AI Review Summaries and Cross-Platform Pricing: Evidence from Indian E-Commerce

Dheer Avashia*†

May 13, 2026

Abstract

AI-generated review summaries are among the first large-scale consumer-facing applications of generative AI in digital markets. Existing work documents effects on reviews, engagement, and purchase behavior; the cross-platform equilibrium pricing response remains largely unstudied. I construct a novel monthly matched-product price panel across Amazon India and Flipkart—1,196 identical products tracked over 86 months (2019–2026)—combining commercial Amazon histories (Keepa) with scraped, archived, and backfilled Flipkart prices. The empirical design exploits Amazon India's December 2023 rollout of AI review summaries against Flipkart's lack of an analogous product-detail-page feature.

The preferred strict-support specification estimates a 2.5 log-point Amazon-relative price decline, alternative counterfactual estimators generally imply 2–4 log-point declines, the slope-change specification implies a larger initial level shift followed by partial mean reversion, and short local windows with cleaner pre-trends produce small and imprecise near-rollout effects. The mechanism evidence tracks the model's comparative static: a continuous review-corpus interaction is more negative for products with larger pre-treatment review corpora, and the largest category-level estimate falls in Toys & Games, where review narratives carry the most decision-relevant content and documented vertical-entry exposure is low. The Amazon–Flipkart price gap was already moving differentially before treatment, so the aggregate event-study path cannot by itself deliver the causal interpretation. The most credible evidence comes from the strict-support sample, from low-entry/review-dependent categories, and from the cross-sectional heterogeneity that the model predicted ex ante.

*Second-year paper, University of Arizona
†Thank you to Dr. Matthijs Wildenbeest and Dr. Mo Xiao for valuable comments

1 Introduction

Product reviews shape consumer beliefs and equilibrium outcomes in digital markets (Chevalier and Mayzlin 2006; Sun 2012; Luca 2016; Nosko and Tadelis 2014). Their informational value, however, is bounded by what consumers can process. Search costs, limited attention, and information overload prevent consumers from extracting the full content of large review corpora (Nelson 1970; Moraga-González et al. 2017; Seiler 2013; Hu and Krishen 2019). AI-generated review summaries are an institutional response to this friction: they compress an unstructured review corpus into a concise product-level signal at no cost to the consumer. If summaries lower the cost of evaluating quality, they should reshape not only consumer behavior but the equilibrium prices that platforms and sellers can sustain.

This paper studies whether an AI tool that changes the consumer information environment shifts equilibrium prices on the treated platform relative to a close competitor. Existing studies focus on within-Amazon outcomes—purchase behavior, review engagement, and review production (Xie et al. 2025). The natural counterpart is a cross-platform equilibrium price response, identified in a market with a close competing platform that lacks an analogous feature.

A simple evaluation-cost model anchors the analysis. AI summaries lower the marginal cost of processing existing review information; if this raises consumers' propensity to compare outside options, the residual demand elasticity facing treated-platform sellers rises and equilibrium prices fall. The model also delivers a sharper prediction: the price response should be larger where there is more review information to compress and in categories where review narratives are more decision-relevant. This heterogeneity prediction is what allows the reduced-form price evidence to speak to mechanism, not only to direction.

I make three contributions. First, I construct a monthly matched-product price panel across Amazon India and Flipkart—1,196 ASIN–PID pairs over 86 months—that, to my knowledge, is the first cross-platform price asset of this scale for Indian e-commerce. Second, using the December 2023 rollout, I estimate a family of Amazon-relative price effects: the preferred strict-support specification implies a 2.5 log-point decline, alternative counterfactual estimators generally imply 2–4 log-point declines, and slope-change specifications reveal a larger initial decline followed by partial mean reversion. Third, the heterogeneity in this estimate matches the model's comparative static: a continuous interaction with pre-treatment review-corpus size is more negative for products with larger review corpora, and the largest category-level estimate falls in Toys & Games, where reviews carry the most decision-relevant content. The Amazon–Flipkart price gap was already moving differentially before treatment, so the aggregate event-study path cannot by itself deliver the causal interpretation. The most credible evidence comes from the strict-support sample, where the pre-trend problem is least severe, and from the cross-sectional heterogeneity—by review-corpus size and category—that the model predicted ex ante.

2 Related Work

This paper connects three literatures: AI-generated review summaries, the role of product reviews in digital markets, and cognitive constraints on consumer search.

The closest recent work studies Amazon's AI review summaries as a consumer-information intervention. Xie et al. (2025) estimate within-Amazon U.S. purchase responses to generative-AI summaries. Wang and Wang (2025) study sales-rank outcomes within Amazon. Su et al. (2024) examine how AI summaries affect review-writing behavior, while Alavi and Nozari (2025) show that summaries reduce engagement with individual reviews and can homogenize subsequent review content. Lei and Liu (2025) provide survey evidence on perceived usefulness and trust. Relative to these papers, the margin studied here is different: the data record prices for the same products on two competing platforms, not clicks, review reading, or purchase decisions. If AI summaries increase purchase likelihood or reduce review-reading costs, sellers and pricing algorithms may respond through prices; the design here studies that downstream cross-platform price response.

A large literature documents how product reviews shape consumer beliefs and demand. Chevalier and Mayzlin (2006) show that consumers respond to review text, not only star ratings, making review summaries economically relevant. Sun (2012) demonstrates that rating variance can signal niche appeal or disagreement. Jang, Prasad, and Ratchford (2012) find that reviews matter more in the consideration stage than at final choice, and Luca (2016) documents revenue effects of Yelp ratings. Nosko and Tadelis (2014) show how reputation information shapes platform behavior. That reviews matter for demand is necessary but not sufficient for the present question: the outcome here is a downstream price response, not a direct measure of review use.

The relevant friction is not only whether reviews exist but whether consumers can process them at low cost. Nelson (1970) distinguishes search and experience goods. Moraga-González et al. (2017) show how search-cost heterogeneity generates price dispersion; Seiler (2013) estimates limited consumer search in a dynamic framework; and Ariely and Lynch (2000) experimentally show that lower search costs lead to more competitive pricing. In e-commerce, Hu and Krishen (2019) show that large numbers of product reviews can reduce satisfaction through information overload, consistent with bounded working memory in information-rich settings (Sweller 1988). AI summaries may reduce the cost of processing quality information without changing the underlying product; the sign and timing of a price response are empirical questions because sellers, platforms, and consumers may all adjust.

The supply side of this setting is also connected to the literature on algorithmic marketplace pricing. Calvano et al. (2020) and Brown and MacKay (2023) study how pricing algorithms can change competitive conduct and price dynamics, while Aparicio and Misra (2023) survey how AI-based pricing systems use real-time demand learning and automated response, and Aparicio, Metzman, and Rigobon (2024) document high-frequency online retail price adjustment. The intervention I study is demand-side information design rather than a pricing algorithm itself, but the market architecture is similar: third-party sellers, buy-box competition, and automated repricing translate demand-side information shocks into equilibrium prices. This connection matters because an AI summary can affect prices even if it does not directly set prices; it changes the demand environment to which sellers and pricing systems respond.

3 Conceptual Framework

The framework makes two contributions to the empirics. First, it shows that the sign of the price response is not mechanical: it depends on whether AI summaries raise or lower the elasticity of demand facing treated-platform sellers, and either is possible a priori. Second, it generates testable cross-sectional predictions about which products and categories should show larger responses.

Consider a consumer evaluating product i with true quality v on platform p. The consumer observes price P_{ip} and receives indirect utility

$$u_{ip} = v_i - q_i^u - P_{ip} - c_{ip} + \varepsilon_{ip},$$

where q_i^u is residual uncertainty about product quality after processing available information, c_{ip} is the cost of evaluating the product information environment, and ε_{ip} is an idiosyncratic taste shock. Reviews reduce q_i^u, but reading many unstructured reviews is costly. An AI summary lowers the Amazon evaluation cost from c_{iA} to $c_{iA} - \Delta_i$ by making the review corpus easier to process. This is an information-design shock rather than a shock to product quality or seller marginal cost.

The price effect operates through residual demand. Search models establish that lowering information frictions can compress price premia and reduce equilibrium prices through the elasticity channel (Moraga-González, Sándor, and Wildenbeest 2017). Let Amazon residual demand for product i have elasticity $\epsilon_i(\Delta_i) > 1$ over the relevant price range and marginal cost m_i. The seller's pricing problem implies the Lerner condition,

$$\frac{P_{iA} - m_i}{P_{iA}} = \frac{1}{\epsilon_{iA}}.$$

Equivalently,

$$P_{iA}(\Delta_i) = m_i \frac{\epsilon_i(\Delta_i)}{\epsilon_i(\Delta_i) - 1}.$$

Differentiating the Lerner expression gives

$$\frac{\partial P_{iA}}{\partial \Delta_i} = -m_i \frac{\partial \epsilon_i(\Delta_i)/\partial \Delta_i}{\{\epsilon_i(\Delta_i) - 1\}^2}.$$

Two opposing channels can move the sign. Under the elasticity channel, summaries make consumers more willing to compare outside options once quality uncertainty is resolved, so $\partial \epsilon_i/\partial \Delta_i > 0$ and $\partial P_{iA}/\partial \Delta_i < 0$: the treated platform's price compresses relative to the comparison platform. Under a willingness-to-pay channel, summaries primarily raise consumer trust without intensifying cross-platform comparison; the sign is then weakly positive or zero. The empirical question is which channel dominates on average, and the negative estimate reported below is the sign predicted when the elasticity channel does.

The same expression yields a testable heterogeneity prediction. Let R_i denote the size of the pre-treatment review corpus and V_i denote dispersion in review information. If a summary is more valuable when there is more information to compress, then

$$\frac{\partial^2 \epsilon_i}{\partial \Delta_i \partial R_i} > 0 \qquad \text{and} \qquad \frac{\partial^2 \epsilon_i}{\partial \Delta_i \partial V_i} > 0,$$

implying $|\partial P_{iA}/\partial \Delta_i|$ is increasing in R_i and V_i. This is the cross-sectional implication that I take to the data through a continuous review-corpus interaction and through category-level estimates targeting review-dependent products such as Toys & Games. The framework therefore plays a coordinating role: it does not estimate the primitives of a search model, but it disciplines which heterogeneity patterns count as evidence on the elasticity channel and which do not. It connects the reduced-form price estimates to platform-design and online-search models in which information frictions shape elasticities, equilibrium prices, and price dispersion (Ellison and Ellison 2009; Dinerstein et al. 2018; Moraga-González et al. 2017).

4 Institutional Context

To be able to credibly leverage cross-platform pricing effects to identify the effect of an information design shock and extrapolate to implications on consumer search behavior it is beneficial to look at an enviorment where firms are simmilar and consumers are price-sensitive. India is a useful setting because Amazon and Flipkart are close competitors with substantial catalog overlap, frequent platform-wide sales events, and a consumer environment in which cross-platform price comparison is salient. The price-sensitivity argument here is institutional motivation, not an identifying assumption: the empirical design rests on matched products and platform-time variation, not on a behavioral claim that Indian consumers are uniquely price sensitive. The empirical design uses Amazon and Flipkart because they are close competitors in Indian e-commerce and many identical products can be matched across them (Figure 1). The comparison is informative but imperfect: Flipkart is itself an active platform with its own pricing, seller, and product-information dynamics over the sample period.

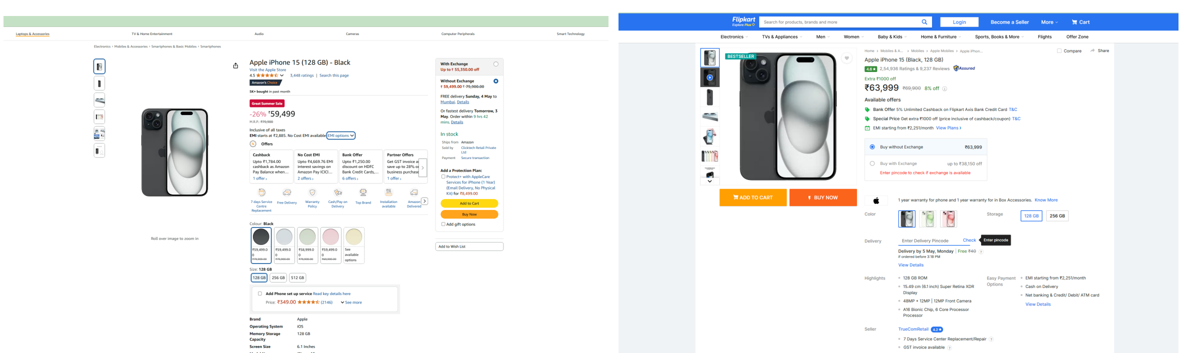


Figure 1: Same Product on Amazon and Flipkart

On November 30, 2023, Amazon announced its rollout of AI-generated review summaries and stated that "[t]he new AI tool will be available in India starting December 12."[1] The announcement fixes a platform-level treatment date but does not reveal product-by-product or interface-specific exposure. Archived page evidence shows that visible summaries appeared at different times across surfaces, consistent with a rollout that may have differed across mobile and desktop or included unobserved pilot exposure. The monthly analysis therefore codes January 2024 onward as the effective post period and reads sharp event-time tests cautiously.

Flipkart introduced other AI shopping tools during the sample period—notably the Flippi shopping assistant in October 2023[2] and multimodal search features tied to Flipkart Immerse.[3] These are not directly analogous to Amazon's review-summary block: they are product-discovery and assistant surfaces, not PDP-level summarizations of the review corpus. Flipkart's own 2023 trends reporting indicates that Flippi engaged roughly four million users in its first month against a platform with hundreds of millions of registered users, so the feature was visible but not plausibly universal in the early treatment window.[4] Archived Flipkart product-page checks in the data-collection workflow do not show a directly analogous PDP-level review summary during the analysis window. If Flippi or related tools nevertheless moved Flipkart prices, the resulting comparison is partial-treatment versus stronger-treatment rather than treatment versus pure control. Under same-sign AI information effects on both platforms, this would attenuate the Amazon-relative estimate toward zero, so the design's potential bias from Flipkart-side AI is weakly conservative.

5 Data

I assemble a matched product-level price panel across Amazon India and Flipkart by combining platform scraping, commercial price histories, and archived price records. The con-

[1] https://www.aboutamazon.in/news/retail/generative-ai-improves-customer-reviews

[2] https://inc42.com/buzz/flipkart-launches-chatgpt-powered-shopping-assistant-flippi/

[3] https://stories.flipkart.com/flipkart-immerse-multi-modal-search

[4] See, for example, Financial Express reporting on Flipkart's 2023 trends.

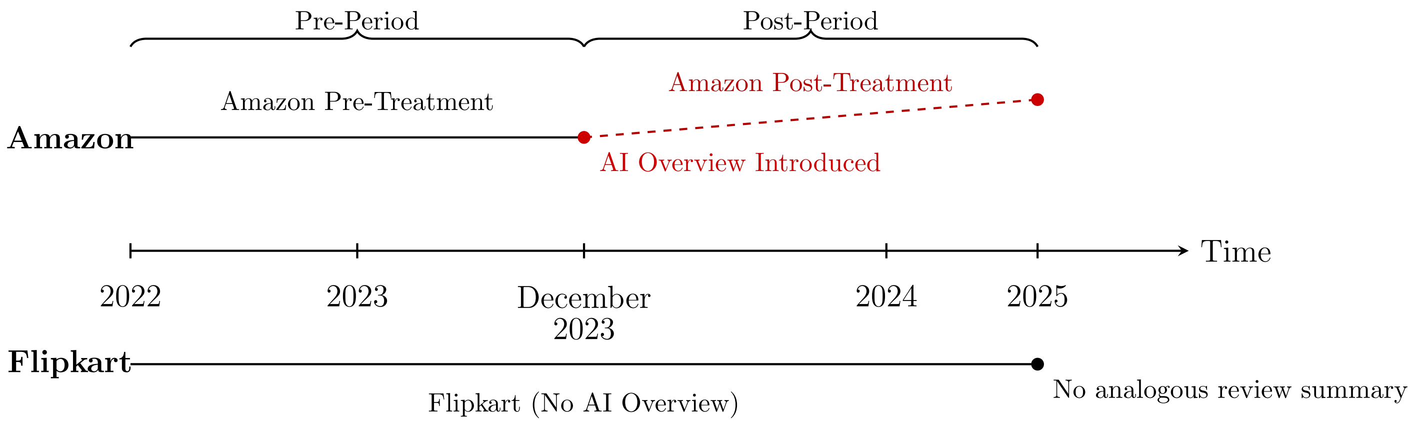


Figure 2: Timeline of AI Overview Rollout on Amazon and Flipkart. The upward slope of the Amazon post-treatment line is purely representational.

struction has two stages. The first identifies product intersections from Amazon.in and Flipkart.com search results and product-detail pages using ASIN–PID links, brand names, titles, URLs, categories, and product attributes. The second assembles monthly historical prices for each matched product–platform cell.

Amazon prices are taken from Keepa daily price histories for Amazon India and aggregated to monthly medians after unit normalization[5]. Flipkart prices require multi-source reconstruction because no comparable commercial longitudinal data was available. The Flipkart side combines five collection channels: direct Flipkart.com product-page and search-page scrapes for identifiers and match metadata; open third-party price-history trackers, primarily PriceHistoryApp and SpendMitra; attempted BuyHatke PID-level queries, which are retained in the collection logs but did not contribute usable explicit point histories to the final panel; Internet Archive/Wayback Machine product-page snapshots with Memento used as an archive fallback; and a small set of archived JSON backfills. The union scripts are also configured to ingest other open price-tracker exports such as Flipshope and TrackProduct histories where available. The final panel retains source flags, proof URLs, collection timestamps, and proof notes, so each observation is traceable to its collection channel. Because sales rank and quantity outcomes are not consistently available across both platforms, price

[5] Keepa reports prices in paise; I normalize to rupees for the analysis.

is the primary outcome.

5.1 Brand Intersections

The matching pipeline begins with five broad product umbrellas with expected cross-platform overlap: electronics, fashion, home, toys, and office products. Within each umbrella, category-specific queries are issued on both platforms. Product titles and identifiers are parsed into brand and product-attribute fields, generating separate Amazon and Flipkart pre-match datasets.

The matching stage is deliberately conservative. Candidate matches must agree on brand and hard product attributes when those attributes are available. For example, a phone is matched only to the same model, color, memory, and storage configuration. When titles do not contain enough information, the pipeline opens the product detail page to recover missing attributes before scoring the candidate pair. Brand and attribute blocking are combined with fuzzy title similarity to generate candidate intersections. The objective is to minimize false positives, even at the cost of false negatives, because a small number of invalid matches can mechanically generate spurious cross-platform price gaps. Candidate matches were subsequently reviewed and flagged mismatches or apparent SKU changes were removed from the final gold panel.

Product overlap percentages are consistent with expected overlap patterns across categories. The broad scraping and matching process produces a large candidate universe, but the analysis uses only matches that survive identifier checks, title/attribute verification, manual review, and later exclusion of flagged mismatches or apparent SKU changes. The final gold analysis panel contains 1,196 ASIN–PID product pairs. Appendix Tables 15 and 16 report the final sample counts and category composition.

5.2 Historical Price Sources and Sparsity

After matching and price-history construction, the resulting monthly product–platform panel covers January 2019 through February 2026. The final gold analysis sample contains 1,196 ASIN–PID pairs. Amazon contributes 28,656 matched observations and Flipkart contributes 17,996 matched observations. The asymmetry is substantively important: Keepa gives relatively dense Amazon histories, while Flipkart histories are reconstructed from noisier and more intermittent sources. The final sample composition reflects this asymmetry. Amazon observations are entirely sourced from Keepa. Flipkart observations come primarily from the PriceHistoryApp/pretrend union and PriceHistoryApp interpolation files, with additional observations from scraped price-history trackers, Archive.org snapshots, and a small archived-JSON backfill. Appendix Table 17 reports these source shares. Flipkart sparsity is therefore a central identification issue.

In the gold sample (matched product sample), 429 of 1,196 Flipkart series are continuous over their observed span, 516 have span fill rates below 80%, and 196 have span fill rates below 50%. The median Flipkart span fill rate is about 0.855, with a 25th percentile of 0.600 and a 10th percentile of 0.375. Missingness can affect the analysis in three ways: it changes which products identify the pre/post contrast, it can create composition-driven movement in the Amazon–Flipkart gap, and it can induce unusual TWFE residual weights in the full unbalanced panel. For this reason, the analysis distinguishes the full gold panel from a strict-support sample requiring at least six pre-treatment and six post-treatment observations on both platforms. This preferred strict-support sample contains 146 ASINs and 14,167 observations.

6 Empirical Design

The missing object is the price Amazon would have charged absent its AI summaries. Each estimator constructs this counterfactual under different restrictions; this section defines the estimands, states the estimating equations, and gives the identifying assumptions under which each estimate has a causal interpretation.

6.1 Potential Outcomes and the Missing Counterfactual

Let i index matched products, $p \in \{\text{Amazon}, \text{Flipkart}\}$ index platform, and t index month. The outcome is log price, $Y_{ipt} = \log(P_{ipt})$. Following the potential-outcomes framework (Rubin 1974; Holland 1986), define

$$Y_{ipt}(1),\ Y_{ipt}(0),$$

where treatment denotes exposure to Amazon's AI review summaries.

Treatment assignment is platform-time specific:

$$D_{pt} = \mathbf{1}\{p = \text{Amazon}\} \cdot \mathbf{1}\{t \geq T_0\},$$

with $T_0 =$ December 12, 2023. Because the outcome panel is monthly and month labels are stored as first-of-month dates, December 2023 is treated as pre-treatment and January 2024 onward is the effective post period in the monthly regressions.

Observed outcomes satisfy

$$Y_{ipt} = D_{pt} Y_{ipt}(1) + (1 - D_{pt}) Y_{ipt}(0).$$

For treated Amazon post-period observations, $Y_{i,\text{Amazon},t}(0)$ is unobserved. Constructing this missing path is the central empirical task; each estimator below imposes different restrictions on how it is built from observed Amazon and Flipkart prices.

6.2 Support Sets and Estimands

The primitive causal object is a group-time effect:

$$ATT_t(\mathcal{S}_t) = \mathbb{E}[Y_{i,\text{Amazon},t}(1) - Y_{i,\text{Amazon},t}(0) \mid i \in \mathcal{S}_t]\,,$$

where $\mathcal{S}_t$ is the product support contributing to identification at month t. The reported specifications estimate weighted aggregates:

$$\theta(w, \mathcal{S}) = \sum_{t \geq T_0} w_t\, ATT_t(\mathcal{S}_t), \qquad \sum_{t \geq T_0} w_t = 1.$$

The preferred estimand is the average post-rollout effect for products with stable bilateral coverage: matched ASIN–PID pairs observed at least six times before and at least six times after the rollout on *both* Amazon and Flipkart. Other specifications change $\mathcal{S}$ or w to learn how the estimate depends on support, weighting, or the counterfactual model. The four support sets serve distinct economic purposes:

1. **Full gold panel.** All matched ASINs observed on both platforms at any point in the cleaned dataset. This maximizes external validity but is most exposed to Flipkart sparsity and changing composition.

2. **Strict-support sample, $k = 6$.** ASINs with at least six pre- and post-treatment monthly price observations on both platforms. This is the preferred support because it forces the treated and comparison platforms to be observed on both sides of the rollout, which is what makes the within-product cross-platform contrast well-defined. The cleaned sample contains 146 ASINs and 14,167 positive-price observations.

3. **Coverage-weighted support.** The same equation with product-platform weights based on observed panel coverage. Following Solon, Haider, and Wooldridge (2015), this is interpreted as estimand-changing rather than as a missing-data correction: it

answers what the average effect is if better-measured product-platform histories receive more weight.

4. **Category and local-window support.** Category-specific and event-window specifications used to test whether the response appears where the model predicts and whether there is a sharp break at the rollout date.

6.3 Two-Way Fixed Effects DiD

The baseline stacked-panel estimator is

$$Y_{ipt} = \alpha_{ip} + \delta_t + \beta D_{pt} + \Gamma' Z_{ipt} + \varepsilon_{ipt},$$

where α_{ip} denotes product×platform fixed effects, δ_t denotes month fixed effects, and Z_{ipt} denotes time-varying controls used in robustness specifications. Product×platform fixed effects absorb permanent Amazon–Flipkart differences for the same ASIN, including platform-level price differences, seller-mix differences, and any time-invariant matching artifact. Month fixed effects absorb shocks common to both platforms—inflation, exchange-rate movements, broad seasonality. The coefficient β therefore captures within-product differential movement in Amazon prices relative to Flipkart prices after the rollout. The main coefficient table reports ASIN-clustered standard errors; the inference table reports ASIN-platform and two-way ASIN-by-month clustering as more conservative dependence structures (Cameron, Gelbach, and Miller 2011).

Because treatment is assigned at the platform level on a single date, the staggered-timing TWFE pathologies studied by Goodman-Bacon (2021), Sun and Abraham (2021), and de Chaisemartin and D'Haultfœuille (2020) do not apply. The weight diagnostic addresses a different concern: unbalanced Flipkart support. In multi-period settings TWFE recovers an implicit weighted average of group-time effects,

$$\operatorname{plim} \hat{\beta}^{TWFE} = \sum_{t \geq T_0} \omega_t^{TWFE} ATT_t,$$

with weights determined by the residualized treatment indicator. I report observation-level treatment-weight diagnostics constructed from the Frisch–Waugh–Lovell residualized treatment variable. The full panel exhibits negative treated weights under unbalanced support; the strict-support restriction eliminates this problem, making the implied $\hat{\beta}^{TWFE}$ an interpretable convex combination of underlying group-time effects.

For β to have a causal interpretation, the key condition is conditional parallel trends in untreated potential outcomes:

$$\mathbb{E}[Y_{i,\text{Amazon},t}(0) - Y_{i,\text{Amazon},s}(0) - \{Y_{i,\text{Flipkart},t}(0) - Y_{i,\text{Flipkart},s}(0)\} \mid i \in \mathcal{S}] = 0$$

for post-treatment months t and pre-treatment reference months s, conditional on product×platform fixed effects, month fixed effects, and the relevant support $\mathcal{S}$. Intuitively, absent AI summaries, the within-product Amazon–Flipkart log price gap would have evolved like its untreated comparison path. This condition is more demanding in the full unbalanced panel, where Flipkart coverage changes over time, than in the strict-support sample, where the same matched products are observed on both platforms before and after the rollout.

6.4 Support, Weighting, and Paired-Ratio Specifications

The strict-support estimator uses the same TWFE equation but changes $\mathcal{S}$. The gain is cleaner support: the same matched products must be observed on both platforms before and after treatment. The cost is a smaller and less externally representative sample.

The coverage-weighted estimator keeps the TWFE structure but changes the aggregation weights. Let f_{ip} denote the within-span fill rate of product i on platform p and let c_{ip} denote coverage over the global monthly panel. The implemented weights are proportional to $f_{ip}c_{ip}$ and normalized to have mean one in the estimation sample. Following Solon, Haider, and Wooldridge (2015), these weights are used as an estimand-changing robustness diagnostic, not as a cure for missing data.

The analysis also estimates a direct paired-ratio specification:

$$R_{it} = \log(P_{i,\text{Amazon},t}) - \log(P_{i,\text{Flipkart},t}), \qquad R_{it} = \eta_i + \rho Post_t + u_{it}.$$

This is the most direct cross-platform object because the outcome is exactly the Amazon–Flipkart price gap for the same ASIN-month. Because this requirement discards many sparse Flipkart months, the paired-ratio estimates have much lower power and are used as diagnostics.

6.5 Dynamic, Category, and Local-Window Specifications

Event studies replace D_{pt} with relative-month interactions:

$$Y_{ipt} = \alpha_{ip} + \delta_t + \sum_{\ell \neq -1} \beta_\ell \mathbf{1}\{t - T_0 = \ell\}\mathbf{1}\{p = \text{Amazon}\} + \varepsilon_{ipt}.$$

Pre-treatment coefficients test whether the Amazon–Flipkart gap was already moving differentially before the rollout. These are diagnostic tests, not separate causal estimates. Category-specific DiD estimates apply the same logic within mechanism-relevant product groups. Local-window DiD restricts the event window around treatment and is designed to detect immediate breaks. In this data the local-window estimates are small and insignificant, so the timing evidence is interpreted as gradual rather than as a sharp one-month discontinuity.

6.6 Imputation, Matrix Completion, and Interactive Fixed Effects

TWFE assumes untreated outcomes are additively separable into product-platform and month components. The analysis also implements the imputation estimator of Borusyak, Jaravel, and Spiess (2024). The imputation approach estimates an untreated-outcome model using untreated observations, predicts $\widehat{Y}_{ipt}(0)$ for treated Amazon post-period cells, and averages the observed-minus-imputed gaps. In this application the first-stage untreated-outcome model includes product-platform and month fixed effects, so the identifying content is close to TWFE but the estimator avoids using already-treated observations to fit the untreated potential-outcome surface.

Matrix completion and interactive fixed effects relax additive separability by allowing unobserved latent factors:

$$Y_{ipt}(0) = \alpha_{ip} + \delta_t + \lambda_{ip}' f_t + u_{ipt}.$$

The estimator uses untreated observations to recover the low-rank component $\lambda_{ip}' f_t$ and then predicts the missing counterfactual $\widehat{Y}_{i,\text{Amazon},t}(0)$ for treated Amazon post-period cells (Bai 2009; Athey et al. 2021; Liu, Wang, and Xu 2022). The corresponding ATT-type estimand is

$$\theta^{MC} = \frac{1}{|\mathcal{T}|} \sum_{(i,t)\in\mathcal{T}} \left(Y_{i,\text{Amazon},t} - \widehat{Y}_{i,\text{Amazon},t}(0) \right).$$

These estimators change the counterfactual model rather than the underlying research design. BJS imputation is consistent if the untreated-outcome model fitted on never-treated and pre-treatment observations extrapolates correctly to treated Amazon post-period cells. Matrix completion and interactive fixed effects are consistent if untreated prices admit a stable low-rank factor structure conditional on fixed effects and missingness preserves enough information to recover the latent factors. These restrictions accommodate richer latent demand, category, or platform shocks than additive parallel trends. Crucially, however, they do not solve the pre-trend problem: a low-rank counterfactual model can still inherit the same pre-treatment movement in the Amazon–Flipkart gap that drives concern about TWFE. Agreement across estimators is therefore evidence of robustness to the functional form of the counterfactual, not evidence that pre-trend violations are immaterial.

6.7 Treatment Timing

The monthly panel codes January 2024 as the first effective post-treatment month. Identification requires no anticipation: pre-treatment Amazon prices were not already responding to the upcoming feature in a way that fixed effects do not absorb. This condition can fail if Amazon piloted summaries before the public announcement or if exposure differed across mobile and desktop. Such timing uncertainty rules out a strict event-date interpretation; I therefore report treatment-timing variants, donut specifications, and local-window estimates.

6.8 Cross-Platform Interference

The cross-platform design is also vulnerable to interference. Many Indian e-commerce sellers list on both Amazon and Flipkart, and some operate common repricing systems. If Amazon's AI summaries shift Amazon-side demand and a common seller adjusts Flipkart prices in response, Flipkart is partially treated. Such spillovers attenuate the relative estimate if both platforms move in the same direction, but could amplify it under differential repricing; the estimand shifts in either case. The coefficient is therefore properly read as a cross-platform relative equilibrium price response, not as a SUTVA-clean direct effect. The available data do not consistently identify Flipkart sellers, so common-seller overlap cannot be measured directly.

6.9 Identification Diagnostics

I assess the credibility of the identifying conditions through three diagnostics. First, event studies and binned pre-trend tests examine whether the Amazon–Flipkart gap was already moving differentially before the rollout. Second, the partial-identification framework of Rambachan and Roth (2023) characterizes how conclusions change as one allows bounded violations of parallel trends, using both relative-magnitude restrictions (post-treatment violations bounded by a multiple $\bar{M}$ of pre-period violations) and smoothness restrictions (bounded rate of change). Third, randomization-based sign-flip inference in the strict-support sample tests whether the observed Amazon-relative decline is unusually large under random platform relabeling within matched ASINs (Fisher 1935; Rosenbaum 2002).

7 Results

7.1 Main Estimates

Table 1 reports the main estimates. In the preferred strict-support sample (146 ASINs, 14,167 observations), the estimate is $\hat{\beta} = -0.025$: Amazon prices fell by 2.5 log points relative to Flipkart for the same product after the rollout. With ASIN-clustered standard errors the p-value is 0.044; under ASIN-by-month two-way clustering (Table 3), it is 0.069.

The estimate is similar in the broader matched panel. The full gold panel produces a 2.4 log-point decline ($\hat{\beta} = -0.024$, $p = 0.044$), which rules out the concern that the strict-support estimate is an artifact of the support restriction. The full panel uses essentially every matched ASIN but is also the most exposed to Flipkart sparsity and changing composition, which is why I treat it as supporting evidence rather than as the main finding.

The strict-support estimate is not a clean event-study result on its own. The Rambachan–Roth relative-magnitude confidence interval for the strict-support average-post target includes zero even at the no-deviation benchmark. What gives the design its identifying content is two things: the price movement is similar across TWFE, BJS imputation, matrix completion, and interactive fixed effects (Table 13), and the cross-sectional heterogeneity matches what the model predicts ex ante. Neither of these depends on a single event-study coefficient.

The remaining columns of Table 1 show how the estimate responds to changes in support structure. Entry-cohort controls, outlier-cell removal, sale-calendar controls, and platform-by-category trends preserve sign and broad magnitude. Coverage weighting moves the estimate toward zero and eliminates conventional significance: this is a real qualification, and I interpret the defensible magnitude as a 2–3 log-point decline rather than a precisely identified scalar. Local-window estimates around the rollout date are near zero, consistent with a gradual rather than discontinuous adjustment.

Table 1: Main Estimates: Amazon Prices Relative to Flipkart After AI Review Summaries

Specification	Support	$\hat{\beta}$	% effect	SE	p-value	ASINs
Full gold TWFE	All matched products	-0.024	-2.37	0.012	0.044	1,196
Strict-support TWFE	$k = 6$ both platforms	-0.025	-2.48	0.012	0.044	146
Entry-cohort controls	$k = 6$ both platforms	-0.025	-2.45	0.013	0.050	146
Outlier-cell cleaned	$k = 6$ both platforms	-0.025	-2.48	0.013	0.048	146
Historical sale controls	All matched products	-0.024	-2.38	0.012	0.043	1,196
Platform×category trends	Rich category support	-0.028	-2.72	0.011	0.013	1,185
Coverage-weighted TWFE	$k = 6$ both platforms	-0.021	-2.09	0.014	0.128	146
Local window $[-12, +18]$	Event-window support	0.004	0.42	0.010	0.680	1,041

Notes: Outcome is log price. Treatment is Amazon×post after Amazon India's AI review-summary rollout. Product×platform and month fixed effects are included unless otherwise stated. The strict-support sample $k = 6$ contains products with at least six pre-treatment and six post-treatment observations on both platforms. Percent effects are $100(\exp(\hat{\beta}) - 1)$. Standard errors are clustered by ASIN. Historical sale controls include platform-specific indicators for Amazon Prime Day, Amazon Great Indian Festival, and Flipkart Big Billion Days months from 2019–2025; they do not exhaust every promotional event in Indian e-commerce. The local-window row is included to show that the effect is not an immediate one-month break around treatment.

The treatment-date convention does not drive the main coefficient. Table 2 reports three monthly codings of the announcement month: December 2023 as pre-treatment, December 2023 as treated, and December 2023 dropped. The full-panel and strict-support estimates lie between -0.024 and -0.026 across these choices. Stability matters here because the announcement was on December 12 while the panel is monthly, and product-level exposure may have differed across mobile, desktop, or pilot surfaces.

Table 3 separates magnitude from inference. The point estimate is unchanged across clustering choices, but the standard error grows under more permissive dependence structures. In the strict-support sample, ASIN-platform clustering preserves the coefficient with a

Table 2: Treatment Timing Variants Around December 2023

Treatment timing	Sample	$\hat{\beta}$	% effect	SE	p-value	Obs.	ASINs
Dec. pre / Jan. post	Full gold	-0.024	-2.37	0.012	0.044	46,504	1,196
Dec. treated	Full gold	-0.025	-2.43	0.012	0.042	46,504	1,196
Dec. dropped	Full gold	-0.025	-2.43	0.012	0.043	45,942	1,196
Dec. pre / Jan. post	Strict support $k = 6$	-0.025	-2.48	0.012	0.044	14,167	146
Dec. treated	Strict support $k = 6$	-0.026	-2.52	0.012	0.040	14,167	146
Dec. dropped	Strict support $k = 6$	-0.026	-2.54	0.012	0.041	13,972	146

Notes: Outcome is log price. All specifications include product-by-platform and month fixed effects. Standard errors are clustered by ASIN. The preferred monthly convention treats December 2023 as pre-treatment because the public rollout date was December 12 and monthly dates are stored as first-of-month dates.

larger standard error, and two-way ASIN-by-month clustering following Cameron, Gelbach, and Miller (2011) produces a p-value of 0.069. The estimated magnitude is therefore stable, while conventional 5% significance is not robust to every plausible residual-dependence assumption. I read the estimate as marginally significant at conventional levels and clearly distinguishable from zero in magnitude.

Table 3: Inference Sensitivity for the Stacked TWFE Estimate

Support	Clustering	$\hat{\beta}$	SE	p-value	Observations
Full gold	ASIN	-0.024	0.012	0.044	46,504
Full gold	ASIN-platform	-0.024	0.014	0.087	46,504
Full gold	ASIN and month	-0.024	0.014	0.083	46,504
Strict support $k = 6$	ASIN	-0.025	0.012	0.044	14,167
Strict support $k = 6$	ASIN-platform	-0.025	0.019	0.182	14,167
Strict support $k = 6$	ASIN and month	-0.025	0.014	0.069	14,167

Notes: The coefficient is mechanically unchanged across rows; only the variance estimator changes. ASIN-platform clustering is more conservative because it allows residual dependence to differ across the Amazon and Flipkart series for a matched product. Two-way clustering follows Cameron, Gelbach, and Miller (2011). The table shows that the magnitude is stable, while conventional 5% significance is sensitive to conservative dependence assumptions.

As a direct gap diagnostic, I also estimate the within-product paired log-price ratio, $R_{it} = \log P_{it}^{Amazon} - \log P_{it}^{Flipkart}$, using only ASIN-months observed simultaneously on both platforms. This object visualizes the cross-platform gap most directly, but the same-month observability requirement discards a substantial share of the panel, leaving the strict-support paired-ratio estimate underpowered (-0.017, $p = 0.58$). I therefore treat it as a diagnostic, not the main estimator to interpret and conclude from. Figure 3 plots the monthly median paired log-ratio with bootstrap bands. The figure documents the direction of the cross-platform gap over time, but early months rest on few same-month pairs and should not be read as a substitute for the stacked-panel estimates.

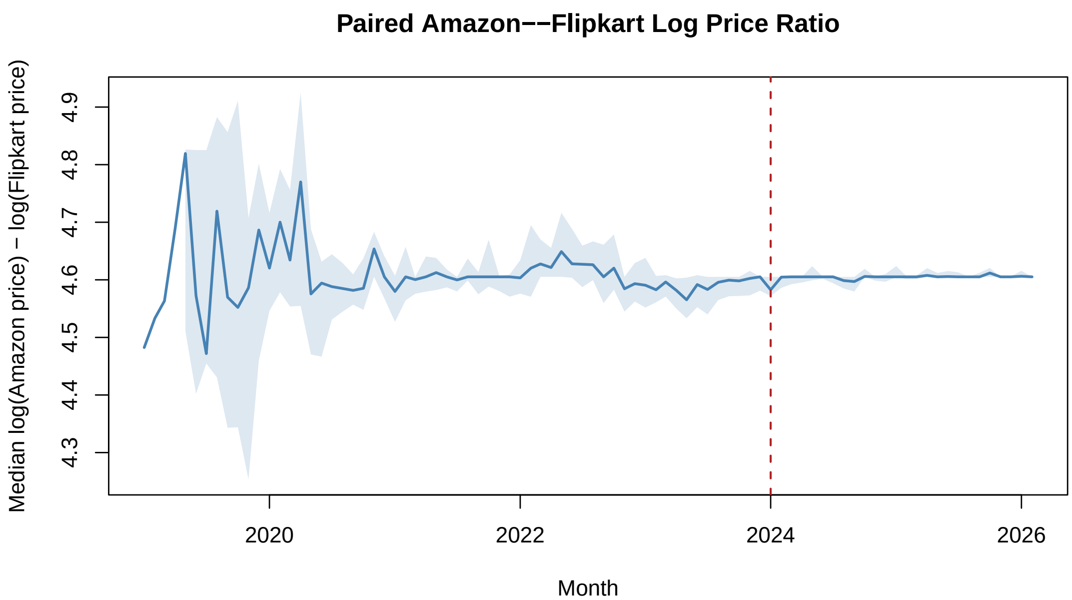


Figure 3: Monthly paired Amazon–Flipkart log price ratio

7.2 Dynamic Patterns and Diagnostics

The event-study evidence informs but does not by itself identify the aggregate causal estimand. Figure 5 shows differential pre-treatment movement in the Amazon–Flipkart gap, particularly in the far pre-period. In the strict-support binned pre-trend test, the far-pre bin is positive (0.080) and the mid-pre bin is close to zero (0.007); the joint test rejects flat pre-trends ($p < 0.001$). This failure is unsurprising given non-stationarity in Indian e-commerce over 2019–2026 and unevenly expanding Flipkart coverage that changes the contributing product mix over time.

The remaining diagnostics quantify but do not eliminate this concern. Platform-by-category trend controls produce similar or slightly larger negative estimates ($\hat{\beta} = -0.028$

unweighted, −0.026 weighted), indicating that the result is not absorbed by flexible category-specific platform trends. Local-window estimates around treatment are small and statistically insignificant, weakening a sharp event-date interpretation. Donut specifications that drop the immediate treatment-window months remain close to the strict-support estimate, consistent with gradual price adjustment.

This pattern reflects a genuine support–validity tradeoff. Starting the panel after the most disruptive COVID-era logistics period, for example in late 2021 or early 2022, leaves the strict-support estimate negative and economically similar in magnitude to the baseline (−0.019 to −0.024, compared with −0.025 in the full strict-support window). This stability is informative: the headline magnitude is not mechanically driven by the earliest COVID-period observations. However, the available pre-period in these medium windows still rejects flat pre-trends. Starting the window late enough for formal pre-trend tests to pass, around late 2022 or early 2023, sharply reduces the identifying support to roughly 80–100 strict-support ASINs and produces small, imprecise coefficients that are near zero or positive. I therefore do not interpret the clean short-window estimates as evidence against any price response. They identify a narrow near-rollout discontinuity with limited support, whereas the medium-window estimates identify a more gradual relative price adjustment but require stronger counterfactual assumptions. The empirical design is consequently informative about a medium-run cross-platform price movement, not a sharp one-month event-study break.

A slope-change diagnostic gives the same message in a parametric form. I augment the strict-support specification with an Amazon-specific linear trend and an Amazon-specific post-rollout trend shift, centering time at January 2024. This separates the rollout level shift from pre-existing Amazon-relative convergence. In the strict-support sample, the implied level shift is about −0.041 to −0.042 ($p < 0.02$), while the post-rollout slope change is positive, about 0.0028 to 0.0031 log points per month. Thus the average post-period effect remains negative, but the dynamics are not monotone: the data are consistent with an initial Amazon-relative price decline followed by partial upward drift or mean reversion.

This decomposition also explains why the main average estimate is smaller than the initial level shift. In the strict-support slope-change model, the month-0 level shift is roughly -4.1 log points. The positive post-rollout slope implies that by month 12 the implied effect has attenuated to approximately -0.4 to -0.8 log points, while the average over months 0–12 is approximately -2.3 to -2.5 log points. This arithmetic maps closely to the headline strict-support estimate. I therefore interpret the main coefficient as an average medium-run relative price response with dynamics, rather than as a permanent one-time compression in the Amazon–Flipkart gap.

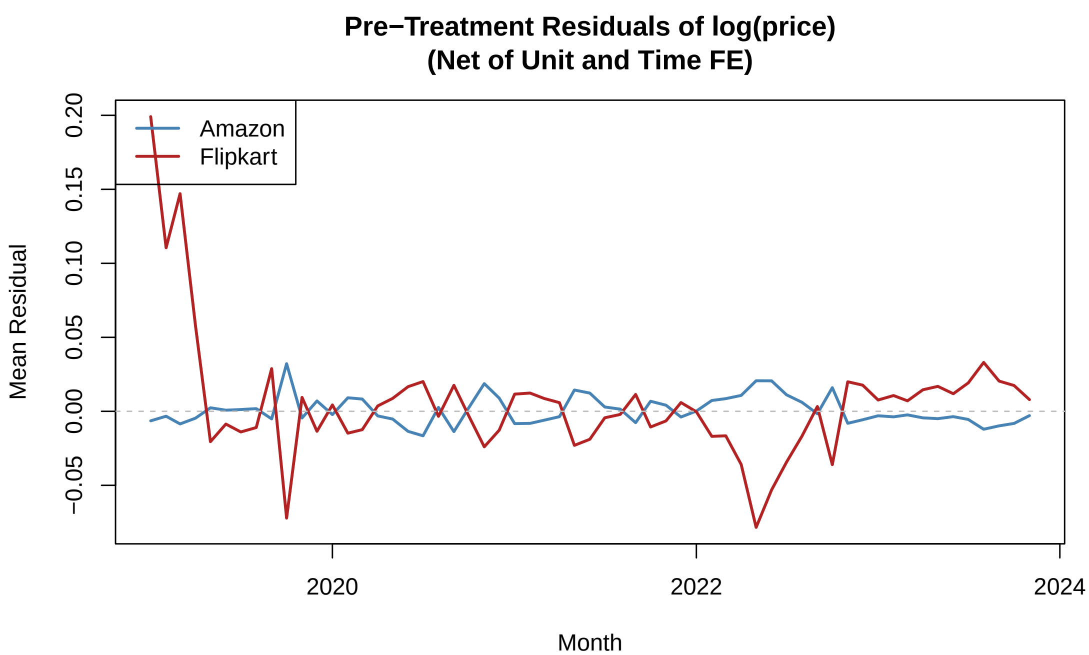


Figure 4: Pre-treatment residuals

Figures 7–8 and Tables 4–5 report Rambachan and Roth (2023) sensitivity results, implemented with the `HonestDiD` package. The two restrictions identify different objects. Under the relative-magnitude restriction, $\bar{M} = 0$ corresponds to exact parallel trends; positive $\bar{M}$ allows post-treatment violations bounded by a multiple of pre-period violations. Under the smoothness restriction, $M = 0$ instead permits a linear continuation of any pre-period differential trend, so $M = 0$ is *not* parallel trends. With January 2024 aligned as event month 0,

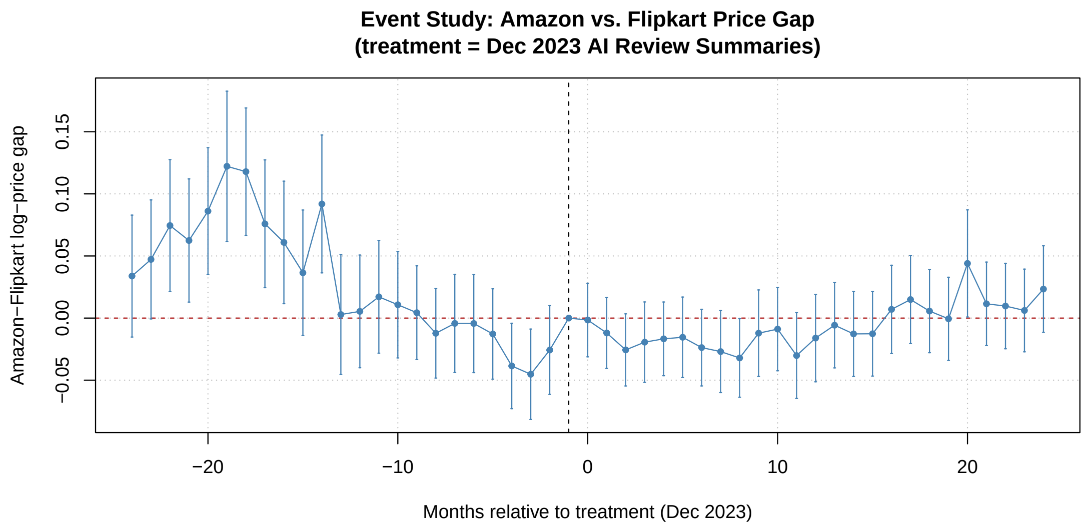


Figure 5: Event study around Amazon India's AI review-summary rollout

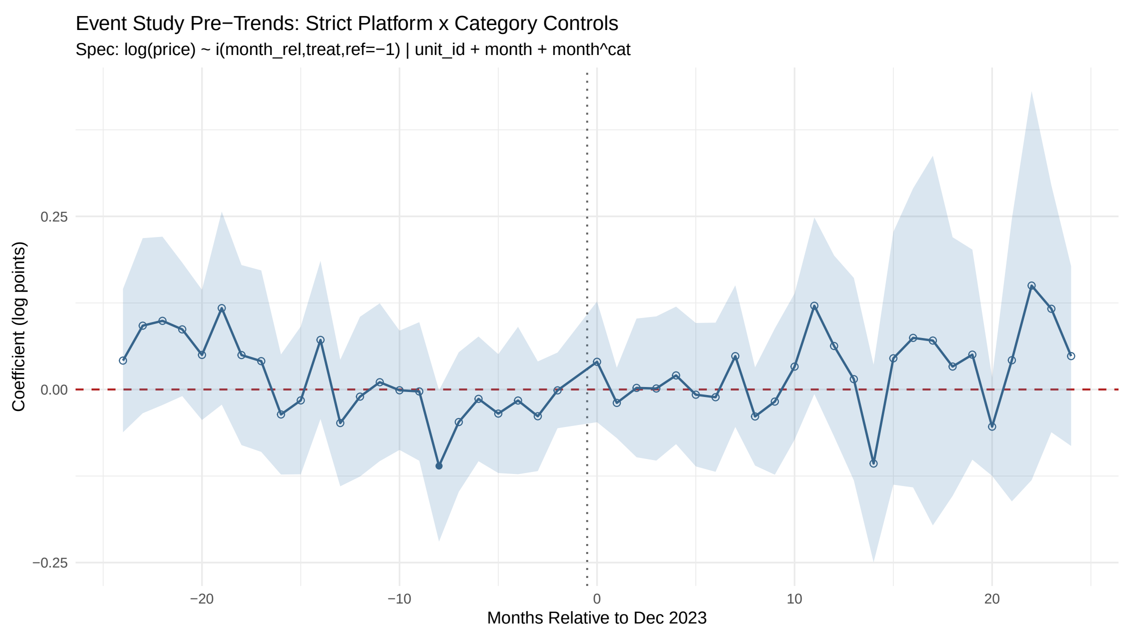


Figure 6: Event-study diagnostics with platform-by-category trend structure

the strict-support average-post confidence set includes zero even at $\bar{M} = 0$ ($[-0.037, 0.017]$), which is what the local-window evidence already showed: there is no sharp immediate break. The Toys & Games sample is different. For the average effect over event months 0–12, the relative-magnitude confidence set excludes zero at $\bar{M} = 0$ ($[-0.197, -0.043]$) and includes zero once $\bar{M}$ reaches 0.25. The smoothness confidence sets include zero for both samples, by construction: they allow the observed pre-period differential trend to continue linearly into the post-period. The sensitivity analysis therefore does not overturn the pre-trend concern but locates it: the Toys & Games result holds under parallel-trends-based identification but is not robust to moderate extrapolated violations of that assumption.

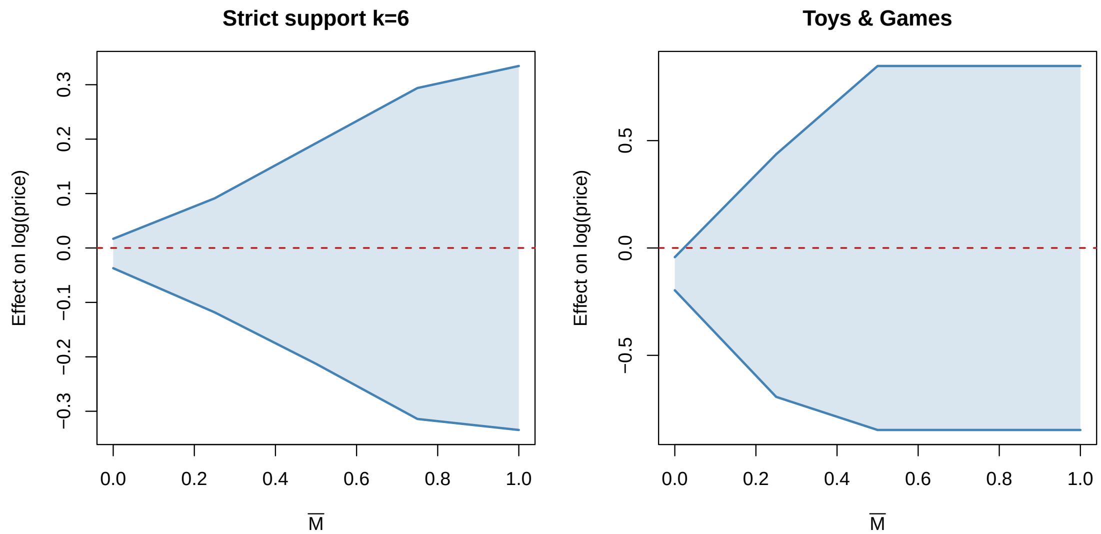


Figure 7: Rambachan–Roth relative-magnitude sensitivity in strict-support and Toys & Games samples

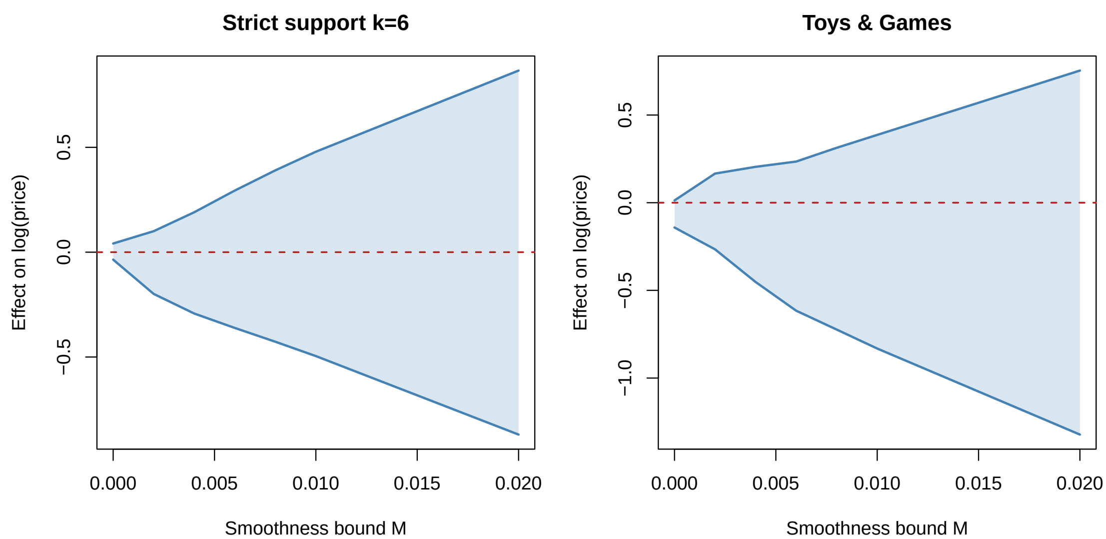


Figure 8: Rambachan–Roth smoothness sensitivity in strict-support and Toys & Games samples

Table 4: Rambachan–Roth Relative-Magnitude Sensitivity Across Samples

Sample	$\bar{M}$	Lower CI	Upper CI	Includes zero
Strict support k=6	0.00	-0.037	0.017	Yes
Strict support k=6	0.25	-0.118	0.091	Yes
Strict support k=6	0.50	-0.213	0.193	Yes
Strict support k=6	0.75	-0.314	0.294	Yes
Strict support k=6	1.00	-0.334	0.334	Yes
Toys & Games	0.00	-0.197	-0.043	No
Toys & Games	0.25	-0.693	0.437	Yes
Toys & Games	0.50	-0.847	0.847	Yes
Toys & Games	0.75	-0.847	0.847	Yes
Toys & Games	1.00	-0.847	0.847	Yes

The table reports 95% Rambachan–Roth confidence sets, computed with the `HonestDiD` package, under the relative-magnitude restriction for the average post-treatment effect over event months 0–12. $\bar{M} = 0$ corresponds to no deviation from parallel trends; larger values allow post-treatment violations proportional to the largest observed pre-period violation. The inversion uses 100 grid points for computational tractability.

Table 5: Rambachan–Roth Smoothness Sensitivity Across Samples

Sample	M	Lower CI	Upper CI	Includes zero
Strict support k=6	0.000	-0.036	0.041	Yes
Strict support k=6	0.002	-0.199	0.100	Yes
Strict support k=6	0.004	-0.293	0.190	Yes
Strict support k=6	0.006	-0.361	0.294	Yes
Strict support k=6	0.008	-0.427	0.390	Yes
Strict support k=6	0.010	-0.496	0.479	Yes
Strict support k=6	0.020	-0.870	0.866	Yes
Toys & Games	0.000	-0.141	0.013	Yes
Toys & Games	0.002	-0.266	0.166	Yes
Toys & Games	0.004	-0.452	0.205	Yes
Toys & Games	0.006	-0.616	0.235	Yes
Toys & Games	0.008	-0.723	0.313	Yes
Toys & Games	0.010	-0.832	0.387	Yes
Toys & Games	0.020	-1.322	0.753	Yes

The table reports 95% Rambachan–Roth confidence sets, computed with the `HonestDiD` package, under the smoothness restriction for the average post-treatment effect over event months 0–12. Event month 0 is January 2024; December 2023 is the omitted reference month. Under the smoothness restriction, $M = 0$ permits a linear continuation of any pre-period differential trend; it is not the same as imposing exact parallel trends.

7.3 Competitive Entry as an Alternative Explanation

The main alternative economic explanation is that Amazon–Flipkart price compression reflects competitive entry in Indian e-commerce rather than AI review summaries. To examine this channel directly, I construct a source-backed event file of major vertical-entry and scaling events over the sample period: Nykaa's public-market scale milestone in beauty, Ajio/Reliance Retail scaling in apparel, Zepto and Blinkit in quick commerce, Tata Neu's super-app launch, Meesho Superstore, and JioMart's WhatsApp shopping integration. Each event is assigned to exposed product categories using the event's stated business scope and the product taxonomy in the matched panel. I then construct, for each category-month, the number of documented entrant events active by that month and interact this exposure with the Amazon indicator.

Table 6 reports the diagnostic. The source-backed entry controls do not make the far-pre-period trend disappear. In the strict-support sample, adding Amazon $\times$ active entrant-event count reduces some early imbalance in simple unit-month specifications, but the far-pre pretrend still rejects; with month $\times$ category fixed effects, the control has little additional explanatory power. This means documented vertical entry is part of the competitive environment, but it is not a sufficient explanation for the full pre-period movement. The window-start frontier gives the same message: starting in 2021 and controlling for entry leaves negative medium-window estimates but still rejects flat pretrends; starting late enough for pretrends to pass produces small, imprecise near-rollout estimates because support falls sharply.

The entry-exposure split is more informative for mechanism. If vertical entry alone drove the result, the largest Amazon-relative declines should appear in categories with high documented entrant exposure. Instead, low-entry categories in the source-backed event file produce a strict-support coefficient of about -0.031 ($p = 0.035$). Toys & Games is central to this group: in the documented entrant-event file, it has no major vertical-entry shock analogous to Nykaa in beauty, Ajio in fashion, or quick-commerce/JioMart in household

Table 6: Source-Backed Competitive Entry Diagnostics

Panel A. DiD with source-backed entrant exposure count, month × category FE					
Sample	$\hat{\beta}$	Percent	SE	p-value	ASINs
full gold	-0.020	-2.002	0.019	0.297	1,196
strict k6	-0.021	-2.071	0.027	0.434	146
Panel B. Strict-support estimates by documented-entry exposure tier					
Entry tier	$\hat{\beta}$	Percent	SE	p-value	ASINs
High documented entry	-0.027	-2.689	0.023	0.243	54
Medium documented entry	0.070	7.226	0.040	0.114	11
Low documented entry	-0.031	-3.015	0.014	0.035	81
Panel C. Window-start frontier with entry controls, strict support, month × category FE					
Start month	$\hat{\beta}$	Percent	SE	Pretrend p	ASINs
2021-01-01	-0.012	-1.173	0.028	<0.001	140
2021-07-01	-0.006	-0.553	0.028	<0.001	136
2022-12-01	-0.005	-0.497	0.035	0.255	80
2023-01-01	-0.002	-0.225	0.034	0.293	79

Notes: Source-backed entrant exposure is constructed from documented Indian e-commerce entrant events listed in the external event file. Panel A reports the Amazon-post coefficient after controlling for Amazon × active entrant-event count. Panel B splits strict-support ASINs by the number of documented entrant events matched to their category. Panel C shows that entry controls and later starts improve pretrend diagnostics only once the window becomes short and support falls.

essentials. Relative to those exposed categories, Toys & Games is therefore closer to a direct Amazon–Flipkart comparison with fewer observed third-platform shocks. This makes the Toys & Games result a cleaner test of the review-information channel, while still stopping short of claiming that the category is literally isolated from all competitive pressure.

7.4 Mechanism Diagnostics

The category-level evidence matches the model's prediction that $|\partial P/\partial \Delta|$ is larger where review information is more valuable to compress. Four broad categories pass the category pre-trend screen: Toys & Games, Electronics, Beauty, and Outdoor Living. The cleanest point estimate is in Toys & Games, where the category-specific DiD coefficient is -0.086 ($p = 0.020$). A leave-one-ASIN-out diagnostic produces estimates between -0.100 and -0.072, ruling out a single-product driver. This is also a category in which review narratives carry decision-relevant information about safety, age appropriateness, durability, and parent experiences. The Rambachan–Roth sensitivity locates the strength of this result: Toys & Games clears the relative-magnitude no-deviation benchmark but does not survive moderate extrapolated violations. Electronics and Beauty are directionally negative but underpowered; Outdoor Living clears the pre-trend screen but does not produce a meaningful negative estimate.

Table 7: Category-level pre-trend screen

cat_top	n_asins	n_obs	pre_joint_p_unweighted	pre_joint_p_weighted	pass_pretrend_screen
Office Products	304	10751	8.740e-09	3.380e-09	FALSE
Home & Kitchen	202	6920	0.0432	0.0499	FALSE
Computers & Accessories	39	1033	0	0	FALSE
Toys & Games	224	6603	0.5728	0.5232	TRUE
Electronics	123	3067	0.2638	0.2601	TRUE
Beauty	101	4163	0.6418	0.4192	TRUE
Outdoor Living	33	745	0.4737	0.5155	TRUE

The sharper mechanism test uses pre-treatment review-corpus size as a continuous proxy for R_i in the framework above. Table 9 interacts the Amazon-post indicator with the standardized log of the product's pre-treatment Amazon review count. In the strict-support

Table 8: Timing and Category Diagnostics

Diagnostic	Estimate	SE	p-value	ASINs
Donut: drop months 0–1, post starts month 2	-0.025	0.013	0.044	146
Donut: drop months 0–2, post starts month 3	-0.025	0.013	0.051	146
Donut: drop months -1–2, post starts month 3	-0.026	0.013	0.046	146
Toys & Games baseline	-0.091	0.034	0.008	225
Toys & Games leave-one-out range	[-0.100, -0.072]	—	—	224

Notes: Donut specifications use the strict-support sample and remove immediate treatment-window months before re-estimating the stacked TWFE coefficient. The Toys & Games leave-one-out row reports the minimum and maximum coefficient after dropping one ASIN at a time. These diagnostics support a gradual and category-specific pattern but do not separately identify consumer learning from seller or algorithmic price adjustment.

sample, the interaction is -0.025 (SE 0.014, $p = 0.064$): a one-standard-deviation larger pre-treatment review-count proxy is associated with an additional 2.5 log-point Amazon-relative price decline. This is the sign predicted by the model—summaries are more valuable when there is more review information to compress—and it directly tests the cross-sectional implication $\partial^2\epsilon_i/\partial\Delta_i\partial R_i > 0$. The full-panel interaction is also negative but less precise. I read this result as mechanism evidence, not as an additional source of causal identification.

Table 9: Review-Corpus Heterogeneity

Sample	Coefficient	$\hat{\beta}$	% effect	SE	p-value	Obs.	ASINs
Full gold	Post	-0.016	-1.57	0.013	0.213	29,058	445
Full gold	Post × review intensity	-0.014		0.009	0.131	29,058	445
Strict support $k = 6$	Post	-0.007	-0.67	0.016	0.671	12,955	134
Strict support $k = 6$	Post × review intensity	-0.025		0.014	0.064	12,955	134

Notes: Outcome is log price. Review intensity is the ASIN-level standardized log of the pre-treatment mean Amazon review count, $\log(1+\overline{reviews}_{i,A,pre})$. Specifications include product-by-platform and month fixed effects; standard errors are clustered by ASIN. A negative interaction means the relative Amazon price decline is larger for products with richer pre-treatment review corpora.

A natural additional implication of search-cost models is price-dispersion compression. I test this directly using ASIN-months with same-month prices on both platforms, measuring dispersion as the absolute deviation of the Amazon–Flipkart log price gap from that month's cross-sectional median. Table 10 does not show dispersion compression. In the strict-support paired sample, post-rollout dispersion rises modestly across specifications. This does not contradict the negative mean effect. The mean and dispersion of the cross-platform gap are distinct moments: the average gap can fall while treatment effects become more heterogeneous across products. A descriptive decomposition by pre-treatment review-corpus quartile is consistent with this interpretation. In the highest review-corpus quartile, the average post-minus-pre gap change is the most negative (-0.036 log points), and within-month gap dispersion rises the most (about 0.047 log points), while lower-review quartiles move less. These quartile patterns are not precise enough to constitute separate identification, but they align with the continuous interaction in Table 9: products with richer review corpora experience larger Amazon-relative price declines. The dispersion result therefore qualifies the homogeneous-compression version of the mechanism, not the heterogeneous treatment-effect version.

Table 10: Cross-Platform Price Dispersion Diagnostics

Specification	$\hat{\beta}_{Post}$	SE	p-value	Pre mean
Abs. deviation from monthly median, ASIN FE + month-of-year FE	0.020	0.009	0.029	0.192
Abs. deviation from monthly median, plus trend	0.025	0.010	0.017	0.192
Squared deviation from monthly median, ASIN FE + month-of-year FE	0.023	0.009	0.017	0.097

Notes: The table uses ASIN-months with same-month Amazon and Flipkart prices in the strict-support sample. Disp measured as deviation of the paired Amazon–Flipkart log price gap from that month's cross-sectional median. Negative co indicate compression in cross-platform price dispersion after the rollout. Standard errors are clustered by ASIN.

As a demand-side diagnostic, in the 22 ASINs with usable Keepa Best Sellers Rank (BSR) histories (Chevalier and Goolsbee 2003), log BSR falls by 0.53 after the rollout (SE 0.29, $p = 0.079$), pointing to improved Amazon-side demand. This is triangulation, not identification, since Flipkart rank histories are unavailable.

Table 11: Amazon Best Sellers Rank Diagnostic

Specification	$\hat{\beta}$	% rank change	SE	p-value	Obs.	ASINs
Amazon BSR pre/post with ASIN FE	-0.529	-41.06	0.287	0.079	1213	22
Amazon BSR pre/post with ASIN FE + calendar-month FE	-0.534	-41.39	0.289	0.079	1213	22

Notes: Lower BSR means better sales rank. This is an Amazon-only diagnostic, not a cross-platform DiD, because Flipkart sales-rank histories are unavailable. The result should be read as demand-side triangulation only.

Table 12 summarizes Amazon-side seller diagnostics. The matched-panel Amazon products are overwhelmingly third-party listings with FBA fulfillment and multiple seller histories, so the estimate reflects marketplace-seller repricing rather than Amazon Retail pricing policy.

Table 12: Amazon Supply-Side Diagnostics from Keepa

Measure	Count/value	Percent
Gold ASINs with raw Keepa JSON	1196	
Current buy box is Amazon Retail	0	0.0
Current buy box is FBA	695	58.1
Any Amazon offer in retrieved offers	0	0.0
Any third-party offer in retrieved offers	1184	99.0
Median current total offer count	1.0	
Median unique sellers in retrieved offers	2.0	
Median unique buy-box sellers in history	2.0	

Notes: Diagnostics are product-level summaries from raw Keepa JSON for gold-sample Amazon ASINs. They describe Amazon-side supply structure only; Flipkart seller identities are not observed in the final panel, so common cross-platform seller overlap cannot be measured here.

7.5 Alternative Counterfactuals and Randomization Inference

Table 13 shows that the price decline is not specific to the additive TWFE counterfactual (Baker et al. 2025). BJS imputation gives -0.037 in the full panel and -0.032 in the strict-support sample; matrix completion gives -0.030 to -0.031; interactive fixed effects gives -0.025 to -0.026. As discussed in Section 6, this agreement indicates robustness to the functional form of the counterfactual, not robustness to pre-trend violations.

Randomization-based inference adds a complementary diagnostic. In the strict-support sample, a TWFE sign-flip test that randomly reassigns the treated platform within ASINs produces a two-sided p-value of 0.045 and a one-sided p-value of 0.023. The placebo-time permutation is not significant ($p = 0.319$), confirming that the treatment month is not a sharp break relative to other candidate dates: the post-treatment Amazon–Flipkart contrast is unusually negative under random platform relabeling, but the timing of the response is gradual.

Table 13: Alternative Counterfactual Constructions

Estimator	Support	Estimate	SE	p-value	ASINs
BJS imputation	Full gold	-0.037	0.012	0.003	1,196
BJS imputation	Strict support $k = 6$	-0.032	0.012	0.006	146
Matrix completion	Full gold	-0.030	—	—	1,196
Interactive fixed effects	Full gold	-0.025	—	—	1,196
Matrix completion	Strict support $k = 6$	-0.031	—	—	146
Interactive fixed effects	Strict support $k = 6$	-0.026	—	—	146

Notes: Estimates are ATT-type effects on log price. BJS imputation follows Borusyak, Jaravel, and Spiess (2024), estimating untreated potential outcomes from untreated and not-yet-treated observations with product-platform and month fixed effects. Matrix completion and interactive fixed effects construct $\widehat{Y}_{i,Amazon,t}(0)$ using low-rank latent factor structure rather than only additive fixed effects. The BJS standard errors are clustered by ASIN. Matrix-completion/IFE standard-error calculation is unstable in this application, so those rows are reported as counterfactual-model diagnostics rather than independent confirmatory inference.

Table 14: Randomization-Based Inference in the Strict-Support Sample

Test	Observed effect	Two-sided p	One-sided p	Permutations
TWFE sign-flip	-0.025	0.045	0.023	5,000
ASIN-level sign-flip	-0.033	0.005	0.002	10,000
Placebo treatment date	-0.025	0.319	0.319	1,240

Notes: The sign-flip tests randomly swap the treated-platform label within matched ASINs. They test whether the Amazon-relative post-treatment decline is unusually negative relative to platform relabeling in the same matched support. The placebo-date test asks whether the December 2023 timing is unusually sharp relative to other candidate dates; it does not reject, consistent with gradual rather than discontinuous adjustment.

7.6 Limits of the Design

Four features of the design shape interpretation. First, Flipkart price histories are sparse: the Flipkart panel is built from live scraping, archived pages, and third-party trackers rather than from a commercial longitudinal feed, which generates changing support and exposes the full unbalanced panel to composition-driven movement. Second, cross-platform interference is plausible. Many Indian e-commerce products are sold by third-party sellers, and some sellers or repricing systems operate across platforms; if Amazon-side demand changes lead a common seller to update Flipkart prices, the Flipkart side is partially treated. Third, the aggregate Amazon–Flipkart gap moves before treatment, so the strongest claims cannot rest on a flat-pre-trends argument. The source-backed entry analysis sharpens this limitation rather than eliminating it: documented entrant exposure does not explain the full far-pre-period movement, but the negative estimate survives in low-entry categories, including Toys & Games, where third-platform vertical-entry shocks are less visible in the event file. Fourth, Flipkart introduced AI and personalization tools during the sample period, but the available evidence indicates that none was a directly analogous PDP-level review summary. To the extent these tools nevertheless affected Flipkart shopping behavior, the cross-platform contrast compares Amazon's review-summary intervention against a partially changing competitor environment.

These limitations do not invalidate the design, but they sharpen what it identifies. The estimate is best read as a relative cross-platform price response under explicit counterfactual assumptions, not as a clean average causal effect across all matched products. It is most informative where support is strongest and confounding pressure is weakest: in the strict-support sample, in low documented-entry categories such as Toys & Games, and among products with larger pre-treatment review corpora.

7.7 Interpretation

The economic chain has four steps. First, the feature lowers the cost of processing Amazon's review corpus. Second, this raises consumers' propensity to compare outside options, increasing the residual demand elasticity facing Amazon sellers—directionally tracked by the small-sample BSR result, in which Amazon log-rank falls by 0.53 ($p = 0.079$) post-rollout in the 22 products with usable rank histories. Third, prices adjust through marketplace-seller and buy-box repricing, consistent with the supply-side environment shown in Table 12. The slope-change estimates fit naturally at this step: the larger initial level shift followed by partial mean reversion is consistent with sellers or repricing systems adjusting quickly to a changed demand environment and then partially converging as the new pricing equilibrium settles. Fourth, the realized outcome is a relative Amazon–Flipkart price movement, attenuated to the extent that overlapping sellers or Flipkart-side tools move Flipkart prices in the same direction. The data identify the reduced-form relative price response and document heterogeneity that tracks this chain; the individual links remain mechanism statements rather than separately estimated structural parameters.

Read as a whole, the price movement appears across counterfactual constructions, varies with pre-treatment review information in the direction the model predicts, and is largest in the category where review narratives carry the most decision-relevant content. These patterns are most consistent with an elasticity-channel interpretation of the reduced-form estimate.

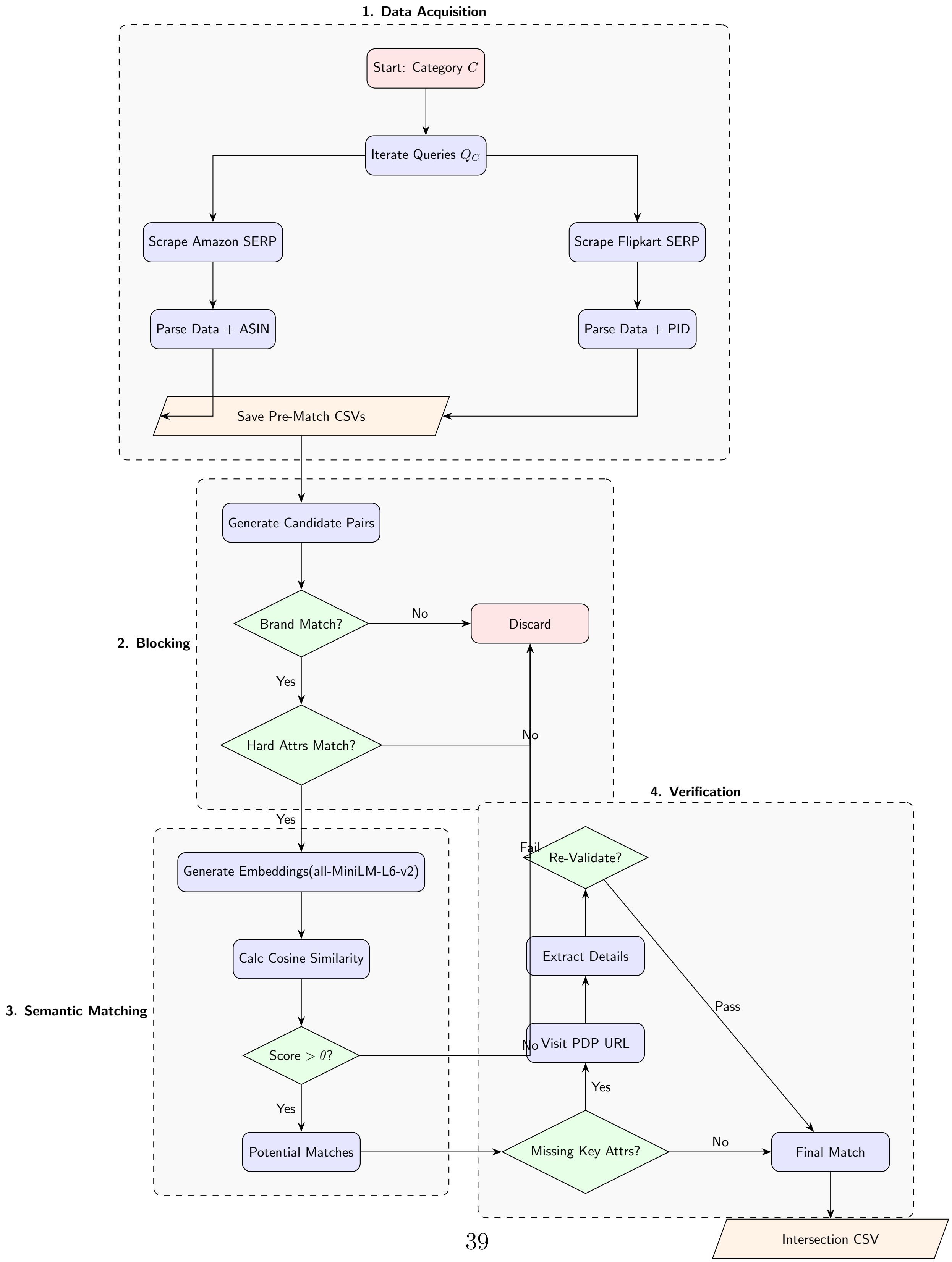
8 Conclusion

This paper assembles a monthly matched product-level panel across Amazon India and Flipkart (1,196 ASINs, 2019–2026) and exploits Amazon India's December 2023 rollout of AI review summaries as a platform-time shock. The preferred strict-support specification implies an Amazon-relative price decline of approximately 2.5 log points under the maintained counterfactual assumptions. The broader coefficient family is similar in sign but not identical in estimand: alternative counterfactual estimators generally imply 2–4 log-point declines, while the slope-change specification implies a larger initial level shift followed by partial mean reversion. Aggregate pre-trends and small local-window estimates indicate a gradual rather than discontinuous response; the sharpest evidence comes from the strict-support sample and from Toys & Games, the category in which review narratives are most decision-relevant, documented vertical-entry exposure is low, and the Rambachan–Roth no-deviation benchmark is independently satisfied.

The broader implication is that generative-AI tools that summarize existing information can have equilibrium consequences without changing product quality, seller costs, or the underlying review corpus. The substantive claim is narrow but specific: AI-mediated information design can lower relative prices when it increases consumer comparison and the residual demand elasticity facing sellers, and the cross-platform matched panel constructed here is one way to observe that response directly.

9 Appendix

9.1 Matching Pipeline



9.2 Treatment Timing

The exact page-level timing of AI-summary visibility is not observed for every Amazon India product. Amazon's public announcement states that the feature would be available in India starting December 12, 2023.[6] Random archived snapshots suggest that some desktop pages show visible summaries only in early 2024. A staged rollout across mobile and desktop interfaces, or pilot availability for a subset of products, cannot be ruled out. Because the empirical panel is monthly, the main regressions treat January 2024 onward as the effective post period.

If some Amazon pages received visible summaries later than the announcement date, the treatment indicator misclassifies some untreated Amazon-months as treated. For the negative price effects estimated in the paper, this timing misclassification should attenuate the magnitude toward zero under the usual monotone-treatment interpretation. It does not, however, solve the broader concern that the pricing response may be gradual rather than a sharp treatment-date discontinuity. This timing uncertainty is one reason the local-window and Rambachan–Roth sensitivity results are interpreted cautiously.

9.3 Sample and Composition Tables

Table 15: Analysis Sample Summary

Sample	Pairs	Obs.	Amazon obs.	Flipkart obs.	Months	Med. obs./pair-platforn
All cleaned product histories	1,360	48,094	29,430	18,664	86	13.(
Gold matched panel	1,196	46,652	28,656	17,996	86	13.(
Strict support k=6	146	14,167	7,779	6,388	86	46.(

Notes: The gold matched panel contains ASIN–PID pairs observed on both platforms at some point in the cleaned data. Stri pre-treatment and six post-treatment monthly observations on both platforms. Descriptive Amazon prices are divided by 10 units to rupees; regression estimates are invariant to this platform-specific level normalization because specifications include

[6] https://www.aboutamazon.in/news/retail/generative-ai-improves-customer-reviews

Table 16: Category Composition of the Matched Panel

Category	Gold pairs	Gold obs.	Strict pairs	Strict obs.
Office Products	307	13,964	60	5,805
Toys & Games	225	7,789	11	1,051
Home & Kitchen	204	8,598	28	2,731
Electronics	123	3,610	6	440
Beauty	104	5,144	23	2,479
Computers & Accessories	39	1,136	5	320
Outdoor Living	34	928	2	121
Sports, Fitness & Outdoors	20	564	2	202
Health & Personal Care	19	739	2	237
Home Improvement	17	392	1	67
Industrial & Scientific	14	428	1	74
Smartphones	12	142	0	0
Traditional Laptops	12	98	0	0
Video Games	9	238	0	0
Stick Ballpoint Pens	6	677	4	510
Gel Ink Rollerball Pens	5	395	0	0
Split-System Air Conditioners	4	50	0	0
Body Lotions	3	215	1	130
Lipsticks	3	140	0	0
Personal Computer	3	103	0	0
Other categories	33	1,302	0	0

Notes: Categories are the top-level category labels parsed from the final matched panel. The table reports the 20 largest categories by number of gold-panel ASIN–PID pairs; remaining categories are grouped as "Other categories."

Table 17: Observation Sources in the Gold Matched Panel

Platform	Source	Obs.	Share
Amazon	Keepa	28,656	61.4%
Flipkart	PriceHistoryApp / pretrend union	10,484	22.5%
Flipkart	PriceHistoryApp interpolation	5,999	12.9%
Flipkart	scraped price-history tracker	1,043	2.2%
Flipkart	Archive.org	462	1.0%
Flipkart	archived JSON backfill	8	0.0%

Notes: Source labels are retained from the final data-construction pipeline. Amazon observations come from Keepa; Flipkart observations combine scraped, archived, and open price-history tracker sources.

10 References

References

[1] Aparicio, Diego, and Kanishka Misra. Artificial intelligence and pricing. In K. Sudhir and Olivier Toubia, editors, *Artificial Intelligence in Marketing*, Review of Marketing Research, volume 20, pages 103–124. Emerald Publishing Limited, 2023.

[2] Aparicio, Diego, Zachary Metzman, and Roberto Rigobon. The pricing strategies of online grocery retailers. *Quantitative Marketing and Economics*, 22(1):1–21, 2024.

[3] Lynch Jr, John G., and Dan Ariely. Wine online: Search costs affect competition on price, quality, and distribution. *Marketing Science*, 19(1):83–103, 2000.

[4] Alavi, Soogand, and Salar Nozari. AI product review summaries and content homogenization: The changing landscape of UGC. SSRN working paper, 2025.

[5] Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021.

[6] Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730, 2021.

[7] Bai, Jushan. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

[8] Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285, 2024.

[9] Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. Difference-in-differences designs: A practitioner's guide. 2025.

[10] Bundorf, M. Kate, Maria Polyakova, and Ming Tai-Seale. How do humans interact with algorithms? Experimental evidence from health insurance. No. w25976, National Bureau of Economic Research, 2019.

[11] Brown, Zach Y., and Alexander MacKay. Competition in pricing algorithms. *American Economic Journal: Microeconomics*, 15(2):109–156, 2023.

[12] Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020.

[13] Chevalier, Judith A., and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.

[14] Chevalier, Judith, and Austan Goolsbee. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1:203–222, 2003.

[15] Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011.

[16] de Chaisemartin, Clément, and Xavier D'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

[17] Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan. Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–1859, 2018.

[18] Ellison, Glenn, and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the Internet. *Econometrica*, 77(2):427–452, 2009.

[19] Fisher, Ronald A. *The Design of Experiments*. Oliver and Boyd, 1935.

[20] Goodman-Bacon, Andrew. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

[21] Holland, Paul W. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[22] Hu, Han-fen, and Anjala S. Krishen. When is enough, enough? Investigating product reviews and information overload from a consumer empowerment perspective. *Journal of Business Research*, 100:27–37, 2019.

[23] Jang, Sungha, Ashutosh Prasad, and Brian T. Ratchford. How consumers use product reviews in the purchase decision process. *Marketing Letters*, 23:825–838, 2012.

[24] Liu, Licheng, Ye Wang, and Yiqing Xu. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 2022.

[25] Lei, Ke, and Yixuan Liu. When AI becomes a shopping advisor: A study on the impact of generative AI review on consumer purchase decision. *SAGE Open*, 15(3), 2025.

[26] Luca, Michael. Reviews, reputation, and revenue: The case of Yelp.com. Harvard Business School NOM Unit Working Paper 12-016, 2016.

[27] Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455, 2014.

[28] Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest. Prices and heterogeneous search costs. *The RAND Journal of Economics*, 48(1):125–146, 2017.

[29] Nelson, Phillip. Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329, 1970.

[30] Nosko, Chris, and Steven Tadelis. The limits of reputation in platform markets: An empirical analysis and field experiment. No. w20830, National Bureau of Economic Research, 2015.

[31] Rosenbaum, Paul R. *Observational Studies*. Springer, 2nd edition, 2002.

[32] Rubin, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[33] Rambachan, Ashesh, and Jonathan Roth. A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5):2555–2591, 2023.

[34] Seiler, Stephan. The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics*, 11:155–203, 2013.

[35] Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. What are we weighting for? *Journal of Human Resources*, 50(2):301–316, 2015.

[36] Sun, Monic. How does the variance of product ratings matter? *Management Science*, 58(4):696–707, 2012.

[37] Sun, Liyang, and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

[38] Su, Yi, Qili Wang, Liangfei Qiu, and Runyu Chen. Navigating the sea of reviews: Unveiling the effects of introducing AI-generated summaries in e-commerce. SSRN working paper, 2024.

[39] Sweller, John. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988.

[40] Wang, Huike, and Tianmei Wang. Does AI-Generated Review Summarization Affect Consumer Purchasing Behavior?—An Empirical Study Based on the Amazon Platform. 2025.

[41] Xie, Jianing, Mingqian Li, Rong Du, and David Townsend. The impact of generative AI models on consumer purchase behavior in e-commerce platforms: Evidence from a quasi-natural experiment. AIS eLibrary, 2025.