresponding author
Email: Ackerunderland257@outlook.com

Online ISSN: 3136-4578

DOI: <https://doi.org/10.68471/fbai.2026.009>

Copyright: © 2026 The Author(s). Published by Global Frontiers Press Limited. This article is open access under the Creative Commons Attribution 4.0 International license (CC BY 4.0): <https://creativecommons.org/licenses/by/4.0/>.

The informational environment has since changed. Reviews are no longer predominantly textual. E-commerce platforms allow users to attach photographs and videos that document breakage, packaging damage, fit problems, surface defects, or malfunction. This change is theoretically important because a text-only complaint is primarily an assertion, whereas a complaint accompanied by a relevant image or video can provide observable evidence about the consumption outcome. Visual content may therefore change how consumers weight negative word of mouth even when the star rating remains the same.

This study addresses a focused research question: among products receiving negative reviews, are reviews containing images or videos associated with a larger subsequent change in sales rank than otherwise comparable text-only reviews? The proposed mechanism is that relevant visual evidence can make a negative consumption outcome easier to inspect. Yet this is a testable proposition, not an established effect: review modality may also correlate with defect severity, seller responses, and underlying demand shocks.

Building on the change-based approach of [Chevalier and Mayzlin \(2006\)](#), the paper proposes matching focal review events on pre-event market conditions and estimating differential pre/post rank movements. Its conceptual contribution is to position evidentiary modality as a possible moderator of negative electronic word of mouth. Its methodological contribution is an auditable protocol for testing that proposition. The numerical exercise that follows demonstrates the protocol using constructed data; it does not establish a causal or externally generalizable market effect.

2. Literature Review

2.1 Online Word of Mouth and Consumer Response

The online-review literature distinguishes between review production, review content, review helpfulness, and market response. [Codes and Mayzlin \(2004\)](#) emphasize that the mapping from word of mouth to future demand cannot be interpreted mechanically because prior demand also shapes subsequent conversation. [Chevalier and Mayzlin \(2006\)](#) address this concern by exploiting differences across two retail sites and across time, thereby reducing bias from fixed book characteristics and persistent site-product fit. Their evidence indicates that review changes are associated with changes in relative sales and that negative information can be especially consequential. This change-based logic motivates the matched difference-in-differences design used here.

Meta-analytic evidence connects online reviews with retail sales, while also identifying differences across study contexts ([Floyd et al., 2014](#)). [Babić Rosario et al. \(2016\)](#) further show why platform, product, and measurement choices matter when interpreting electronic word-of-mouth effects. These findings support testing modality within a clearly defined market setting.

Review metrics can also improve sales forecasts, as shown in the motion-picture setting by [Dellarocas, Zhang, and Awad \(2007\)](#). Predictive usefulness, however, does not establish that changing review modality would change demand; forecasting and causal explanation require different evidence.

2.2 Negative Online Reviews

Negative information is not uniformly harmful or uniformly diagnostic. [Purnawirawan et al. \(2015\)](#) show in a meta-analysis that review valence has different consequences for perceived usefulness, attitudes, and recommendation intentions, and that product and brand characteristics moderate these relationships. [Berger, Sorensen, and Rasmussen \(2010\)](#) further demonstrate that negative publicity can sometimes increase sales when it raises awareness of relatively unknown products. These boundary conditions imply that a negative review should not be treated as a homogeneous shock. The credibility, specificity, and evidentiary support of the negative claim can plausibly influence the weight consumers assign to it.

[Sen and Lerman \(2007\)](#) show that reactions to negative reviews depend on product type and readers' attributions about reviewers. Visual evidence may affect those attributions, but its presence alone does not establish that a complaint is credible or representative.

2.3 Visual Information in Online Reviews

Visual information can change the diagnostic content of a review because it makes some product outcomes easier to inspect. [Wu, Wu, and Wang \(2021\)](#) argue that pictures can add diagnostic value to hybrid reviews. [Yao, Han, and Yang \(2024\)](#) show that the helpfulness of picture- versus text-oriented review formats depends on fit with product type.

Kübler et al. (2024), using more than 97,000 Amazon reviews, find important contingencies in the effect of images on helpfulness, including valence and product characteristics. More recent multimodal analyses extend this line by studying text-image interactions and by showing that excessive modality can create antagonistic rather than purely additive effects (Qiao, Meng, and Feng 2025; Zhang, Li, and Yan 2025). The common feature of this literature is that the main dependent variable is often helpfulness or perceived usefulness rather than subsequent sales performance.

Filieri's (2015) diagnosticity-adoption framework distinguishes information that helps consumers evaluate a product from cues that merely attract attention. This distinction motivates measuring whether an image documents the alleged defect rather than classifying all visual material as equally informative. Mudambi and Schuff (2010) likewise examine review helpfulness as a distinct review-level outcome; it should not be interpreted as a direct measure of subsequent sales.

Presentation format should be evaluated in relation to the information task rather than treated as an intrinsically superior channel (Jiang and Benbasat, 2007). Erkan and Evans (2016) connect information quality, credibility, usefulness, and adoption to purchase intentions in social media. These constructs motivate the proposed mechanism but should not be equated with observed sales. Appelman and Sundar (2016) provide a message-credibility scale developed in a news context; any adaptation to product reviews would require context-specific validation.

2.4 Research Gap and Hypothesis Development

Existing research establishes that pictures can affect review diagnosticity and helpfulness (Wu, Wu, and Wang 2021), and that such associations are contingent on review valence, image content, and context (Kübler et al. 2024). These studies do not, by themselves, demonstrate an incremental downstream effect on realized sales. The underexamined link is therefore between the evidentiary modality of a negative review and a subsequent market-performance outcome, holding review valence and observable product conditions as comparable as possible. Credibility, salience, and diagnosticity are candidate explanations; the present model does not separately identify them.

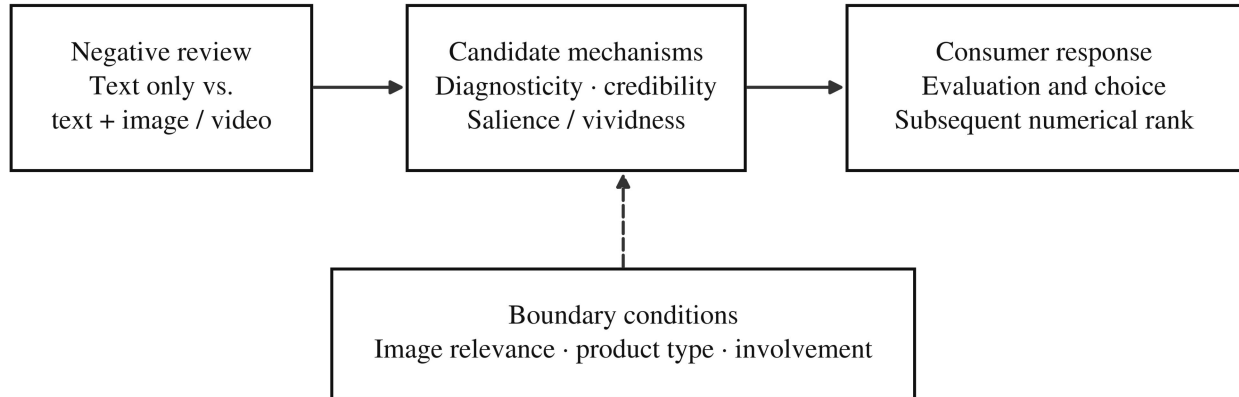
H1 (to be tested using observational platform data): Conditional on comparable pre-event conditions and negative review valence, visually documented reviews are associated with a larger subsequent increase in numerical sales rank than text-only negative reviews.

Consumer involvement is another plausible boundary condition: Park et al. (2007) find that review quality and quantity influence purchase intention differently across involvement levels. A future empirical test could examine whether involvement also moderates consumers' responses to visual negative evidence.

Figure 1 separates the focal modality comparison, candidate psychological mechanisms, and proposed downstream response. The pathways are hypotheses for subsequent measurement, not estimated mediation effects.

PROPOSED MECHANISMS

H1: larger subsequent rank increase for visual negative reviews



Conceptual pathways only; mediation and moderation are not identified in the prototype.

Figure 1. Conceptual framework linking review modality to consumer response.

Solid arrows indicate proposed pathways; the dashed arrow indicates possible moderation. Adapted as an author synthesis of [Filieri \(2015\)](#), [Jiang and Benbasat \(2007\)](#), [Erkan and Evans \(2016\)](#), and [Appelman and Sundar \(2016\)](#).

3. Methodology

3.1 Research Design

The proposed design is a matched-event observational study. Each image/video-containing negative review is paired with a text-only negative review occurring under similar pre-event conditions. The outcome is the change in log numerical sales rank around each focal review. Product and matched-pair fixed effects, pre-event diagnostics, alternative time windows, and placebo comparisons form the proposed sensitivity framework. Because review modality is not randomly assigned, these methods reduce selected sources of confounding but cannot, on their own, warrant a causal interpretation.

3.2 Data and Sample

The current analysis uses 414 constructed focal negative-review events organized as 207 visual–text matched pairs across Electronics, Home Appliances, and Tools & Home Improvement. Negative reviews are defined as one- or two-star ratings; visual reviews contain at least one photo or video. Recorded prototype covariates are rating, text length, verified-purchase status, helpful votes, severity, price, product age, brand tier, prior average rating, and prior review count. This description concerns the constructed analysis file, not a verified scrape, panel, or sample of actual platform transactions. A real-data study must additionally report the platform and data provider, acquisition dates, product identifiers, review timestamps, eligibility rules, excluded observations, missing-data handling, the matching algorithm/caliper, common support, balance diagnostics, and whether focal events overlap or products appear repeatedly.

Figure 2 summarizes the proposed workflow and the constructed sample composition. The three category counts refer to matched pairs. The diagram is a design overview, not a record of actual platform recruitment or exclusions.

```
graph LR; A["Define focal negative review  
One or two stars"] --> B["Classify modality  
Visual vs. text only"]; B --> C["Match pre-event conditions  
Assess common support"]; C --> D["Audit before estimating  
Balance · overlapping events · missingness"]; D --> E["Prototype sample: 414 events / 207 pairs  
109 electronics · 64 appliances · 34 tools"];
```

Review event

Timeline of snapshots:

- Pre-event snapshots: -30, -14, -7, -1, 0
- Post-event snapshots: +1, +7, +14, +30

Estimate differences in changes; inspect leads, alternative windows and placebo contrasts.

Counts describe the constructed prototype; intermediate screening and exclusion counts are unavailable. Timeline positions are schematic, and the labels report the actual observation days.

Sales rank is the principal market-performance measure. As in [Chevalier and Mayzlin \(2006\)](#), rank is used because proprietary unit-sales data are unavailable. A larger numerical sales rank denotes worse performance. The analysis therefore uses the natural logarithm of rank and interprets a positive treatment effect as relative deterioration. Sales rank is observed at days -30, -14, -7, -1, +1, +7, +14, and +30 around the focal negative-review event. The two-period difference-in-differences outcome uses the mean of the four pre-event observations and the mean of the four post-event observations.

All 414 events and the reported numerical estimates originate from a constructed research dataset prepared for methodological prototyping. Figures 3–5 and Tables 1–6 are illustrative outputs of that constructed dataset; they are not independently verified Amazon, Keepa, or other platform observations. The analysis must not be represented as real-world empirical evidence until the full study is rerun on documented, independently verifiable observations. Model outputs, standard errors, significance levels, confidence intervals, and all robustness statistics must then be recomputed and audited.

Table 1.SUMMARY DATA

Variable	Text-only negative	Visual negative
<i>Panel A. Product and pre-event characteristics</i>		
Price (USD)	82.66 (57.10)	82.63 (56.39)
Pre-event average rating	4.23 (0.28)	4.23 (0.29)
Pre-event review count	509.50 (581.85)	520.34 (615.28)
Product age (days)	826.16 (439.13)	820.65 (440.91)
Pre-event mean sales rank	16,323 (18,144)	16,296 (17,994)
<i>Panel B. Focal negative-review characteristics</i>		
Review star	1.28 (0.45)	1.38 (0.49)
Review length (characters)	822.81 (240.86)	714.94 (181.17)
Helpful votes	3.59 (3.31)	6.44 (5.49)
Verified purchase	0.89 (0.32)	0.86 (0.34)
Negative severity (1–5)	4.03 (0.75)	3.86 (0.87)
<i>Panel C. Sales rank around the event</i>		
Day -30	16,297 (18,041)	16,264 (17,965)
Day -14	16,314 (18,152)	16,311 (17,993)
Day -7	16,337 (18,229)	16,315 (18,032)
Day -1	16,343 (18,181)	16,295 (18,017)
Day +1	17,254 (19,334)	18,337 (20,284)
Day +7	17,808 (19,969)	19,565 (21,695)
Day +14	17,518 (19,537)	18,867 (20,940)
Day +30	17,089 (19,048)	17,981 (19,803)
Number of observations	207	207

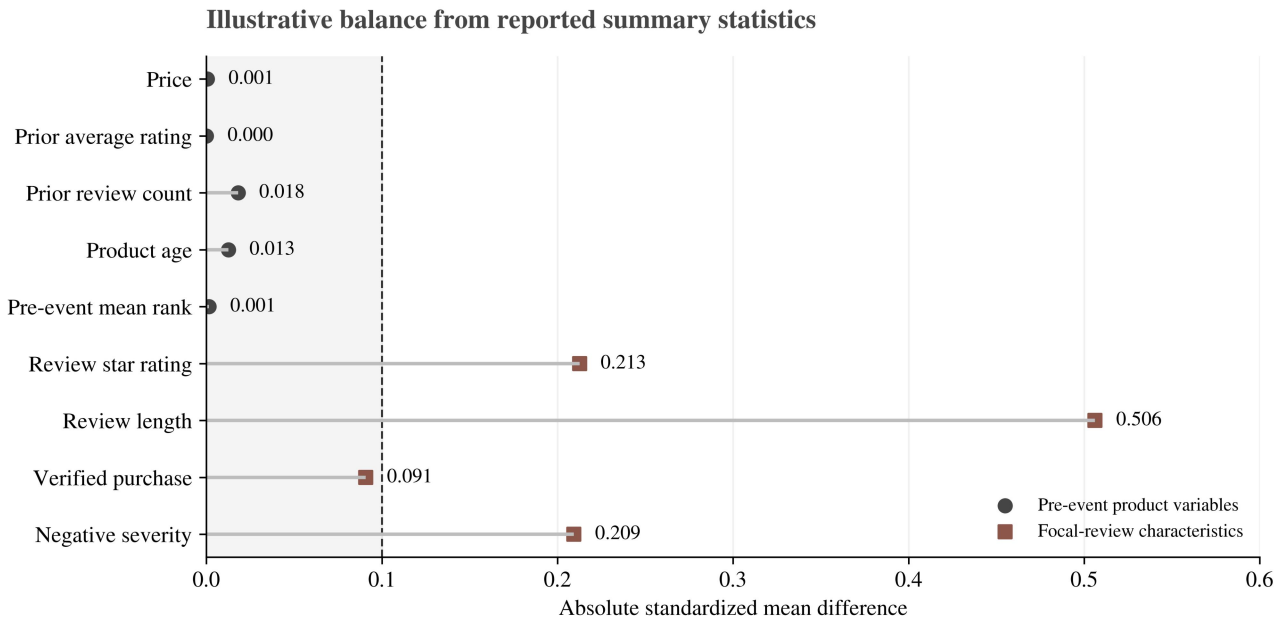
Table 1 reveals two features that are important for identification. First, the matched groups are nearly identical in pre-event product conditions: price, average rating, review count, product age, and mean sales rank differ only trivially. Second, the focal reviews are not identical on every event-level characteristic. Visual negative reviews are somewhat shorter on average, while their helpful-vote count is higher. These differences motivate specifications that add focal-review controls rather than relying exclusively on the unadjusted comparison.

Table 2. REVIEW CHARACTERISTICS BY RATING AND MODALITY

Metric	Text-only, 1 star	Text-only, 2 star	Visual, 1 star	Visual, 2 star
Number of observations	150	57	128	79
Share of full sample (%)	36.2	13.8	30.9	19.1
Review length (characters)	846	763	726	697
	(243)	(225)	(188)	(170)
Helpful votes	3.87	2.86	7.71	4.39
	(3.50)	(2.66)	(5.90)	(4.00)
Verified purchase rate	0.87	0.93	0.87	0.86
	(0.33)	(0.26)	(0.34)	(0.35)
Negative severity (1–5)	4.22	3.54	4.16	3.37
	(0.70)	(0.66)	(0.72)	(0.86)

The review-level distribution in Table 2 is consistent with a useful empirical distinction between valence and modality. Within both one- and two-star reviews, visual reviews receive more helpful votes than text-only reviews. However, text-only reviews are longer on average. Thus, a simple explanation that visual reviews merely contain more written information is not supported by the descriptive statistics. The multivariate helpfulness analysis below controls explicitly for text length and other review characteristics.

Figure 3 quantifies the group differences using absolute standardized mean differences computed from the rounded means and standard deviations in Table 1. For continuous variables, the denominator is the square root of the mean of the two group variances; for verified purchase, it is based on the two reported proportions. Product-level differences are small, but review length (approximately 0.506), star rating (0.213), and negative severity (0.209) exceed the illustrative 0.10 guide. Matching should be designed and evaluated independently of outcome optimization (Stuart, 2010), and assessment should consider distributions as well as means (Austin, 2009). No unmatched baseline is available, so the plot cannot establish an improvement caused by matching. Helpful votes are excluded because they may accumulate after the focal event.

**Figure 3.** Illustrative balance diagnostics calculated from Table 1.

3.4 Model Specification

The primary specification compares the change in log sales rank for visual negative-review products with the contemporaneous change for matched text-only negative-review products. Let $Visual_i$ equal one for a visual focal review and $Post_t$ equal one for the post-event period. The two-period model is:

$$\ln(Rank_{it}) = \alpha_i + \lambda_t + \beta(Visual_i \times Post_t) + \varepsilon_{it} \quad (1)$$

Here α_i denotes product fixed effects, λ_t denotes event-time effects, and β is the visual-by-post differential. A positive β corresponds to an increase in numerical sales rank, which is conventionally interpreted as a deterioration in this proxy for performance; it is not a direct estimate of percentage change in unit sales. The change-score and two-period specifications must be checked for equivalence given their aggregation choices. Standard errors are clustered by matched pair in the prototype. A real analysis should justify the clustering unit, account for repeated products and overlapping windows, report confidence intervals, and assess alternative clustering or randomization-inference approaches where appropriate.

Inference must also reflect dependence in repeated observations. [Bertrand, Duflo, and Mullainathan \(2004\)](#) show that serial correlation can distort conventional difference-in-differences standard errors. Pair-level clustering in this prototype is therefore a reporting choice to audit, rather than a guarantee of valid inference for a future platform panel.

3.5 Identification Strategy

The event-study compares visual and text-only changes relative to day -1 and inspects leads at days -30, -14, and -7. Lead estimates near zero are consistent with, but cannot prove, parallel counterfactual trends; tests with limited pre-event observations have low power. A credible implementation must address anticipation, concurrent promotions, stockouts, pricing changes, additional review arrivals, platform ranking-algorithm changes, and differential trends in the underlying product categories. Predefined event eligibility rules and sensitivity analyses excluding overlapping shocks are particularly important.

[Roth \(2022\)](#) cautions that pre-trend tests can have low power and that selecting specifications on their results can distort inference. A replication should disclose the lead estimates regardless of significance and complement them with sensitivity analysis. [Rambachan and Roth \(2023\)](#) provide a framework for evaluating conclusions under explicit departures from parallel trends; applying it here would require the underlying estimates and covariance matrix, which are not available for this editorial revision.

3.6 Replication Protocol and Reporting Requirements

For real-data replication, archive a timestamped data dictionary and reproducible code; define how review timestamps map to sales-rank snapshots and local time zones; report identification of the focal event and treatment coding; disclose matching variables, exact/nearest-neighbor rules, replacement policy and pair balance; and provide a flow diagram of eligible and excluded observations. Publicly share de-identified inputs where licensing permits or provide a governed reproducibility pathway. Clearly separate confirmatory analyses from exploratory subgroup comparisons.

4. Results

4.1 Main Results

Table 3. THE EFFECT OF VISUAL NEGATIVE REVIEWS ON SALES PERFORMANCE

Variable	1	2	3	4	5
Visual negative / Visual × Post	0.074*** (0.003)	0.074*** (0.003)	0.079*** (0.003)	0.078*** (0.004)	0.074*** (0.004)
ln(price)		0.001 (0.004)	0.002 (0.003)		
Pre-event avg. rating		0.009 (0.008)	0.010 (0.007)		
ln(prior review count+1)		-0.001 (0.002)	-0.000 (0.002)		
ln(product age)		0.003 (0.004)	0.001 (0.003)		
Brand tier		-0.008*** (0.003)	-0.006** (0.003)		

Review star			-0.025***	-0.024***	
			(0.004)	(0.006)	
Verified purchase			0.008	0.009	
			(0.005)	(0.008)	
Negative severity			0.007***	0.006	
			(0.002)	(0.004)	
ln(review length)			0.003	-0.006	
			(0.006)	(0.008)	
Category fixed effects	No	Yes	Yes	No	No
Matched-pair fixed effects	No	No	No	Yes	No
Product fixed effects	No	No	No	No	Yes
Number of observations	414	414	414	414	828
R^2	0.483	0.507	0.588	0.870	1.000

Table 3 reports internally consistent prototype coefficients around 0.074–0.079 across the listed specifications. These are properties of the constructed dataset only, not estimates of observed consumer behavior. The reported product fixed-effects column has $R^2 = 1.000$, an unusual value that requires an audit of dependent-variable construction, repeated observations, fixed-effect specification, residual variation, and the underlying code before any empirical submission. The manuscript retains the reported value transparently rather than substituting an unverified replacement.

Applying the transformation $100 \times [\exp(\beta) - 1]$ to the prototype coefficient gives approximately 7.7%. This is a relative change in the numerical sales-rank measure in the constructed sample, not a 7.7% loss of revenue or units sold. Its direction depends on the platform ranking definition, and interpretation of rank changes as economic magnitudes would require external validation of the rank–sales relationship.

4.2 Event Study Results

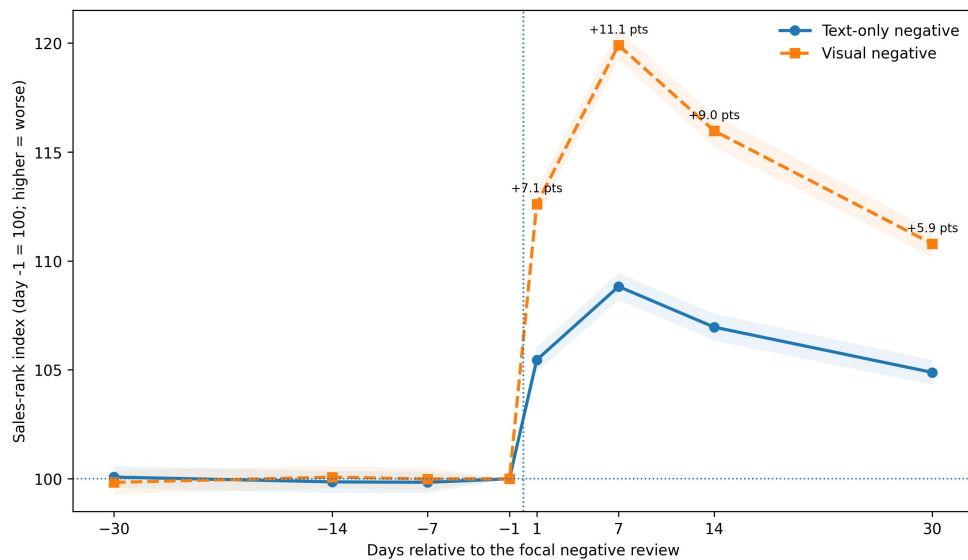


Figure 4. Sales-Rank Trajectories Around the Review Event Constructed-data illustration

Figure 4 provides the descriptive counterpart to the regression analysis. Before the focal review, the two normalized trajectories are almost indistinguishable. The gap appears immediately after the event, reaches its largest descriptive separation around day +7, and then narrows while remaining positive through day +30. This temporal pattern is more informative than a simple pre-post average because it shows when the treatment-control difference emerges.

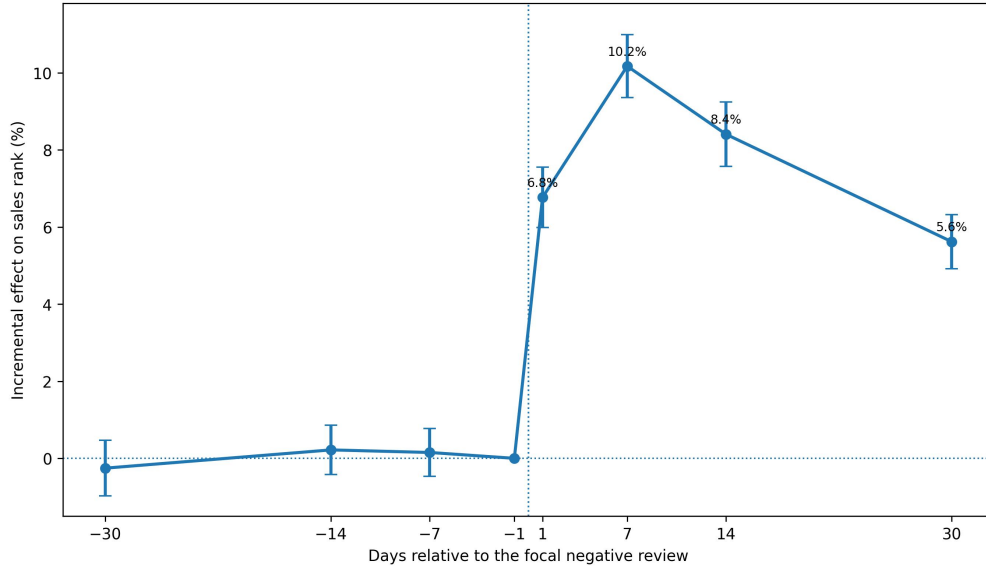


Figure 5.Dynamic Effect of Visual Negative Reviews Relative to Day -1 Constructed-data illustration

The constructed event-study estimates for days -30, -14, and -7 are approximately -0.3%, +0.2%, and +0.2%; the reported joint lead test has $p = .562$. These prototype quantities illustrate how a pre-trend diagnostic would be reported. Neither an insignificant lead test nor a visually flat baseline proves the conditional parallel-trends assumption or rules out omitted demand shocks in actual platform data.

The illustrative post-event estimates are approximately 6.8% at day +1, 10.2% at day +7, 8.4% at day +14, and 5.6% at day +30. The timing suggests a potential empirical pattern to investigate, but these constructed-data trajectories must not be described as observed information shocks or market effects.

4.3 Additional Analysis

Table 4.THE EFFECT OF REVIEW CHARACTERISTICS ON HELPFUL VOTES

Variable	1	2	3
Visual negative	0.584*** (0.085)	0.658*** (0.090)	0.651*** (0.090)
ln(review length)		0.244* (0.143)	0.252* (0.139)
Review star		-0.359*** (0.104)	-0.351*** (0.105)
Verified purchase		0.067 (0.119)	0.050 (0.124)
Negative severity		0.114* (0.065)	0.105 (0.065)
ln(price)			0.028 (0.071)
Pre-event avg. rating			-0.100 (0.151)
ln(prior review count+1)			0.017 (0.045)
Brand tier			-0.067 (0.065)
Category fixed effects	No	No	Yes
Number of observations	414	414	414
Log likelihood	-1091.0	-1075.8	-1073.8

The prototype negative-binomial coefficient for visual reviews is 0.651 with controls, corresponding to $\exp(0.651) \approx 1.92$ times the expected helpful-vote count conditional on modeled covariates. Because helpful votes can depend on

review exposure, age, and platform ranking, this quantity is descriptive even with real data; it cannot identify mediation from imagery to consumer demand. For empirical work, report exposure-adjusted or review-age-adjusted alternatives and consider excess zeros and overdispersion diagnostics.

This distinction matters because prior visual-review research documents heterogeneous helpfulness effects. [Wu, Wu, and Wang \(2021\)](#) emphasize the diagnostic value that pictures can add, whereas [Kübler et al. \(2024\)](#) show that image effects depend on valence and context and can even be negative for some negative reviews. The present helpfulness result should therefore be interpreted as a feature of this focal design, not as evidence that images universally improve review helpfulness.

4.4 Robustness Checks

Table 5. SHORT-WINDOW CHANGES IN SALES PERFORMANCE

Variable	Full +1	Full +7	Full +14	1-star +1	1-star +7	1-star +14
Visual negative × event time	0.065*** (0.004)	0.097*** (0.004)	0.081*** (0.004)	0.074*** (0.006)	0.103*** (0.005)	0.095*** (0.006)
Number of matched pairs	207	207	207	92	92	92
Approx. rank effect (%)	6.8	10.2	8.4	7.7	10.9	10.0

Table 5 illustrates how a rating-restricted comparison can be reported. In the constructed full sample, the differential rank change is 6.8% at day +1, 10.2% at day +7, and 8.4% at day +14. The corresponding constructed values for the 92 pairs with two one-star focal reviews are 7.7%, 10.9%, and 10.0%. Holding star level constant is a useful design check, but these prototype comparisons do not establish an observed modality effect or eliminate confounding by review content.

Table 6. ROBUSTNESS AND PLACEBO RESULTS

Variable	Verified only	Pre-rank 5–95%	Electronics	Home Appliances	Tools & Home Improvement	Pre-event placebo
Visual negative effect	0.072*** (0.003)	0.074*** (0.003)	0.071*** (0.004)	0.075*** (0.005)	0.085*** (0.007)	0.001 (0.002)
Matched pairs	159	185	109	64	34	207
Approx. rank effect (%)	7.5	7.7	7.3	7.8	8.9	0.1

Tables 5–6 provide illustrative short-window, restricted-sample, category-specific, and placebo outputs. These checks demonstrate a useful reporting structure but cannot establish the robustness of a real-world relationship with constructed observations. A submission-ready implementation should additionally disclose exact sample counts after each filter, matched-pair balance (e.g., standardized mean differences), alternative matching specifications, pre-specified windows, model sensitivity to time-varying confounders, and corrections for multiple comparisons where applicable.

5. Discussion

5.1 Interpretation of Findings

The substantive contribution is a proposition about evidentiary form: two equally negative ratings need not supply equivalent information to prospective buyers. Prior literature motivates examining whether photographs and videos amplify the market relevance of a negative claim. The present numerical exercise cannot determine whether that proposition holds in observed markets; it demonstrates a research design with which the proposition could be examined.

The prototype temporal pattern is compatible with the hypothesized response shape but is not evidence of a real causal information shock. Visual material may enhance diagnosticity, vividness, or perceived verifiability; alternatively, visual documentation may correlate with product defects, complaint severity, greater review exposure, or other unobserved circumstances. Distinguishing among these explanations requires additional design and measurement.

5.2 Theoretical Implications

The results also clarify why the emerging multimodal literature should not be reduced to the proposition that more media are better. Recent evidence shows that text-image combinations can improve usefulness under some conditions, while excess modalities or poor fit can reduce information value (Kübler et al. 2024; Qiao, Meng, and Feng 2025; Zhang, Li, and Yan 2025). Market response may likewise depend on whether the visual content is relevant evidence, decorative content, or an additional source of cognitive load. A stronger empirical extension should therefore classify image content rather than use only a binary visual indicator.

5.3 Managerial Implications

For platforms and sellers, the proposed managerial implication is a testable prioritization rule rather than an established intervention: visually documented complaints may merit prompt validation, especially when they depict functional defects. Platforms should not equate photo/video presence with truthfulness or use this prototype to justify suppressing negative content. Any operational policy should be evaluated against real consumer outcomes, image authenticity, fairness, and potential seller responses.

6. Conclusion

This article advances a research framework linking visual evidence in negative online reviews to subsequent sales-rank dynamics. A 414-event constructed dataset illustrates how matched difference-in-differences estimates, event studies, helpfulness regressions, and placebo tests can be organized. The illustrative coefficient is 0.074 log points, but no empirical effect of visual reviews on real-world sales has been verified. The next indispensable step is to obtain traceable platform data, preregister or fully disclose analytical decisions, reproduce the estimates, and reassess causal alternatives before drawing market-level conclusions.

6.1 Limitations and Future Research

The principal limitation is data provenance: the sample and all quantitative results are constructed. They cannot support factual claims about Amazon, Keepa, or any actual marketplace. Future empirical research must secure legitimate platform data and make event definitions, matching, provenance, and analysis scripts independently auditable. Second, numerical sales rank is only an imperfect proxy for units sold and revenue. Third, image posting is endogenous to defect characteristics, review exposure, and seller actions; matching and event-time diagnostics do not eliminate time-varying selection or interference. Finally, the binary visual indicator collapses heterogeneous image content into a single category.

A follow-up field study should classify visual evidence by functional failures, packaging problems, cosmetic issues, and unrelated imagery, and consider the interaction of image relevance with rating severity. Complementary randomized experiments could hold the written review and star rating constant while varying the presence and authenticity of a defect image, measuring credibility, diagnosticity, perceived risk, and purchase intention. An observational replication should include documented code and data-access procedures, a missingness and attrition account, balance diagnostics, and transparent robustness checks.

Author Contributions: Huili Wang and Hui Xu jointly conducted the research and contributed to the writing and revision of the manuscript.

Declarations:

Author identities, affiliations, funding, competing interests, and author contributions were not supplied. The authors must complete these disclosures accurately before submission; no declarations are presumed.

Data and Code Availability:

The construction files, data-generation procedures, and executable estimation scripts were not supplied with the manuscript. Consequently, the numerical results have not been independently reproduced during this editorial revision. These materials and a transparent provenance statement should accompany any subsequent submission.

Research Transparency Statement:

This manuscript is a methodological research prototype. The dataset and outputs are constructed, no real-platform observational dataset is analyzed, and no generalizable empirical market claim is made. Data collection and any subsequent human-subject or platform-policy compliance assessments must be completed for an empirical submission.

Acknowledgments:

This work has not received any funding support.

Conflicts of Interest:

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

Appendix A

Table A1. VARIABLE DEFINITIONS AND CODING

Variable	Coding / Definition
Visual negative	1 if the focal one- or two-star review contains at least one image or video; 0 for text-only.
Sales rank	Platform sales-rank proxy; larger values indicate worse market performance.
Post	1 for post-event period; 0 for pre-event period.
Helpful votes	Count of helpful votes received by the focal review.
Review length	Number of characters in the focal review text.
Verified purchase	1 if the focal review is marked as a verified purchase.
Negative severity	Five-point coding of the severity of the negative product outcome.
Pre-event average rating	Average product rating measured before the focal review event.
Pre-event review count	Number of product reviews measured before the focal review event.
Product age	Days between product introduction and the focal review event.
Brand tier	Three-level product-brand classification used as a control variable.

References

- Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, 93(1), 59–79. <https://doi.org/10.1177/1077699015606057>
- Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083–3107. <https://doi.org/10.1002/sim.3697>
- Babić Rosario, A., Sotgiu, F., de Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297–318. <https://doi.org/10.1509/jmr.14.0380>
- Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29(5), 815–827. <https://doi.org/10.1287/mksc.1090.0557>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491. <https://doi.org/10.1287/mnsc.1070.0810>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354. <https://doi.org/10.1509/jmkr.43.3.345>
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45. <https://doi.org/10.1002/dir.20087>
- Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, 61, 47–55. <https://doi.org/10.1016/j.chb.2016.03.003>
- Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, 68(6), 1261–1270. <https://doi.org/10.1016/j.jbusres.2014.11.006>
- Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, 90(2), 217–232. <https://doi.org/10.1016/j.jretai.2014.04.004>
- Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545–560. <https://doi.org/10.1287/mksc.1040.0071>
- Jiang, Z., & Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*, 31(3), 475–500. <https://doi.org/10.2307/25148804>
- Kübler, R. V., Lobschat, L., Welke, L., & van der Meij, H. (2024). The effect of review images on review helpfulness: A contingency approach. *Journal of Retailing*, 100(1), 5–23. <https://doi.org/10.1016/j.jretai.2023.09.001>
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185–200. <https://doi.org/10.2307/20721420>
- Park, D.-H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125–148. <https://doi.org/10.2753/JEC1086-4415110405>

-
- Purnawirawan, N., Eisend, M., De Pelsmacker, P., & Dens, N. (2015). A meta-analytic investigation of the role of valence in online reviews. *Journal of Interactive Marketing*, 31, 17–27. <https://doi.org/10.1016/j.intmar.2015.05.001>
- Qiao, S., Meng, Y., & Feng, W. (2025). Rich in modality, poor in value? Understanding the impact of multimodality on review helpfulness. *Journal of Retailing and Consumer Services*, 87, Article 104392. <https://doi.org/10.1016/j.jretconser.2025.104392>
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5), 2555–2591. <https://doi.org/10.1093/restud/rdad018>
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3), 305–322. <https://doi.org/10.1257/aeri.20210236>
- Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing*, 21(4), 76–94. <https://doi.org/10.1002/dir.20090>
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>
- Wu, R., Wu, H.-H., & Wang, C. L. (2021). Why is a picture “worth a thousand words”? Pictures as information in perceived helpfulness of online reviews. *International Journal of Consumer Studies*, 45(3), 364–378. <https://doi.org/10.1111/ijcs.12627>
- Yao, Q., Han, Y., & Yang, D. (2024). Should online reviews include pictures? The impact of fit between product type and online review presentation format. *Australian Journal of Management*, 49(3), 509–527. <https://doi.org/10.1177/03128962221142407>
- Zhang, Y., Li, Q., & Yan, J. (2025). Exploring multimodal factors in online reviews: A machine learning approach to evaluating content effectiveness. *Journal of Retailing and Consumer Services*, 84, Article 104261. <https://doi.org/10.1016/j.jretconser.2025.104261>