Do Review Solicitations Elicit Reviews Where They Matter for Sales and Returns?

Hana Choi* Minkyung Kim† Jinsoul Seo‡

November 14, 2025

*Assistant Professor of Marketing, University of Rochester (hana.choi@rochester.edu)

†Assistant Professor of Marketing, Carnegie Mellon University (minkyungkim@cmu.edu)

‡PhD student, University of Rochester (jseo13@simon.rochester.edu)

All authors contributed equally and are listed in alphabetical order. The authors thank the data provider for access to the data used in this study. The analyses, interpretations, and conclusions expressed here are solely those of the authors and do not reflect the views of the data provider. We also appreciate Yufeng Huang, Minjae Kim, Kosuke Uetake, and participants at CODE, Conference on AI/ML/BA , University of Rochester Marketing Seminar, Marketing Science Conference, London Business School Summer Camp, and Virtual Quant Marketing Seminar for their comments and suggestions. All remaining errors are our own.

Abstract

Firms solicit online reviews to enhance sales and reduce mismatches, yet it is unclear whether solicitations generate reviews where buyers benefit most, since they target reviewers while value accrues through buyers. Using rich individual-level panel data on transactions, reviews, and returns from a large apparel e-commerce retailer, we examine alignment between reviewer response and buyer value. Leveraging natural variation in solicitations, we estimate a two-stage framework: Stage 1 measures how solicitations affect review generation across informational states, and Stage 2 quantifies how these induced reviews influence subsequent sales and returns. Additional reviews increase sales and reduce return rates by up to 0.25 percentage points for products with zero-to-low prior reviews. However, solicitations face a first-review barrier, failing to elicit reviews in these high-value states, and thereby creating a reviewer-buyer misalignment. Counterfactual simulations indicate that raising first-review likelihood by two percentage points increases annual net revenue by $120 per product, highlighting the value of targeting zero-review products.

Keywords: online word-of-mouth (WOM), online reviews, review solicitation, review system design, product returns, natural experiments, e-commerce

JEL Classification Codes: D4, L1, L2, L81, M3, M31, M37, C93

1 Introduction

1.1 Overview

Online reviews influence a wide range of consumer decisions, from which products to purchase to whether to try a new restaurant, by enabling more informed choices that improve ex ante quality assessment and product–consumer matching. Recognizing this impact, firms do not passively wait for reviews; they actively solicit them, typically through post-purchase messages inviting buyers to rate and review as illustrated in Figure 1. One industry survey reports that around 80% of consumers have been asked by a seller to leave a review (Paget, 2023). While these messages target consumers who have already experienced the product, firms ultimately care about future market outcomes such as sales and returns, rather than the reviews themselves. The managerial hope is that solicitations generate reviews where they matter most.

This raises our key question: do review solicitations allocate review supply to where buyers need it most, thereby enhancing sales and reducing mismatches? The answer is not obvious, as solicitations target reviewers, but value accrues through buyers. The effectiveness of solicitation thus depends on both where reviewers choose to respond and the incremental informational value that buyers gain, highlighting that its economic value hinges on alignment between reviewer response and buyer value.

In addressing the research question, we highlight product returns as a key outcome measure alongside sales. Theoretically, returns are informative about mismatches, as they provide a direct measure of discrepancies between customer expectations, shaped by reviews, and the product's realized fit. Practically, returns are of first-order importance as they impose substantial costs on both sides of the market. For consumers, returns involve hassle costs (repacking, shipping, and time) that can reduce satisfaction and future purchase intent, even when returns are nominally free. For retailers, returns require reverse logistics, customer ser-

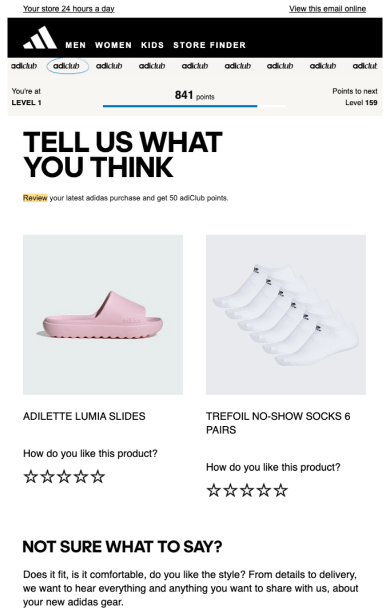


Figure 1: Example of review solicitations

vice, inspection, restocking, and, in the case of apparel, frequent liquidation. The macro stakes are large: U.S. retailers estimate that 16.9% of sales were returned in 2024 (approximately $890 billion), with online channels typically experiencing higher return rates than brick-and-mortar stores (e.g., 17.6% online versus 10.0% in-store in 2023).[1] Reviews speak directly to this problem by shaping expectations and match quality ex ante; richer information should lower returns by reducing uncertainty (Sahoo et al., 2018). Consequently, if review solicitations succeed in generating reviews where information is needed most, they can both raise sales and reduce returns by mitigating post-purchase misfit.

We study this question in partnership with a large apparel e-commerce retailer in South Korea that sells multiple products under a single brand and sends unincentivized SMS review solicitations after delivery. In our empirical setting, solicitation exposure varies naturally due to the retailer's monthly free allotment of SMS messages, which is managed by a third-party review service provider. This institutional setting generates plausibly exogenous

[1]Source: https://nrf.com/research/2023-consumer-returns-retail-industry, https://nrf.com/research/2024-consumer-returns-retail-industry

variation in which customers receive review solicitations, shifting review supply independently of demand-side preferences or shocks and thus providing an ideal context for causal identification. Combined with rich, individual-level panel data covering website visits, purchases, solicitations, reviews, as well as returns, our dataset provides a rare opportunity to examine how review solicitations asymmetrically affect reviewers and buyers, thereby shaping downstream market outcomes. It enables us to address endogeneity in review generation, disentangle multiple sources of selection bias, and assess the informational role of reviews in reducing buyer uncertainty, improving product matching, and evaluating the effectiveness of marketing levers such as review solicitation.

Our empirical strategy mirrors the reviewer–buyer link and proceeds in two stages. Stage 1 (reviewer response) estimates how solicitations affect the likelihood of review submission and the distribution of ratings, conditional on a product's existing review state (e.g., number of prior reviews and average rating). Stage 2 (buyer value) leverages the solicitation-induced variation in review supply from Stage 1 to estimate the causal, state-contingent marginal effects of an additional review on subsequent orders and return rates. Comparing where reviews are generated with where they yield the greatest impact on downstream sales and returns reveals our central object of interest: the degree of reviewer–buyer (mis)alignment.

We begin by showing that for buyers in Stage 2, additional reviews causally improve market outcomes, with effects strongest where existing information is scarcest or noisiest. Leveraging solicitation induced changes in review volume, we find that an additional review in the prior week increases orders in the subsequent week by about 15% and reduces return rates by 0.1 percentage point, consistent with prior studies that online reviews can meaningfully influence market outcomes.(e.g., Chevalier and Mayzlin, 2006; Cabral and Hortaçsu, 2010; Anderson and Magruder, 2012; Sun, 2012; Zhao et al., 2013; You et al., 2015; Luca, 2016; Sahoo et al., 2018; Reimers and Waldfogel, 2021). Additionally, we find the reductions in returns are particularly pronounced at zero-to-low reviews and for products where prior information is noisy, reaching up to a 0.25 percentage point decrease. This empirical pattern we document point to an information-based mechanism: reviews clarify size, material, and usage expectations up front, reducing ex-post corrections through returns. While sales effects follow a similar pattern, the asymmetry is sharper for returns, indicating that the primary channel operates through misfit reduction rather than demand stimulation alone. When existing information is scarce or when signals are yet noisy, an additional review provides substantial marginal value by reducing buyer uncertainty. Conversely, we find that additional reviews for products with many existing reviews have a much smaller impact on sales and almost no effect on returns. Accordingly, the economic value of solicitation depends on its ability to effectively nudge reviewers to respond where buyers face the greatest uncertainty, that is in the early review state, whereas for products with many existing reviews, an additional review adds little incremental benefit.

Looking at reviewer responses in Stage 1, solicitations expand review supply, increasing the likelihood of writing a review by about one percentage point. The conditional probability of a five-star rating declines modestly (by 1.5 percentage points), consistent with solicitations drawing in marginal reviewers with less extreme experiences (e.g., Askalidis et al., 2017; Karaman, 2021; Brandes et al., 2022). Yet, these average effects mask important state dependence.

A closer examination of state contingency, juxtaposing reviewer response with buyer value, reveals a pronounced misalignment: solicitations fail to generate reviews in the informational states where they would be most valuable. The effect on review incidence is negligible for products with no prior reviews, suggesting a "first-review barrier", but becomes strongest once a minimal review base exists, particularly in noisy information environment (i.e., below-median average rating in our context). Writing the very first review appear to entail a higher psychological cost, which our interventions are insufficient to overcome. This misalignment is amplified for high-priced items, which exhibit even sharper state dependence with larger return reductions, especially at zero-reviews. Early reviews thus carry disproportionate value for expensive goods, where financial risks are higher, information is scarce due to slower sales velocity, and retailer margins are large. Once a minimal review base is established though, high-priced products respond at least as strongly as low-priced ones, indicating that overcoming the zero-review state for expensive items could yield substantial gains both for buyers and the retailer. Shifts in rating composition are modest and rarely large enough to move product-level means in economically meaningful ways, implying that the extensive margin dominates. Overall, solicitations primarily expand review volume where a minimal review base already exists, rather than in early-review states where buyers would derive the greatest benefit.

Therefore, "reviewer–buyer misalignment" leaves substantial value on the table. We thus quantify its revenue consequences using a counterfactual exercise focused on the first review. By varying the solicitation effect at the zero-review state and feeding the resulting review supply into the Stage 2 outcome model, we simulate changes in orders, return rates, and net revenue. An increase in the first review likelihood by 2 percentage points translates into an annual net revenue increase of $120 per product, driven by the high marginal value of reviews when information is scarce and noisy. This exercise demonstrates that even modest improvements in eliciting first reviews can materially reduce mismatches and returns, underscoring the managerial importance of targeting zero-review states where incremental alignment delivers the largest economic payoff.

1.2 Relevant Research

While a substantial body of literature examines the dynamics of review generation and its impact on market outcomes, relatively little attention has been paid to how marketing interventions create value by allocating review supply and shaping these outcomes.

Broadly, this paper contributes to the literature on the dynamics of review generation.

Reviews are subject to biases arising from buyer selection and the voluntary nature of review writing. Because review generation depends on buyers, accumulated review information is shaped by sales velocity and who buys when (e.g., Li and Hitt, 2008; Godes and Silva, 2012; Park et al., 2021), as well as by the selection of who chooses to write reviews and their underlying motivations (e.g., Hu et al., 2009; Moe and Schweidel, 2012; Muchnik et al., 2013; Hu et al., 2017; Schoenmueller et al., 2020; Chakraborty et al., 2022; Sunada, 2025). Building on this literature, we examine how review generation depends on existing review information states. Even after accounting for sales velocity, review generation remains highly state dependent. We document a first-review barrier, wherein initiating a review in a previously unreviewed state entails substantial (psychological) costs, particularly for products requiring complex fit assessments. Our findings highlight frictions in early review formation and inform strategies to overcome them.

A growing body of research examines how firms' interventions shape the dynamics of review generation, for example, through managerial responses (e.g., Proserpio and Zervas, 2017; Chevalier et al., 2018; Wang and Chaudhry, 2018), various incentive schemes (e.g., Khern-am nuai et al., 2018; Sun et al., 2017; Burtch et al., 2018; Woolley and Sharif, 2021), or the purchase of artificial reviews (e.g., Mayzlin et al., 2014; Luca and Zervas, 2016; He et al., 2022; Gandhi et al., 2025). We focus on one such intervention: solicitation messages, which have been shown to reliably increase review incidence and reduce extremity bias (e.g., Askalidis et al., 2017; Karaman, 2021; Brandes et al., 2022; Gao et al., 2025). However, our results reveal that the solicitation effects documented in prior studies do not extend to the zero-review state (where buyer value is highest), but instead emerge only once a minimal review base exists, with the effect particularly pronounced in noisy information environments.

For marketers, the dynamics of review generation are highly relevant and of great interest, as accumulated review information has been shown to influence downstream sales and revenues across various categories and contexts (e.g., Chevalier and Mayzlin, 2006; Cabral and

Hortaçsu, 2010; Chintagunta et al., 2010). When reviews improve market outcomes by signaling vertical quality and horizontal fit, thereby reducing uncertainty and enhancing matching, their impact is expected to be state-contingent: an additional review matters most where prior information is scarce, dispersed, or less credible (e.g., Forman et al., 2008; Anderson and Magruder, 2012; Sun, 2012; Zhao et al., 2013; You et al., 2015; Luca, 2016; Reimers and Waldfogel, 2021). More broadly, Godes and Mayzlin (2009) show how firms' exogenously created word-of-mouth conversations affect sales depending on product's existing information, which our paper complements with specific focus on online reviews. Specifically focusing on firms' interventions in review systems and their impact on downstream outcomes, Fradkin and Holtz (2023) investigate the effect of monetary review incentives for stays with no prior reviews. They find that the treatment generated additional, more negative reviews but did not increase sales or revenue. They suggest several reasons for this outcome, including supply constraints (listings are booked even without reviews and transactions occur quickly for control listings), Airbnb's policy of hiding average ratings for few-review listings, and capacity limits (only one buyer can book a listing per night). Our study extends this literature by examining solicitation messages in a context largely free from supply or capacity constraints, showing that even unincentivized interventions can meaningfully influence both sales and returns. Importantly, we find that these downstream effects are strongly state-dependent: buyers derive the greatest value from additional reviews when prior review information is absent, limited, or dispersed; yet, a pronounced first-review barrier renders solicitation efforts least effective precisely in these high-value states.

Lastly, our study contributes to the important yet largely understudied literature on firms' interventions regarding product returns, such as shipping and return policies (e.g., Wood, 2001; Anderson et al., 2009; Janakiraman et al., 2016; Shehu et al., 2020). Review information plays a central role in shaping returns by influencing consumers' pre-purchase expectations regarding vertical quality and horizontal fit, thereby reducing discrepancies between expectations and post-purchase experiences. Most closely related to our work, Sahoo et al. (2018) show that richer review information mitigates uncertainty and, in turn, lowers returns. We extend this literature by examining review solicitations as a firm intervention and by decomposing their effects into reviewer and buyer responses, contingent on existing review information states. Our results reveal a reviewer–buyer misalignment that is particularly pronounced for returns when prior information is scarce or noisy. While firms may have incentives to further bias reviews (e.g., through fake reviews), our findings highlight a countervailing motivation: firms may use marketing interventions to reduce informational bias and self-correct, especially when returns impose substantial costs on both consumers and firms.

1.3 Organization

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and data. Section 3 describes the exogenous natural of our review solicitation and presents the two-stage empirical strategy. Section 4 reports the main results on reviewer response and buyer value and documents the reviewer–buyer misalignment, including heterogeneity by price. Section 5 develops counterfactual simulations to quantify revenue impacts of improved reviewer-buyer alignment. Section 6 concludes with implications for managing review supply.

2 Data and Review Environment

This section describes the data used in this study and the review environment that writers and buyers encounter along their consumer journey in our research context.

2.1 Data

The data come from a large clothing retailer in South Korea that operates its own independent e-commerce website (similar to Nike operating nike.com). The dataset captures user activities in 2019, along their consumer journey, including visits, purchases, returns, and review behaviors. The data are organized as follows:

1. Visit Data: Contains each user's visit and its timestamp.

2. Transaction Data: Each observation includes user ID, order ID, product ID, price paid, purchase date, a solicitation dummy (indicating whether a review was solicited), a review dummy (indicating whether a review was posted), and a return dummy (indicating whether the product was subsequently returned).

3. Review Data: For observations with posted reviews, we also observe the buyer's rating on a five-star discrete scale, the review text (currently being collected), and whether a photo was included with the review.

4. Product Data: Regarding product attributes, we observe the product name, category, launch date, and importantly the cumulative number of reviews posted for each of the five rating levels up to a given product-day. These data allow us to construct measures such as the total number of reviews, average rating, share of five-star ratings, and variance of ratings, essentially the key review information visible to a buyer on any given product-day when making a purchase decision.

5. User Data: Contains user-level characteristics, including recency, frequency, and monetary (RFM) metrics from the pre-sample period. Specifically, we observe the number of orders, total purchase amount, and purchase dates for 2018, if any purchases were made in the year prior to our analysis period.

Taken together, these components form a panel dataset that tracks visits, purchases, reviews, and returns across users and products.

It is worth noting several unique features of our data that make it particularly well-suited to addressing our research questions. First, unlike most scraped datasets, ours includes both non-buyers (users who visited but did not purchase) and non-reviewers (buyers who purchased but did not post reviews). This distinction is critical for addressing self-selection biases among online reviewers. Hu et al. (2017) identify two of such biases. The first is acquisition bias, where consumers with favorable predispositions toward a product are more likely to purchase and subsequently review it. The second is under-reporting bias, where consumers with extremely satisfied or dissatisfied experiences, rather than moderate ones, are more likely to generate reviews and ratings, a pattern commonly referred to as the J-curve. Having data on both non-buyers and non-reviewers enables us to disentangle these two sources of selection and to examine how marketing levers, such as review solicitation, affect both.[2] Second, we observe individual-level purchases as well as return outcomes. This is crucial for quantifying the informational value that reviews provide. When changes in review information affect not only sales but also returns, it reveals the role of reviews in reducing buyer uncertainty and improving product matching. Returns provide a direct and observable measure of how review information influences consumers' decisions, and in turn, affects the retailer's market outcomes. Third, our dataset contains natural variation in review solicitation, which exogenously shifts the review states of products (see Subsection 3.1 for details). Because reviews are inherently endogenous, for example, popular products with high unobserved demand shocks tend to receive both more reviews and higher sales, this exogenous variation enables us to effectively address endogeneity concerns.

During our sample period, 83,409 users made 321,420 purchases of 3,112 products on

[2] In this draft version, our analysis focuses primarily on under-reporting bias, while our ongoing work extends to incorporate acquisition bias as well.

the retailer's website. Most products are sold infrequently, with the majority of revenue generated by a small subset. We therefore focus on the top products by revenue and exclude three with an extreme number of reviews ($\geq 1,200$) to reduce potential outlier bias in our regression estimates. The resulting 170 focal products account for 57.5% of the retailer's revenue and 50.0% of all reviews generated in 2019. We further restrict our analysis to users who purchased at least one of these focal products during the sample period, following a common approach in scanner data studies that filter customers based on minimum category purchases (e.g., Guadagni and Little, 1983; Gupta, 1988). This results in 749,543 visits, 153,429 purchases, 13,949 reviews, and 7,341 returns for the 170 products by 63,102 users. Details on sample construction and the operationalization of product prices are provided in the online Appendix A.

2.2 Review environment

2.2.1 Review information displayed to buyers

Consumers considering a purchase can view a summary of review information prominently displayed on each product detail page. Figure 2 illustrates a typical summary of review information displayed on a product detail page of our retailer's website. In our empirical setting, this summary includes the total number of reviews, the average rating (calculated as the simple average of all cumulative ratings to date and rounded to one decimal place), the share of five-star ratings, and the cumulative number of reviews posted for each of the five rating levels (as shown on the right panel of Figure 2). Below this summary section, consumers can scroll down further to read individual reviews in detail.

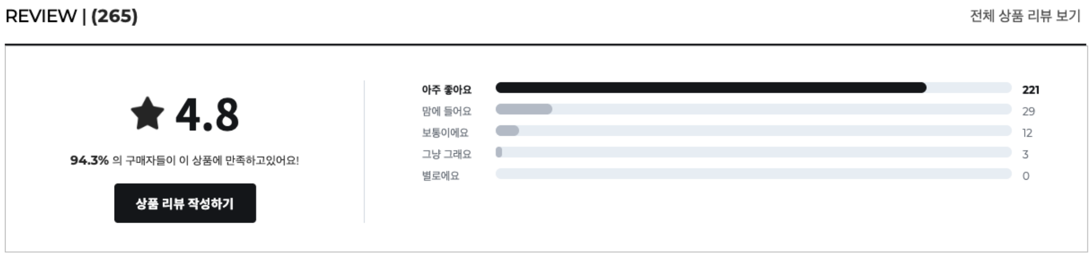


Figure 2: Summary of reviews displayed to buyers

2.2.2 Review writing process and solicitation

Consumers can leave a review after purchasing a product. Upon delivery, the retailer sends an unincentivized SMS message, without any monetary rewards or promotions, to remind customers to share feedback and to encourage the submission of reviews by streamlining the process. As shown in Figure 3, the message typically includes a brief prompt such as "How is the product?" along with a convenient button that directs users to the review submission page without requiring them to log in to the retailer's website. In addition to submitting reviews via this SMS link, consumers can also log in directly to the website to write reviews after receiving their orders. When submitting a review, consumers are asked to provide a star rating on a five-point scale, a written comment and optionally upload a photo. They are also invited to fill in additional fields, such as height, weight, and usual clothing size. Together, these components form each individual review displayed on the product detail pages.[3]

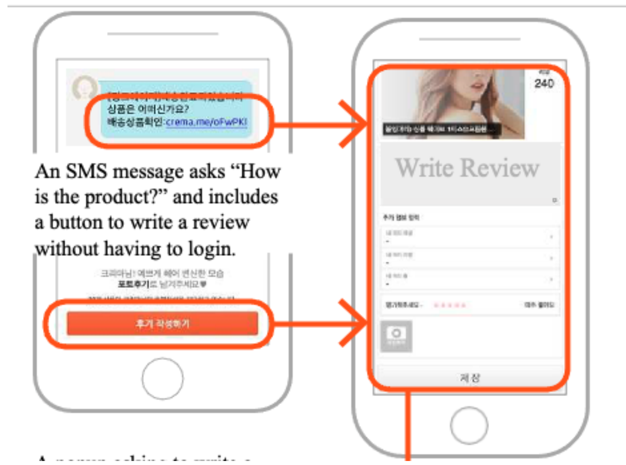
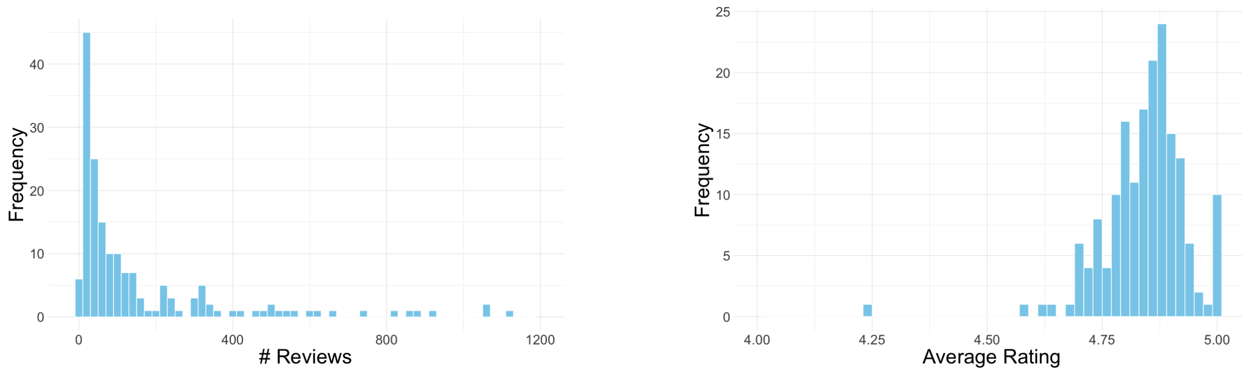


Figure 3: Review solicitation via SMS message

Table 1: Summary Statistics

	Mean	SD
User-level activities in 2019		
# Purchases	2.58	2.15
# Reviews	0.24	0.85
Product-day level flow metrics		
# Purchases	6.32	9.39
% Returns (=returns/purchases)	5.40	14.39
# Reviews	0.59	1.20
% Five-star rating	89.72	26.76
Product-day level cumulative metrics to date		
# Reviews	274.83	513.82
Average rating	4.86	0.10
% Five-star rating	90.46	6.68
Variance of ratings	0.18	0.10



(a) Number of Reviews

(b) Average Rating

Figure 4: Distribution of #reviews and average rating

2.3 Data description

Table 1 presents the summary statistics. The top panel reports user activity: each user makes about 2.6 purchases on average and writes 0.24 reviews over the sample period, corresponding to an approximately 9.3% probability of writing a review. Given the limited number of reviews per user, our identification relies primarily on cross-sectional (across-user) variation, leaving limited scope to analyze within-user effects. The second panel reports daily flow metrics at the product-day level. On average, each product receives 6.32 purchases per day, of which 5.4% are returned, with considerable variation particularly in the return rate. Products also receive an average of 0.59 reviews per day.

The third panel reports review information aggregated to date, corresponding to the summary review information visible to buyers on the product detail page (Figure 2). A closer look at the sanpshot from December 31, 2019, is shown in Figure 4. The left panel displays the distribution of the number of reviews across products. While the mean is 274.8,

[3]Our data include all reviews visible to consumers, each linked to an actual order containing detailed information such as user ID, order ID, product ID, price paid, delivery date, and delivery address. We therefore consider these to be authentic reviews, with no evidence suggesting the presence of fake reviews. Since our setting involves an e-commerce retailer selling its own products on its independent website, the incentive to purchase fake reviews is likely minimal, unlike third-party sellers on platforms like Amazon, who face intense competition and may use inflated ratings to gain visibility or outrank rivals (Gandhi et al., 2025).

the distribution exhibits substantial variation and is heavily right-skewed. Many products have very few reviews, highlighting the potential for marketing interventions, such as review solicitation, to encourage buyers to share feedback, while also raising the question of whether such interventions are effective. The right panel presents the distribution of average ratings across products. The mean average rating is 4.86, with roughly 90% of these ratings being five stars. This extremely positively skewed pattern suggests strong self-selection among reviewers and underscores the importance of accounting for such selection when examining how review solicitations influence market outcomes.

It is important to note that our data do not exhibit the well-known J-curve (Hu et al., 2009; Dellarocas and Wood, 2008), but rather follow an exponential pattern, as also observed in Sunada (2025). The highly positively skewed, exponential-shaped distribution implies that lower average ratings are associated with higher variance.[4] Indeed, there is a strong negative correlation of -0.71 between average rating and variance. Consequently, we focus on average rating as our primary outcome.

To illustrate the variation in review information accumulated across products, we plot existing review volume and price by product category in Figure 5. Higher-priced items that generate greater revenue when sold and products in the Outers and Set categories, which likely involve greater purchase uncertainty due to more complex fit assessments (compared to basic items such as T-shirts or pants), tend to accumulate fewer reviews. This pattern highlights opportunities for review solicitations to reduce uncertainty through enhanced information availability and improve revenue generation for high-margin products. Ideally, we would hope marketing interventions to be most effective where reviews are most scarce yet most needed. In what follows, we empirically assess this question in detail.

[4]When the underlying selection produces a J-curve, as in Karaman (2021), review solicitation often encourages consumers with moderate experiences to write reviews. In such cases, the impact on the average rating is ambiguous, typically minimal, while variance decreases as more moderate reviews are added instead of extreme ratings.

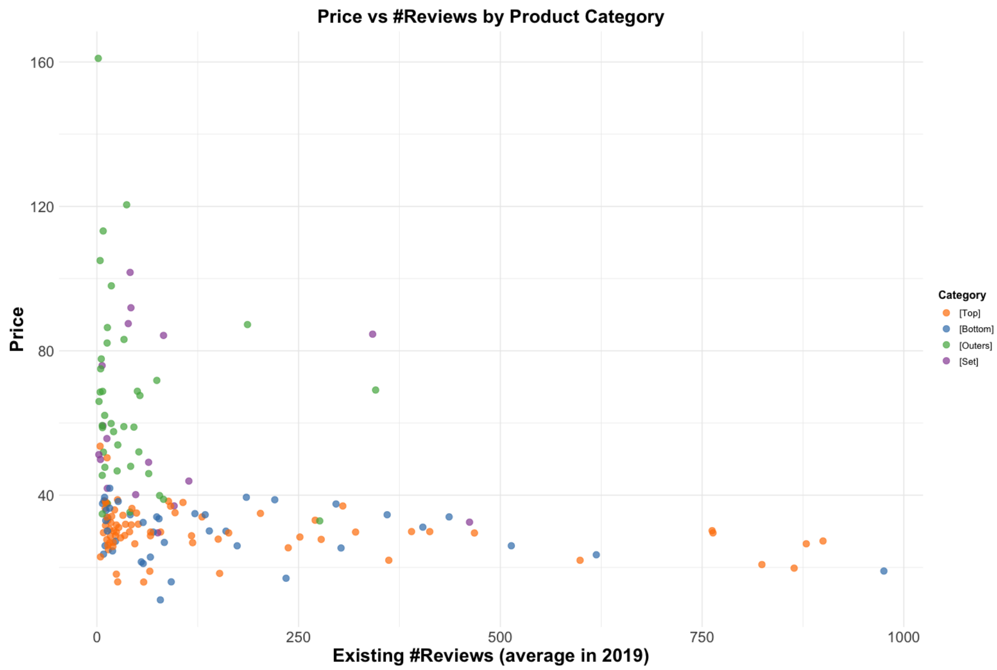


Figure 5: Reviews and product characteristics

3 Empirical Strategy

In this section, we first discuss the exogenous nature of our review solicitation, and then present the empirical framework used to measure the heterogeneous effects of review solicitations on reviewer and subsequent buyer responses.

3.1 Natural experimental variation in review solicitation

The retailer in our study, as is typical for independent e-commerce websites, integrates a third-party solution into its website to collect, manage, and display customer reviews.[5] Under the retailer's monthly subscription plan with the review service provider, a certain number of review solicitation messages are included at no additional cost. Although the retailer could allocate additional marketing budget beyond the monthly subscription fee to

[5]Popular solutions in the U.S. e-commerce industry include Yotpo, Judge.me, and Growave, which offer features such as photo and video reviews, automated email requests, and integrations with marketing tools.

send more SMS review reminders once the free allotment is exhausted, it did not do so. This setup creates natural, exogenous variation in which customers receive review solicitations.

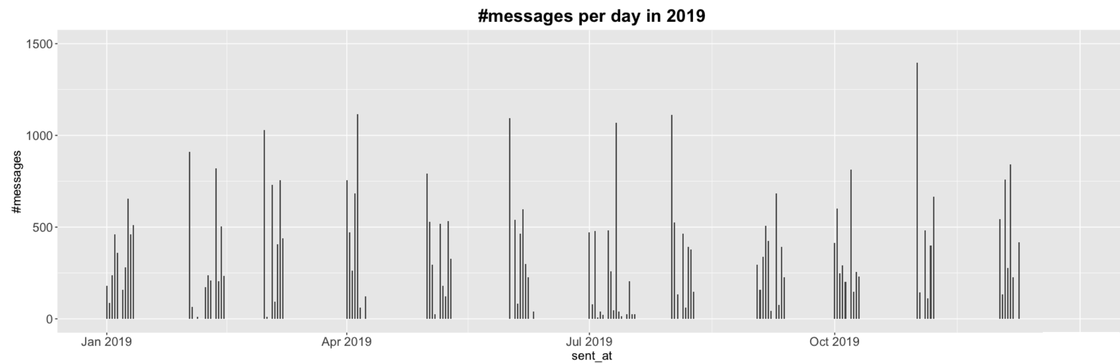


Figure 6: Natural experimental variation in review solicitation

Table 2: Balance Check

	Solicited	Control
Price ($)	30.91 (13.62)	30.90 (13.73)
Product age (days)	296.54 (255.53)	293.41 (256.81)
# Reviews	444.60 (705.95)	446.52 (720.71)
Average rating	4.87 (0.078)	4.87 (0.096)
Variance in ratings	0.171 (0.083)	0.172 (0.088)

Mean (SD) across orders: level of treatment

Because of this setup, review solicitations are sent for all orders placed from the beginning of each month until the free message quota is depleted. Figure 6 displays the daily number of review solicitations sent in 2019. On average, about 2,600 solicitations are sent each month, typically within the first ten days.[6] Overall, 28.6% of orders received a solicitation. At the user level, approximately 38% of buyers received at least one review solicitation, while the remaining 62% never received any. Since review solicitations are triggered by actual orders, both the daily spikes and the number of solicitation days per month reflect fluctuations in consumer demand. Table 2 reports the balance between orders with and without review

[6]The monthly quota of free messages is not constant across months, as the provider occasionally offers random giveaways, for instance, when system glitches occur during updates or when the review section experiences performance issues.

solicitations. Across multiple dimensions of product characteristics, including price, product age, and review information (number of reviews, average rating, and rating variance displayed at purchase), we find no significant differences, indicating that the two groups are well balanced. Of note, these solicitations are reminder messages with no monetary incentives or promotions, and the retailer does not label them, meaning buyers cannot distinguish solicited from unsolicited reviews on the product page.

3.2 Framework

This subsection lays out how we use post–delivery review solicitations as an exogenous lever to (i) measure how reviewers respond across information states (Stage 1) and (ii) identify the causal, state–contingent marginal effect of additional reviews on next–period orders and return rates (Stage 2), eventually to compare reviewer response and buyer value by each state of a product. The design separates the supply of reviews from their demand-side value by first estimating how solicitations change review incidence and composition and then mapping those solicitation–induced reviews into market outcomes observed by subsequent buyers. Throughout, we exploit the fact that solicitations are sent after delivery, contain no promotions, and are not visible to future buyers, so their effect on orders and returns operates only through review creation. We implement both average and heterogeneous effects by the product's *review state*.

We proxy the information environment buyers face with a product's *review state* observed up to the day (Stage 1) or week (Stage 2) before the solicitation window. States are defined by the cross of review–count bins and rating bins. In our setting, average rating is strongly negatively correlated with rating dispersion. The below–median rating bin therefore also captures a higher–variance (noisier) information state. We index states by $k \in \mathcal{K}$ and write $S^{(k)}_{jt-1} = 1$ if product j was in state k through period $t-1$. All state dummies are measured pre–solicitation to avoid mechanical feedback.

3.2.1 Stage 1: Reviewer response

Stage 1 estimates how solicitations change individual review behavior, allowing the effect to vary with the pre–existing review state. The unit is user i–product j–day t. Outcomes are (i) whether a review is written and (ii) conditional on writing, whether the rating is five stars. The baseline specification is

$$Y_{ijt} = \alpha + \sum_{k \in \mathcal{K}} \gamma_k \big(\mathsf{Solicited}_{ijt} \times S^{(k)}_{jt-1}\big) + \sum_{k \in \mathcal{K}} \beta_k S^{(k)}_{jt-1} + X'_{jt}\theta + \lambda_{m(t)} + \varepsilon_{ijt}, \tag{3.1}$$

where $\mathsf{Solicited}_{ijt} \in \{0,1\}$ indicates receipt of the post–delivery SMS request; X_{jt} includes product age and a new–release indicator (price is added in robustness); and $\lambda_{m(t)}$ are month fixed effects.[7] Coefficients γ_k are the state–specific treatment effects on review incidence (or on five–star probability conditional on writing). We estimate (3.1) with heteroskedasticity–robust standard errors; in robustness we cluster by product. We allow the treatment effects to vary by the review state k.

From Stage 1 to a predicted review stock. For each product–week (j,t) we aggregate the state–specific treatment effects into a *predicted* number of reviews caused by solicitations,

$$\widehat{N}^{\,\text{pred}}_{jt} = \sum_{i \in \mathcal{O}_{jt}} \sum_{k \in \mathcal{K}} \widehat{\gamma}_k \; \mathbf{1}\{\mathsf{Solicited}_{ijt} = 1\} \; \mathbf{1}\{S^{(k)}_{jt-1} = 1\}, \tag{3.2}$$

where $\mathcal{O}_{jt}$ is the set of orders of product j during week t that were eligible to receive a solicitation. We also compute state–specific components $\widehat{N}^{\,\text{pred}}_{jt}(k)$ that assign each eligible order to its pre–week state k. These objects summarize the exogenous variation in review

[7] We do not include product fixed effects in the main specification, because the products are highly correlated with review states $S^k_{j,t-1}$, e.g., high-priced items do not have many transactions so have a low number of existing reviews. In the appendix, we show the qualitatively similar results with product fixed effects included, but we believe the specification without product fixed effects is more effective to show our main story about the state-contingent effect in reviewer response and buyer value.

supply induced by solicitations that we carry into Stage 2.

3.2.2 Stage 2: Buyer value

Stage 2 maps solicitation–induced reviews into market outcomes for subsequent buyers. The unit is product j–week t. To avoid mechanical contemporaneous correlations, we use non–overlapping windows: review shocks pertain to week t, while outcomes are measured in week $t+1$. We study two outcomes:

$$Y_{j,t+1} \in \left\{ \log\left(\mathsf{Orders}_{j,t+1} + 1\right),\ \mathsf{ReturnRate}_{j,t+1} \right\}.$$

Our preferred specification instruments the *observed* number of reviews in week t, N_{jt}^{obs}, with the constructed shocks $\widehat{N}_{jt}^{\text{pred}}$ from (3.2) and estimates the following second stage:

$$Y_{j,t+1} = \tilde{\alpha} + \sum_{k \in \mathcal{K}} \delta_k \big(N_{jt}^{\text{obs}} \times S_{j,t-1}^{(k)} \big) + \sum_{k \in \mathcal{K}} \tilde{\beta}_k S_{j,t-1}^{(k)} + X_{jt}' \tilde{\theta} + \lambda_{m(t)} + \tilde{\varepsilon}_{jt}. \tag{3.3}$$

The δ_k parameters are the state–contingent marginal effects of one additional review during week t on next–week outcomes, for products that comply with solicitation–induced changes in review supply. For comparison, we also report (i) a reduced–form "intent–to–treat" regression that replaces N_{jt}^{obs} with $\widehat{N}_{jt}^{\text{pred}}$, and (ii) an OLS specification that uses N_{jt}^{obs} directly. Month fixed effects absorb seasonality; X_{jt} includes the same controls as in Stage 1. Again, we allow the Stage–2 effects to vary by the review state k, which in combination with the results in Stage 1, directly tests whether review supply is allocated to states where buyer value is highest.

3.3 Identification

The identification of (3.3) hinges on the exclusion restriction that solicitations affect $Y_{j,t+1}$ only through their impact on review creation. Three institutional features support this restriction: (i) messages are sent *after* delivery, (ii) they contain no promotions or price information, and (iii) buyers cannot tell whether a posted review was solicited. Moreover, we separate the solicitation window (week t) from the outcome window (week $t+1$) and condition on the pre–week review state $S^{(k)}_{j,t-1}$ to account for baseline heterogeneity in information. Month fixed effects ensure that identification comes from within–month assignment of solicitations across orders. We omit product fixed effects in our preferred specification because the review–state indicators $S^{(k)}_{j,t-1}$ (and their interactions with the instrumented change in reviews) are highly persistent and largely product–specific over weekly horizons; including product fixed effects would absorb most cross–sectional variation in states and severely reduce power. Re-estimating (3.3) with product fixed effects yields the same qualitative patterns, larger effects at zero/low counts and below-median ratings, but with attenuated magnitudes and lower precision, consistent with limited within–product movement across states.

3.4 Link to counterfactual policy analysis

The two–stage design delivers the key sufficient statistics for policy: state–specific solicitation lifts $\widehat{\gamma}_k$ from Stage 1 and state–specific buyer–value parameters $\widehat{\delta}_k$ from Stage 2. In Section 5 we combine these objects to simulate counterfactual solicitation policies that raise the first–review response and reallocate solicitation effort toward high–value states, translating additional reviews into orders and returns using the estimated $\widehat{\delta}_k$. We then convert outcome changes into revenues considering both orders and returns.

4 Results

This section reports how post–delivery solicitations affect reviewer behavior (Stage 1) and how solicitation-induced reviews affect market outcomes (Stage 2). For ease of interpretation, we start with Stage 2 results and see if reviewer response in Stage 1 aligns with buyer value in Stage 2. We emphasize heterogeneity across the product's existing review state at the time of purchase.

4.1 Stage 2: Buyer value

Table 3 estimates the marginal effect of additional reviews on market outcomes, next-week orders and return rate. Here, the return rate is the number of returns divided by the number of orders over the corresponding 7 days. Two specifications per outcome are reported. The 2SLS specifications in models (1) and (3) use the *predicted* number of reviews in the past seven days constructed from Stage 1 and estimated by two-stage least squares, and the OLS specifications in models (2) and (4) use the *observed* number of reviews in the past seven days.

Table 3: Average marginal effect of reviews on orders and return rate

	$\log(\text{Orders}_{t+1})$		Return rate_{t+1}	
	2SLS	OLS	2SLS	OLS
Model:	(1)	(2)	(3)	(4)
# Reviews_t	0.1522^{***}	0.1240^{***}	-0.1054^{***}	-0.0863^{***}
	(0.0020)	(0.0022)	(0.0072)	(0.0063)
Month FE	Yes	Yes	Yes	Yes
N	22,708	22,708	22,036	22,036
R^2	0.45909	0.37044	0.01271	0.01175
Within R^2	0.42773	0.33394	0.00461	0.00364

Heteroskedasticity-robust standard errors in parentheses.
Return rate: no. returns / no. orders over the 7 days.
2SLS instruments reviews with solicitation-induced variation.
Signif.: $^{***}p < 0.01$, $^{**}p < 0.05$.

The 2SLS estimates indicate that additional reviews in the preceding week lead to higher orders in the next week of about 15% and lower return rates of about 0.105, after month fixed effects are included. The negative effect on return rate is consistent with the interpretation that more reviews reduce ex ante uncertainty and improve matching, a mechanism emphasized in prior work on reviews and returns (Sahoo et al., 2018). Although not reported here, the effects are robust to adding alternative controls such as price and product age. Heteroskedacity-robust standard errors are reported. The OLS estimates show qualitatively similar results.

Table 4 reports heterogeneous marginal effects of reviews by existing review state. We report 2SLS results in models (1) and (3) using predicted number of reviews in week t that interact with review-state indicators as of the beginning of t, and OLS results in models (2) and (4) that interact observed number of reviews in week t with the same indicators. Coefficients measure the incremental effect of one additional review during the past week on next-week outcomes, by the product's existing review state. Here, we define the existing review states by review counts and rating. Review count is partitioned into terciles of the empirical distribution: (0–43], [43–185), and [185+), which we call T1 count, T2 count and T3 count, respectively. Rating is split at the sample median, 4.85. Importantly, in our data the rating split doubles as a dispersion split: lower-than-median ratings coincide with much noisier information. For instance, the mean variance in the low-rating group is about 0.27 (versus about 0.12 in the high-rating group), and the median variance is roughly 0.25 in the low-rating group (versus roughly 0.15 in the high-rating group). We therefore interpret "low rating" as "higher-variance/noisier information" throughout. We use the same definition of review states throughout this paper.

Two patterns connect directly to the logic that prior information determines buyer value of a review. First, the marginal effect of an additional review on log orders declines with existing review volume within each rating group. Effects are largest when the information stock is low (0 review; T1 count and T2 counts) and attenuate as a product accumulates reviews.

Table 4: Marginal effect of reviews on orders and return rate: heterogeneity by existing review state

Dependent variables:	log(Orders$_{t+1}$)		Return rate$_{t+1}$	
	2SLS	OLS	2SLS	OLS
Model:	(1)	(2)	(3)	(4)
# Reviews$_t$ × I(0 reviews)	0.1557***	0.1922***	−0.2558***	−0.2297***
	(0.0100)	(0.0135)	(0.0440)	(0.0761)
# Reviews$_t$ × I(T1 count & Low rating)	0.2030***	0.1966***	−0.1885***	−0.2222***
	(0.0093)	(0.0056)	(0.0319)	(0.0344)
# Reviews$_t$ × I(T2 count & Low rating)	0.1764***	0.1485***	−0.1449***	−0.0935***
	(0.0054)	(0.0056)	(0.0236)	(0.0203)
# Reviews$_t$ × I(T3 count & Low rating)	0.1168***	0.0975***	−0.0785***	−0.0645***
	(0.0032)	(0.0031)	(0.0132)	(0.0113)
# Reviews$_t$ × I(T1 count & High rating)	0.1719***	0.1599***	−0.1173***	−0.1147***
	(0.0069)	(0.0051)	(0.0203)	(0.0218)
# Reviews$_t$ × I(T2 count & High rating)	0.1509***	0.1315***	−0.0440***	−0.0632***
	(0.0045)	(0.0044)	(0.0142)	(0.0137)
# Reviews$_t$ × I(T3 count & High rating)	0.1454***	0.1185***	−0.0914***	−0.0650***
	(0.0035)	(0.0044)	(0.0131)	(0.0118)
Review state dummies	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
N	22,708	22,708	22,036	22,036
R^2	0.47249	0.40213	0.01593	0.01497
Within R^2	0.44190	0.36747	0.00785	0.00689

All review states are those at the beginning of t.

T1/T2/T3 count bins refer to the first, second and third tertile of review counts; $(0, 43]$, $[43, 185)$, and $[185, +)$.

Low rating is < 4.85; high rating is $[4.85, 5.00]$.

Heteroskedasticity-robust standard errors in parentheses.

Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

T3 count states have the lowest effect, consistent with diminishing returns to information. Second, effects are larger when the average rating is below the median, which is a noisier state in our context, conditional on the review count. When signals are dispersed and reputation is not yet firmly established, a new review carries more information and buyers respond accordingly. These patterns hold for return rates as well and are even more pronounced: the largest reductions occur at the zero review state, followed by T1 count bin, and the effect is generally larger in the low rating (higher variance) group, with a small exception at T3. This asymmetry is natural because returns reflect misfit. Early reviews that clarify fit-relevant attributes, for example, sizing, material, and color fidelity, are effective to prevent ex post corrections.

4.2 Stage 1: Reviewer response

Table 5 presents average effects of receiving a solicitation on two reviewer responses: whether the buyer wrote any review and, conditional on writing, whether the posted rating was five stars. For clarity, *Solicited* is the main treatment variable that indicates that the order received the post-delivery SMS request to review. We also include indicators for the product's existing review state at the time of purchase and product characteristics such as whether the product is a new product (launched within the last 30 days), product age and price as controls, and month fixed effects.

Models (1) and (5) are the baseline specifications and are the focus, and the remaining columns serve as robustness checks. On average, a solicitation increases the probability that a buyer writes a review by about one percentage point ($p < 0.01$). Conditional on writing, the probability of giving a five-star rating declines by about 1.5 percentage points ($p < 0.01$). These results indicate that solicitations expand review supply by drawing in marginal reviewers whose ratings are, on average, slightly less likely to be a perfect five. Adding controls leaves the magnitudes essentially unchanged, suggesting that the main effect of solicitations

Table 5: Average effect of solicitation on review writing and rating

Dependent variables:	Review dummy				Five-star dummy (conditional on review)			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solicited	0.0103***	0.0104***	0.0103***	0.0105***	−0.0152***	−0.0157***	−0.0156***	−0.0160***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0057)	(0.0057)	(0.0057)	(0.0057)
New product dummy			0.0001	0.0017			−0.0023	−0.0031
Product age (days)			6.85×10^{-6}	-2.69×10^{-5}			-5.6×10^{-5}	−0.0002***
Product age^2			4.14×10^{-9}	3.18×10^{-8}			7.5×10^{-9}	8.31×10^{-8}
Price			-1.9×10^{-5}	1.4×10^{-5}			−0.0003	−0.0002
Review state dummies	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	153,429	153,429	153,429	153,429	13,949	13,949	13,949	13,949
R^2	0.00791	0.00808	0.00797	0.00811	0.00236	0.00414	0.00372	0.00632

All review states are those at the time of purchase.
T1/T2/T3 count bins are $(0, 43]$, $[43, 185)$, and $[185, +)$.
Low rating is < 4.85; high rating is $[4.85, 5.00]$.
For review state controls, the baseline level is I(T1 count & Low rating).
Heteroskedasticity-robust standard errors in parentheses.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

is not driven by correlated changes in new product status, product age, or price.

Table 6 examines whether reviewer response varies with the existing review states, which represent product information already available to buyers. As before, we focus on models (1) and (4) and the remaining models are robustness checks. When a product has zero prior reviews, the estimated lift in review likelihood is small and statistically insignificant across specifications, indicating a first review barrier that solicitations rarely generate the very first review even though early reviews have high informational value to buyers. Once a small base of information exists, e.g., T1 count for both rating groups, solicitations are markedly more effective. For low rating products, the largest increases occur in the T2 count bin between 43 and 185. Coefficients for review incidence are around 1.7 percentage points in these states. For products with high ratings, the increases are more even across count bins, suggesting that where reputation is already very established (T3 count), an additional solicitation induces reviews at a broadly similar rate as at lower counts (T1 and T2).

Conditional on writing, the decline in the five-star probability is statistically significant but modest when the reviews are somewhat established (T1 count in low rating, and T2 count in high rating). The magnitude is small, considering already established review counts and ratings for those products, indicating that these composition shifts are unlikely to move product-level average ratings in a meaningful way. This pattern helps interpret our earlier Stage-2 results about buyer value: the economically important margin for market outcomes mainly comes from the increase in the number of reviews rather than shifts in average valence.

Table 6: Effect of solicitations on review writing and rating: heterogeneity by existing review state

Dependent variable	Review dummy			Five-star dummy (conditional on review)		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Solicited × I(0 review)	0.0046	0.0045	0.0045	−0.0387	−0.0384	−0.0390
	(0.0087)	(0.0087)	(0.0087)	(0.0300)	(0.0300)	(0.0300)
Solicited × I(T1 count & Low rating)	0.0112**	0.0115**	0.0115**	−0.0388	−0.0378	−0.0376
	(0.0056)	(0.0056)	(0.0056)	(0.0201)	(0.0201)	(0.0201)
Solicited × I(T2 count & Low rating)	0.0173***	0.0173***	0.0173***	−0.0230	−0.0228	−0.0229
	(0.0047)	(0.0047)	(0.0047)	(0.0171)	(0.0171)	(0.0171)
Solicited × I(T3 count & Low rating)	0.0045	0.0045	0.0045	0.0087	0.0105	0.0103
	(0.0045)	(0.0045)	(0.0045)	(0.0166)	(0.0166)	(0.0166)
Solicited × I(T1 count & High rating)	0.0111***	0.0112***	0.0112***	−0.0101	−0.0106	−0.0107
	(0.0042)	(0.0042)	(0.0042)	(0.0134)	(0.0134)	(0.0134)
Solicited × I(T2 count & High rating)	0.0077**	0.0077**	0.0077**	−0.0337**	−0.0344**	−0.0346**
	(0.0037)	(0.0037)	(0.0037)	(0.0137)	(0.0137)	(0.0137)
Solicited × I(T3 count & High rating)	0.0122***	0.0122***	0.0122***	−0.0055	−0.0069	−0.0069
	(0.0033)	(0.0033)	(0.0033)	(0.0096)	(0.0096)	(0.0096)
New product		0.0017	0.0017		−0.0027	−0.0036
Product age (days)		-2.67×10^{-5}	-2.70×10^{-5}		−0.0002***	−0.0001***
Product age^2		3.18×10^{-8}	3.22×10^{-8}		8.68×10^{-8}	7.91×10^{-8}
Price			1.29×10^{-5}			−0.0002
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	153,429	153,429	153,429	13,949	13,949	13,949
R^2	0.00811	0.00814	0.00814	0.00473	0.00682	0.00690

All review states are those at the time of purchase.
T1/T2/T3 count bins are $(0, 43]$, $[43, 185)$, and $[185, +)$.
Low rating is < 4.85; high rating is $[4.85, 5.00]$.
Heteroskedasticity-robust standard errors in parentheses.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

4.3 Reconciling reviewer response and buyer value

We examine whether reviewer response to solicitation occurs in the product states where buyers value additional information most. Figure 7 juxtaposes instrumented, state-contingent effects of one more review on next-week orders and return rates from Table 4, models 1 and 3 (top panels) with solicitation effects on review incidence and the five-star probability from Table 6, models 1 and 4 (bottom panels). The top panels split products by rating (low: < 4.85; high: $[4.85, 5]$) and by review count bins (0, T1, T2, T3). An additional review delivers the largest gains when information is scarce or noisy. Positive effects on orders and negative effects on return rates are strongest at zero and low counts, especially in the low rating (higher variance) group. This pattern is consistent with an information-based mechanism: early reviews reduce uncertainty about fit and quality, whereas marginal informativeness diminishes once many similar signals exist.

The bottom panels plot solicitation effects. At zero prior reviews, the lift in review incidence is close to zero and statistically indistinguishable from zero, indicating a first review barrier. Once a small base exists at the T1 and T2 states, solicitations are more effective, with only modest declines in the five-star share. Thus, misalignment arises because reviewer response is weakest exactly where buyers derive the highest marginal value from another review.

Prior work often attributes the scarcity of initial reviews to low traffic and cold-start dynamics. Fewer arrivals mean fewer opportunities to post, reinforcing "rich-get-richer" exposure effects (Moe and Trusov, 2011; Moe and Schweidel, 2012). In our setting, however, solicitation effects are estimated at the order level and leverage exogenous variation in whether a *delivered order* receives a solicitation, so the pool of potential reviewers is held fixed and we effectively control for the arrival rate of buyers. The insignificant and small solicitation effect at the zero review state therefore cannot be attributed to insufficient traffic. Instead, the evidence points to psychological and strategic frictions in initiating review information.

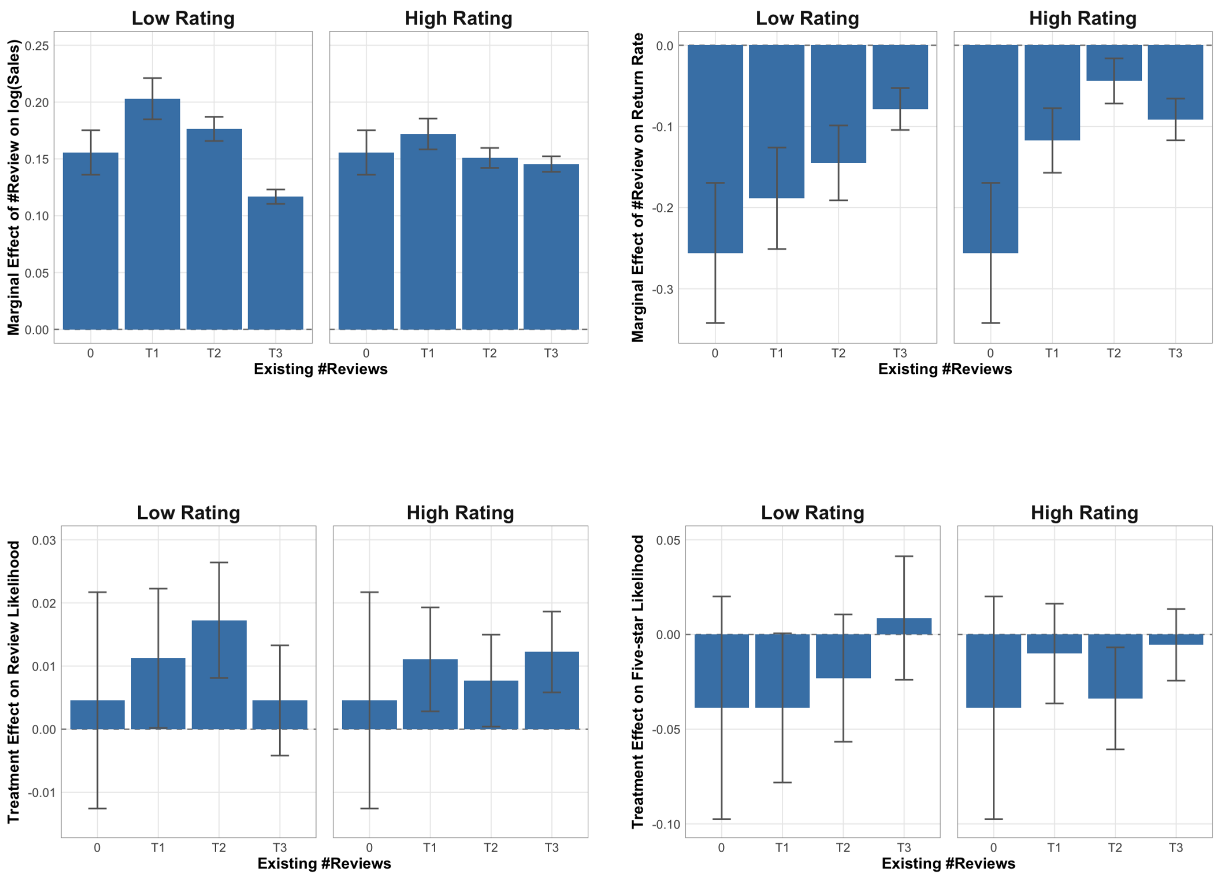


Figure 7: Buyer value (top) and reviewer response (bottom) by existing review state.

Note.
1) Top panels: instrumented, state-contingent marginal effects of additional reviews on log orders (left two subplots) and on next-week return rate (right two subplots), shown separately by rating group (Low: < 4.85, High: $[4.85, 5]$]) and by review-count bins (T1: $(0, 43]$, T2: $[43, 185)$, and T3: $[185, +)$)
2) Bottom panels: solicitation effects on review incidence (left two subplots) and on the probability of a five-star rating conditional on reviewing (right two subplots)
3) Error bars show 95% confidence intervals.

For example, the absence of established review base may raise the perceived cost of "setting the tone". Also, without reference points, writers may face higher cognitive effort to decide what content is useful. These forces lead each potential reviewer to prefer to post once others have done so. This produces a threshold that solicitations, which primarily lower effort, do not overcome at the zero review states. Only once minimal social evidence is present, solicitations become effective, reconciling the strong buyer value of early reviews with weak reviewer response at the very start.

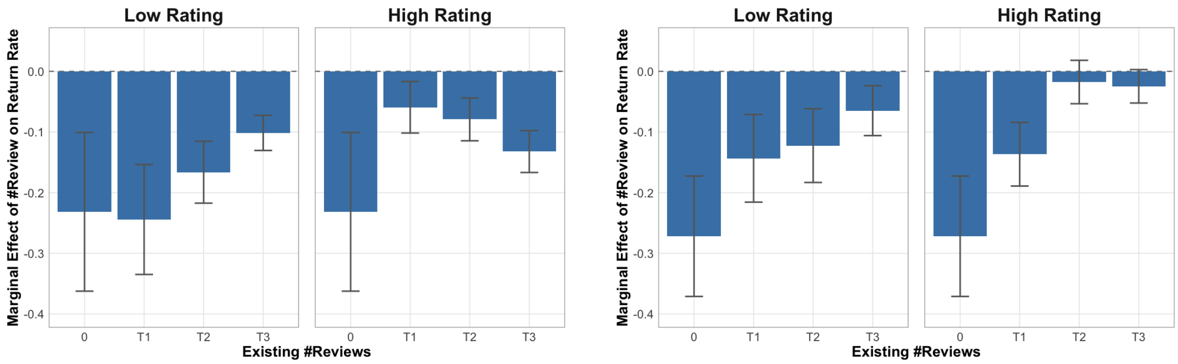
Larger buyer value for new buyers. The reviewer–buyer misalignment is especially consequential when a product is attracting new customers. Table 7 reports estimates from the Stage 2 2SLS specification after splitting outcomes by buyer type, where *existing* buyers do have purchase history before the focal order and *new* buyers do not. Models (1)–(3) show effects on log orders, and models (4)–(6) show effects on return rate, for all customers (replicated from Table 4 models (1) and (3)) in models (1) and (4), for existing buyers in models (2) and (5) and for new buyers in models (3) and (6). On the sales margin, the marginal value of an additional review is broadly similar for existing and new buyers across review states, indicating that incremental reviews increase purchase propensity for both groups. The returns margin tells a different story. Early in the life of a product, one more review reduces new customers' return rates more than existing customers'. At zero prior reviews the return-rate effect is 0.272 for new customers versus 0.232 for existing customers; at T1 counts with high average rating (the lower-variance state), the effect is 0.137 for new customers versus 0.059 for existing customers. Figure 8 visualizes these patterns in return rate: compared to panel (a) plotting effects on existing customers, the panel (b) shows larger negative bars at the zero review state, and at the T1 state when the existing rating is high. Our interpretation is that new customers lack product-specific and brand-specific experience, so the first few reviews supply fit-relevant information that prevents mispurchases especially when the overall signal (high rating/low variance) looks favorable. Existing customers, by contrast, benefit most when the information environment is noisier (lower rating/higher variance) across review count bins, where additional reviews help them discriminate among close substitutes and reduce returns.

Table 7: State-contingent effect of reviews on orders and return rate by buyer type

Dependent variable	log(Orders$_{t+1}$)			Return rate$_{t+1}$		
Model:	All customers (1)	Existing customers (2)	New customers (3)	All customers (4)	Existing customers (5)	New customers (6)
# Reviews × I(0 review)	0.1557***	0.1465***	0.1476***	−0.2558***	−0.2315***	−0.2719***
	(0.0100)	(0.0094)	(0.0090)	(0.0440)	(0.0667)	(0.0506)
# Reviews × I(T1 count & Low rating)	0.2030***	0.1933***	0.1878***	−0.1885***	−0.2443***	−0.1434***
	(0.0093)	(0.0087)	(0.0085)	(0.0319)	(0.0462)	(0.0369)
# Reviews × I(T2 count & Low rating)	0.1764***	0.1693***	0.1654***	−0.1449***	−0.1663***	−0.1224***
	(0.0054)	(0.0052)	(0.0050)	(0.0236)	(0.0260)	(0.0310)
# Reviews × I(T3 count & Low rating)	0.1168***	0.1116***	0.1151***	−0.0785***	−0.1015***	−0.0648***
	(0.0032)	(0.0030)	(0.0032)	(0.0132)	(0.0148)	(0.0210)
# Reviews × I(T1 count & High rating)	0.1719***	0.1680***	0.1566***	−0.1173***	−0.0592***	−0.1366***
	(0.0069)	(0.0068)	(0.0061)	(0.0203)	(0.0217)	(0.0268)
# Reviews × I(T2 count & High rating)	0.1509***	0.1469***	0.1427***	−0.0440***	−0.0792***	−0.0176
	(0.0045)	(0.0043)	(0.0042)	(0.0142)	(0.0180)	(0.0182)
# Reviews × I(T3 count & High rating)	0.1454***	0.1406***	0.1374***	−0.0914***	−0.1322***	−0.0247
	(0.0035)	(0.0033)	(0.0032)	(0.0131)	(0.0176)	(0.0140)
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	22,708	22,708	22,708	22,036	21,339	21,391
R^2	0.47249	0.47271	0.48421	0.01593	0.01367	0.01510
Within R^2	0.44190	0.44180	0.44880	0.00785	0.00911	0.00496

Heteroskedasticity-robust standard errors in parentheses.
T1/T2/T3 count bins are $(0, 43]$, $[43, 185)$, and $[185, +)$.
Low rating is < 4.85; high rating is $[4.85, 5.00]$.
"New" customers have no prior purchase before the focal order.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

From a managerial perspective, these results imply that the revenue at risk from the reviewer–buyer misalignment at the zero review state could be concentrated on products that bring in new customers: the products for which early reviews are hardest to generate are also the products where one more review saves the most in avoidable return costs. Targeting early-review generation at products with a high share of first-time buyers, especially before crossing the first-review and low-count thresholds, should yield disproportionate pay-



(a) Existing buyers

(b) New buyers

Figure 8: Marginal effect of reviews on return rate by buyer type

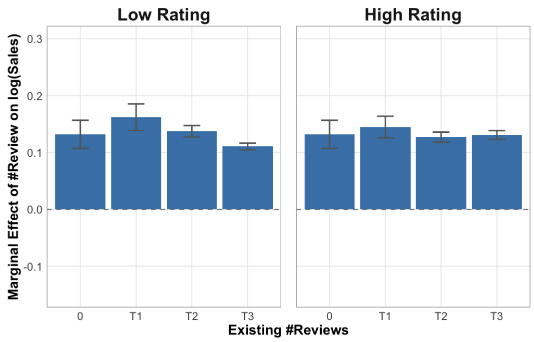
Note.
1) For each buyer type, left: rating <4.85, higher variance; right: rating between $[4.85, 5]$, lower variance.
2) Bars show coefficients for "# Reviews × review state"; whiskers are 95% confidence intervals.

offs. Because solicitation alone underperforms at the zero review state, managers may need complementary, transparent levers that overcome the first-review barrier and are likely to smooth onboarding of new customers.

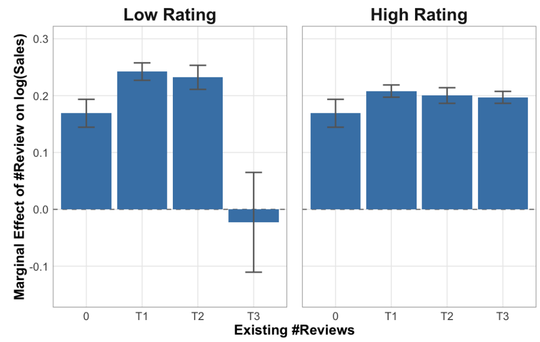
4.4 Reviewer-buyer (mis)alignment by product prices

Next, we examine whether the reviewer–buyer (mis)alignment varies with product price. Price matters for two reasons: (i) stakes are higher at higher prices, scaling both the amount a customer must pay and the economic loss from a mismatch, so the marginal value of pre-purchase information should be greater for expensive goods for both conversion (orders) and post-purchase outcomes (returns), and (ii) price co-varies with category and sales velocity in our setting as in Figure 5, so a given review count can imply different informational richness across price points. Estimating price-contingent effects therefore identifies where misalignment is most costly for the firm.

We split products at the median price and re-estimate the Stage 1 and Stage 2 models within each stratum, keeping the same review-state bins (0, T1: (0–43], T2: [43–185),

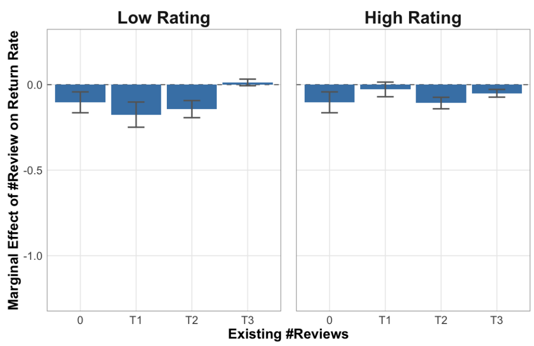


(a) Low Price

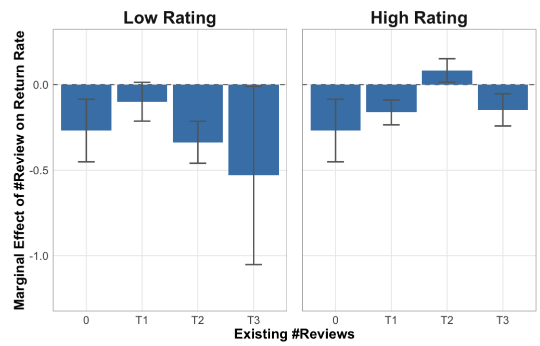


(b) High Price

Figure 9: Marginal effect of reviews on sales by price

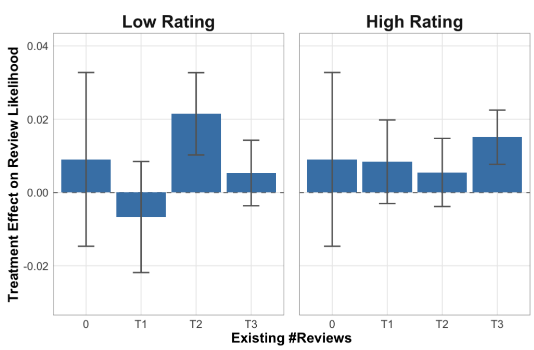


(a) Low Price

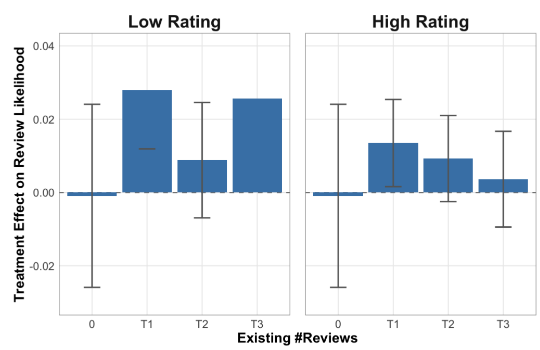


(b) High Price

Figure 10: Marginal effect of reviews on return rate by price



(a) Low Price



(b) High Price

Figure 11: Solicitation effect on review likelihood by price

T3: [185+)) and the same median rating split (4.85). As before, the low-rating bin also corresponds to higher variance in realized ratings.

On the buyer side, Figures 9 and 10 report instrumented, state-contingent effects of one additional review on next-week orders and return rates, respectively. In each figure, panel (a) presents low-priced items, and panel (b) presents high-priced items. Within each panel, the left subgraph shows the low rating (higher variance) group and the right subgraph shows the high rating (lower variance) group, with bars for the 0, T1, T2, and T3 review count bins.

In Figure 9 (orders), effects are generally larger for high-priced items, including at the zero review state. Figure 10 (returns) shows an even sharper contrast: return reductions are larger for high-priced products overall, with especially large differences at zero reviews.[8] Taken together, the within-stratum shapes replicate our baseline pattern, larger effects in noisier, low-information states with diminishing returns as reviews accumulate, while magnitudes are consistently larger for high-priced products. This amplification implies that early reviews deliver disproportionately high buyer value for expensive goods, improving conversion where appropriate and reducing costly mismatches.

Reviewer response is plotted in Figure 11. Solicitations do not reliably produce the very first review in either price stratum: effects at zero reviews are small and statistically indistinguishable from zero, so the first-review barrier persists regardless of price. Conditional on having some reviews, solicitation lifts are generally larger for high-priced products, most notably at T1, whereas for low-priced products the lift concentrates at higher count bins. These pattern shows that solicitations lower effort broadly, but overcoming the first review barrier requires a perceived "established" base of reviews that reduces evaluation apprehension and clarifies posting norms. Because price co-varies with category and sales velocity, the same re-

[8]In the extreme case of high price, low rating, and T3 counts, sales decrease (Figure 9 (b), left subgraph) while returns fall sharply (Figure 10 (b), left subgraph), indicating that additional reviews operate as a screening device, deterring mismatched purchases of poorly rated, expensive products.

view count conveys different informational richness: in slow moving, higher-priced categories a small number of reviews may already be viewed as a normal, adequately informative base that could lower the psychological cost of writing, whereas in fast-moving, lower-priced categories the same count can appear thin and insufficient to trigger participation. Consistent with earlier sections, shifts in the five-star share are modest and statistically weak within both strata, indicating that solicitation effects operate primarily on the extensive margin.

To sum up, the price split amplifies the reviewer–buyer misalignment documented earlier. High-priced products are precisely where an extra review delivers the largest buyer value (especially in low-count, low-rating states), yet solicitations are least effective at generating the first review. Once a minimal base exists, however, high-priced products respond at least as strongly as low-priced products, suggesting that policies targeting the transition out of the zero review state for expensive items should yield outsized reductions in avoidable returns and improved conversion.

5 Quantifying Missed Revenue

This section uses a counterfactual to translate the reviewer–buyer misalignment into dollars. The goal is to approximate how much net revenue the firm misses because solicitations seldom generate the very first review, even though early reviews have the highest buyer value. To emphasize the importance of reducing reviewer-buyer gap at the zero review state, we vary only the first–stage solicitation effect for products at zero reviews and feed the implied review supply into the estimated Stage 2 outcome model to re-predict orders and return rates. We then convert those changes into net revenue per product and aggregate across time. Let γ denote a hypothetical improvement in the first–stage effect of solicitation on the probability of producing the *first* review (i.e., for the zero review state). For each γ in a prespecified grid that spans 0 to 0.02 which includes the realized heterogeneous treatment effect (HTE) at zero reviews (0.0046), the sample average treatment effect (0.0103), and the upper tail of the HTEs we observe (0.0173), we construct counterfactual reviews and propagate them through Stage 2 to obtain counterfactual orders and return rates. As a starting point, the exercise is *static*: we do not model feedback from added reviews to subsequent reviewer responsiveness nor do we update the rating distribution. Here we ask, *holding all else fixed*, how much additional net revenue arises if solicitations were more effective exactly at the zero-to-one margin. This is conservative with respect to dynamic benefits from faster review accumulation and aligns with the "local" allocation question of interest to managers. The simulation procedure is in the appendix.[9]

Figure 12 plots per-product changes in the y-axis as we move the zero-review first–stage effect γ from 0 to 0.02 in the x-axis. Here, $\gamma = 0.02$ means that solicitation increases the first review likelihood by 2 percentage point. The top panel shows the change in net revenue per product, and the bottom panel shows the change in the number of reviews written per product. Dashed vertical markers indicate three salient benchmarks from the in-sample estimation in Stage 2: the realized HTE for no review products (about 0.0046), the sample ATE (0.0103), and the highest HTE observed in our data (0.0173). The curves are monotonically increasing. Increasing the solicitation effectiveness for the first review from 0 to 0.02 increases per-product annual net revenue by 120 USD on average. The distinctly right-skewed pattern reveals substantial heterogeneity in where incremental alignment pays off. Moving from the realized HTE to the ATE produces noticeable revenue gains for a broad subset of products; pushing to the upper tail yields sharply larger gains concentrated among products whose buyer value is particularly sensitive at the no review state.[10]

The bottom panel makes clear why the top panel is increasing and convex in practice. As γ increases from 0 to 0.02, more first reviews are written for products in the cold-start

[9] We plan to implement the dynamic simulation as the next step.

[10] One outlier product with extremely sparse review arrival is excluded to generate this figure to preserve readability of the per-product curves.

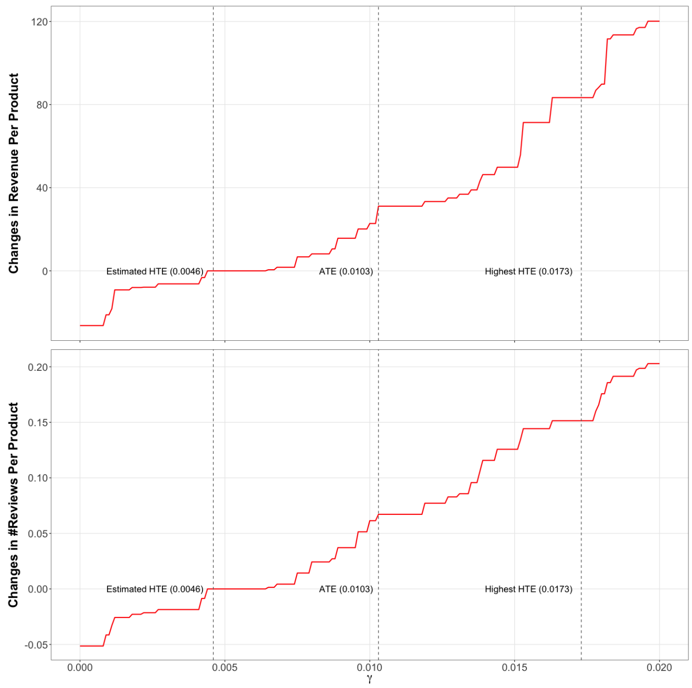


Figure 12: Changes in Net Revenue (top) and Reviews (bottom) under Hypothetical Solicitation Effect

state by 0.2 reviews per product on average. Even small increases in the probability of a first review at those times translate into large net-revenue lifts because Stage 2 returns show that the marginal effect of *one* review is largest on return rates when information is scarcest and noisiest.

This static exercise converts the reviewer–buyer gap into dollars and shows that relatively small improvements in the solicitation effect for products at the no review states can generate economically meaningful net-revenue gains. The managerial implication is to target alignment investments where the Stage 2 buyer value is highest and exposure to the cold-start state is non-trivial. In short, aligning review supply for the first review can reduce avoidable return costs and increase contribution, and the static curves provide a transparent way to quantify and prioritize those gains.

6 Conclusion

In this paper, we ask whether review solicitations allocate review supply to the product states where buyers derive the greatest marginal value. We find that an additional solicitation-induced review raises subsequent orders on the order of 13–16% and lowers next-period return rates by roughly two percentage points. These effects are highly state dependent: the marginal value of a review is largest when information is scarce (zero-to-few prior reviews) and noisier (below-median average rating, which in our setting coincides with higher variance). The asymmetry is most pronounced for returns, consistent with an information-and-matching channel in which early reviews clarify size, material, and usage expectations, thereby preventing ex post returns. Most importantly, reviewer response is not fully aligned with this buyer value. Solicitation lifts review incidence overall, but rarely overcomes the first review barrier. The strongest responses occur only once a minimal base of reviews exists.

These results imply that the value of solicitations comes not from generating as many reviews as possible, but from *where* those reviews are created. The counterfactual analysis to quantify the missed revenue makes this point concrete. Increasing the reviewer response to solicitation for no-review products from the observed heterogeneous treatment effect (about 0.0046) toward the sample average (about 0.0103) and then toward the best observed lift in the data (about 0.0173) produces gains in net revenue per product. Even modest improvements at the zero-review margin translate into economically meaningful increases in steady-state net revenue because they both expand demand and reduce avoidable returns. The exercise is deliberately static: it does not rely on assumptions about the dynamic evolution of reviewing once early reviews arrive, which makes the gains transparent for budgeting. This exercise also highlights that firms may have strong incentives to self-correct informational biases to reduce costly returns - a countervailing force against incentives to inflate or exacerbate bias, for example through fake reviews.

Industry practice is consistent with our interpretation that the first review carries outsized informational value but is costly to elicit. Major retailers pilot targeted early-review programs (e.g., Home Depot's Seeds and Best Buy's Tech Insider), which explicitly subsidize initial reviews for new or pre-release products. These programs are designed to cross the first-review barrier we document and to seed fit-relevant content for prospective buyers.

Managerially, three implications follow. First, targeting should be *state contingent*. Firms should consider prioritizing solicitation budgets for products at zero-to-few reviews and below-median ratings, where marginal buyer value is highest and return reductions are largest. Second, levers should be designed to clear the first review barrier. The most prevalent uniform, low-touch messages are unlikely to suffice. The alternatives could be time-limited nudges for the very first contributors, and structured prompts that request evaluations on specific attributes. Third, success metrics should be market outcomes rather than review counts alone. Because early reviews reduce misfit, measuring lift in orders jointly with reductions in return rates provides a more reliable objective for firms.

Our study has limits that bound generalization. We analyze one seller and category with post-delivery, unincentivized SMS solicitations. Categories with long usage lags, durable goods, or platforms that heavily adjust search and merchandising in response to reviews may exhibit different patterns.

Taken together, the evidence reframes review solicitation as a review-allocation problem. Buyers benefit most from early, informative reviews, yet those reviews are precisely the hardest to elicit with uniform solicitations. Reallocating effort toward low-review states, explicitly managing the zero-to-one transition, and evaluating policies on orders and returns, not review counts alone, can improve both demand and fit.

References

Anderson, E. T., Hansen, K., and Simester, D. (2009). The option value of returns: Theory and empirical evidence. *Marketing Science*, 28(3):405–423.

Anderson, M. and Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122(563):957–989.

Askalidis, G., Kim, S. J., and Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, 97:23–30.

Brandes, L., Godes, D., and Mayzlin, D. (2022). Extremity bias in online reviews: The role of attrition. *Journal of Marketing Research*, 59(4):675–695.

Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64(5):2065–2082.

Cabral, L. and Hortaçsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Chakraborty, I., Kim, M., and Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, 59(3):600–622.

Chevalier, J. A., Dover, Y., and Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, 37(5):688–709.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing science*, 29(5):944–957.

Dellarocas, C. and Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management science*, 54(3):460–476.

Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure. *Information Systems Research*, 19(3):291–313.

Fradkin, A. and Holtz, D. (2023). Do incentives to review help the market? evidence from a field experiment on airbnb. *Marketing Science*, 42(5):853–865.

Gandhi, A., Hollenbeck, B., and Li, Z. (2025). Misinformation and mistrust: The equilibrium effects of fake reviews on amazon.com. *Available at SSRN 5404060*.

Gao, B., Wang, J., Ding, X., and Guo, Y. (2025). The pitfalls of review solicitation: Evidence from a natural experiment on tripadvisor. *Management Science*, 71(2):1671–1691.

Godes, D. and Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing science*, 28(4):721–739.

Godes, D. and Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Science*, 31(3):448–473.

Guadagni, P. M. and Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3):203–238.

Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal* *of Marketing Research*, pages 342–355.

He, S., Hollenbeck, B., and Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Hu, N., Pavlou, P. A., and Zhang, J. (2009). Overcoming the j-shaped distribution of product reviews. *Communications of the ACM*, 52(10):144–147.

Hu, N., Pavlou, P. A., and Zhang, J. (2017). On self-selection biases in online product reviews. *MIS quarterly*, 41(2):449–475.

Janakiraman, R., Syrdal, H. A., and Freling, T. H. (2016). The effect of return policy on consumer purchase and return decisions: A meta-analytic review. *Journal of Retailing*, 92(2):226–235.

Karaman, H. (2021). Online review solicitations reduce extremity bias in online review distributions and increase their representativeness. *Management Science*, 67(7):4420–4445.

Khern-am nuai, W., Kannan, K., and Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research*, 29(4):871–892.

Li, X. and Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information systems research*, 19(4):456–474.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp.com. *Harvard Business School Working Paper*, 12-016.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management science*, 62(12):3412–3427.

Mayzlin, D., Dover, Y., and Chevalier, J. A. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455.

Moe, W. W. and Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3):372–386.

Moe, W. W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of marketing research*, 48(3):444–456.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651.

Paget, S. (2023). Local consumer review survey 2023. Accessed: November 7, 2025, https://www.brightlocal.com/research/local-consumer-review-survey.

Park, S., Shin, W., and Xie, J. (2021). The fateful first consumer review. *Marketing Science*, 40(3):481–507.

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5):645–665.

Reimers, I. and Waldfogel, J. (2021). Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings. *American Economic Review*, 111(6):1944–1971.

Sahoo, N., Dellarocas, C., and Srinivasan, S. (2018). The impact of online product reviews on product returns. *Information Systems Research*, 29(3):723–738.

Schoenmueller, V., Netzer, O., and Stahl, F. (2020). The polarity of online reviews: Preva-

lence, drivers and implications. *Journal of Marketing Research*, 57(5):853–877.
Shehu, E., Papies, D., and Neslin, S. A. (2020). Free shipping promotions and product returns. *Journal of Marketing Research*, 57(4):640–658.
Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4):696–707.
Sun, Y., Dong, X., and McIntyre, S. (2017). Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards. *Marketing Science*, 36(3):329–337.
Sunada, T. (2025). Why do product prices affect review ratings? evidence from a japanese online travel agency. *Forthcoming: Management Science*.
Wang, Y. and Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, 55(2):163–177.
Wood, S. L. (2001). Remote purchase environments: The influence of return policy leniency on two-stage decision processes. *Journal of Marketing Research*, 38(2):157–169.
Woolley, K. and Sharif, M. A. (2021). Incentives increase relative positivity of review content and enjoyment of review writing. *Journal of Marketing Research*, 58(3):539–558.
You, Y., Vadakkepatt, G. G., and Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 79(2):19–39.
Zhao, Y., Yang, S., Narayan, V., and Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing science*, 32(1):153–169.

Online Appendix

A Sample Data Construction

A.1 Focal products

Although over 3,000 products were listed on the retailer's website at some point in 2019, most sold infrequently, with the majority of revenue coming from a much smaller subset. To focus our product-day analyses on economically meaningful and reasonably active products, we first rank products in descending order by (i) total 2019 revenue and (ii) daily revenue, selecting the top 300 from each ranking. Taking the intersection of these two groups yields 173 products. We then exclude three products with an extreme number of reviews ($\geq$1,200) to mitigate potential outlier bias, leaving 170 focal products for our analyses. These products are the most relevant for the retailer's managerial decisions.

A.2 Price construction

Our data records the price paid by consumers, after accounting for retailer promotions (e.g., discounts, coupons) and the use of store credits or rewards, which represent consumers' own resources rather than retailer-provided discounts. Consequently, leveraging product-day variation (especially with product fixed effects) makes the interpretation of price coefficients or elasticities ambiguous and potentially misleading. Second, descriptive statistics show minimal within-product price variation over time. Third, we examine heterogeneity in the effect of solicitations with respect to price by interacting it with the review solicitation treatment. In this approach, treating price as a time-invariant product characteristic captures the financial risk associated with potential product mismatch. Given how the paid price is recorded in our data, using product-level price offers a more stable and conceptually appropriate measure for analyzing price-related heterogeneity

Based on these considerations, we construct a time-invariant product-level price for our analyses as follows. We first compute each product's daily average unit price by averaging across all orders on that day. We then calculate the product's time-invariant price as a quantity-weighted average of these daily prices over the sample period, weighting each day by total units sold. This weighting ensures that the measure reflects the price points at which most transactions occurred, giving greater weight to high-volume days.

B Simulation Procedure

For each product j and day t in 2019:

(i) Baseline predictions. Using the estimated Stage 1 parameters, we predict the number of reviews during the past seven days, $\widehat{\text{Reviews}}_{j,t}^{\text{base}}$, from Stage 1; we then use Stage 2 to obtain baseline predictions for next–week orders and return rates:

$$\widehat{\ln(\text{Orders}_{j,t+1}+1)}^{\text{base}} \quad \text{and} \quad \widehat{\text{ReturnRate}}_{j,t+1}^{\text{base}}.$$

Because the orders equation is estimated on $\ln(\text{Orders}+1)$, we transform it to the number of orders and convert to a daily frequency:

$$\widehat{\text{Orders}}_{j,t}^{\text{base}} = \frac{1}{7}\left\{\exp\left(\widehat{\ln(\text{Orders}_{j,t+1}+1)}^{\text{base}}\right)-1\right\}, \qquad \widehat{\text{Returns}}_{j,t}^{\text{base}} = \widehat{\text{Orders}}_{j,t}^{\text{base}} \cdot \widehat{\text{ReturnRate}}_{j,t+1}^{\text{base}}.$$

Given unit price for each product p_j, baseline net revenue is

$$\pi_{j,t}^{\text{base}} = p_j\left(\widehat{\text{Orders}}_{j,t}^{\text{base}} - \widehat{\text{Returns}}_{j,t}^{\text{base}}\right).$$

(ii) Counterfactual predictions. For a candidate x, we replace the first–stage coefficient for the zero-review state by x while keeping all other first–stage parameters intact, re-predict past–week reviews $\widehat{\text{Reviews}}^{\text{cf}}_{j,t}(x)$, and feed them into Stage 2 to obtain

$$\widehat{\ln(\text{Orders}_{j,t+1}+1)}^{\text{cf}}(x), \qquad \widehat{\text{ReturnRate}}^{\text{cf}}_{j,t+1}(x),$$

which we back-transform and scale to daily values:

$$\widehat{\text{Orders}}^{\text{cf}}_{j,t}(x) \;=\; \frac{1}{7}\Big\{\exp\!\big(\widehat{\ln(\text{Orders}_{j,t+1}+1)}^{\text{cf}}(x)\big)-1\Big\}, \quad \widehat{\text{Returns}}^{\text{cf}}_{j,t}(x) \;=\; \widehat{\text{Orders}}^{\text{cf}}_{j,t}(x)\cdot\widehat{\text{ReturnRate}}^{\text{cf}}_{j,t+1}(x).$$

Counterfactual daily net revenue is

$$\pi^{\text{cf}}_{j,t}(x) \;=\; p_j\left(\widehat{\text{Orders}}^{\text{cf}}_{j,t}(x) - \widehat{\text{Returns}}^{\text{cf}}_{j,t}(x)\right).$$

(iii) Aggregation. We aggregate to annual firm-level profit by summing across products and days and compute per-product net revenue:

$$\Pi^{\text{base}} \;=\; \sum_{j,t}\pi^{\text{base}}_{j,t}, \qquad \Pi^{\text{cf}}(x) \;=\; \sum_{j,t}\pi^{\text{cf}}_{j,t}(x), \qquad \Delta\Pi(x) \;=\; \Pi^{\text{cf}}(x)-\Pi^{\text{base}}.$$

For the policy question that exclusively targets the zero-review margin, it suffices to aggregate over product–days that are at #reviews$=0$.