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

Hana Choi* Minkyung Kim† Jinsoul Seo‡

December 2, 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, Virtual Quant Marketing Seminar, and Business and Marketing Analytics 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, because they target reviewers while value accrues to buyers. Using rich individual-level panel data on transactions, reviews, and returns from an 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 reviews influence subsequent sales and returns. In Stage 1, we find that solicitations increase the likelihood of writing a review, but this response is highly state-dependent: effects are negligible for products with no prior reviews. In Stage 2, we show that reviews increase next-week sales reduce return rates the most when prior reviews are scarce. Taken together, solicitations fail to elicit reviews in high-value states and thereby creating a reviewer–buyer misalignment. This misalignment implies that reviewers do not fully internalize buyer value when deciding whether to write the first review. We interpret the results as evidence of psychological frictions in initiating reviews that solicitations do not overcome. Counterfactual simulations indicate that raising first-review likelihood by two percentage points increases net revenue by $4.88 per solicitation message, highlighting the value of eliciting reviews where they matter for sales and returns.

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

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 service, inspection, restocking, and, in the case of apparel, frequent liquidation. The macro

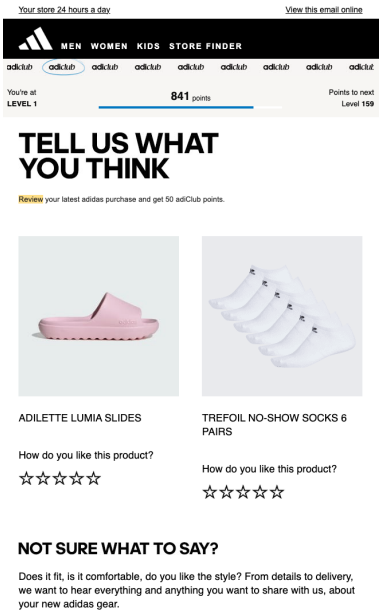


Figure 1: Example of review solicitations

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 variation in which customers receive review solicitations, shifting review supply indepen-

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

dently 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 reviewer responses in Stage 1. Solicitations expand review supply, increasing the likelihood of writing a review by about 1.1 percentage point. The conditional probability of a five-star rating declines modestly (by 1.6 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. 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. Importantly, because our Stage 1 estimations are conducted at the order level and the estimates are robust with respect to controls, we show that the heterogeneous effects are not mechanical artifacts of sales velocity or product mix.

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, 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, 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.

A closer look at how effects vary with existing review states, juxtaposing reviewer response with buyer value, reveals a systematic reviewer–buyer misalignment. In precisely those states where an additional review has the largest causal impact on sales and returns—zero and low counts, especially for lower-rated, higher-variance products—solicitations barely move review incidence. We interpret the misalignment to mean that reviewers do not fully internalize buyer value: if they did, we would expect the first review to be most easily generated where

it yields the largest downstream gains. To sharpen our point, we show the misalignment is most pronounced for high-priced products, where early reviews deliver disproportionately large gains in orders and reductions in returns yet remain hardest to elicit. Instead, our evidence points to a psychological barrier to initiating the very first review, such as evaluation apprehension and the salience of being "the first", that the current solicitations fail to overcome. Overall, solicitations primarily expand review volume on products that have already cleared an initial information threshold, rather than in the 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 a net revenue increase of $4.88 per solicitation, driven by the high marginal value of reviews when information is scarce. 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.

These results connect to and extend several strands of prior work on review generation, firm interventions in review systems, and the role of reviews in shaping sales and returns. In what follows, we situate our contribution within the literature.

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.

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). Karaman (2025) shows that soliciting consumers to post reviews reduce their future spending on the solicited brand. 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. 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.

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 our study, the review environment that writers and buyers encounter along their consumer journey, and the natural variation in review solicitation 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 corresponds to an order and includes order ID, user (buyer) ID, product ID, price paid, purchase date. For each transaction, we also observe 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, a photo review dummy (indicating whether a photo was included), the review text (currently being collected), and its timestamp.

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 2.3 for details). Because reviews are inherently endogenous, for example, popular products with high unobserved demand shocks tend to receive both higher sales and more reviews, this exogenous variation enables us to effectively address endogeneity concerns.

During our sample period, 83,409 users made purchases on the retailer's website and generated 29,953 reviews. Although the retailer offers thousands of products, most are purchased infrequently, with a small subset accounting for the majority of revenue. We therefore focus on the top 152 products by revenue, which together account for 43.5% of all reviews generated in 2019. This sampling criterion yields an estimation sample containing 142,079 purchases, 13,031 reviews, and 6,856 returns by 60,614 users. Additional details on sample construction and the operationalization of product prices are provided in Appendix A.

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

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.

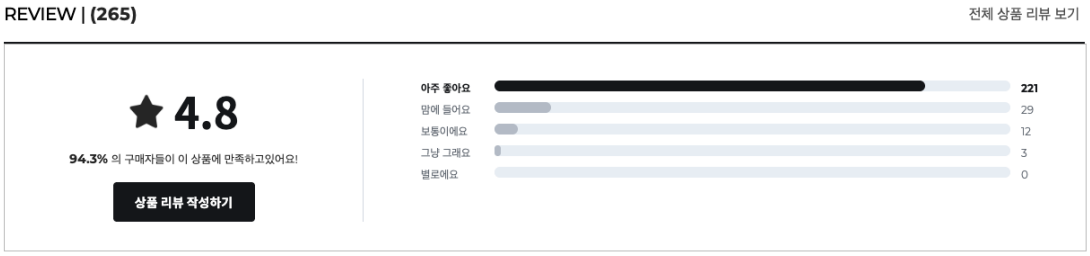


Figure 2: Summary of reviews displayed to buyers

2.2.2 Review writing process

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 notify customers of the delivery and to encourage them to share feedback by streamlining the review submission process. As shown in Figure 3, the message typically includes a brief prompt such as "Your product has been delivered. 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

2.3 Natural experimental variation in review solicitation:

The retailer in our study, similar to many independent e-commerce websites, integrates a third-party solution into its website to collect, manage, and display customer reviews.[4]

[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).

[4]Popular third-party 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 marketing tool integrations.

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 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.

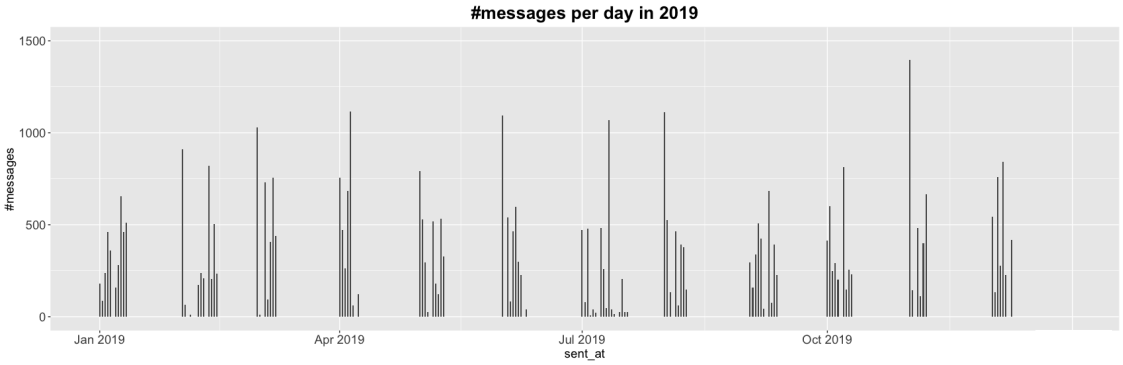


Figure 4: Natural experimental variation in review solicitation

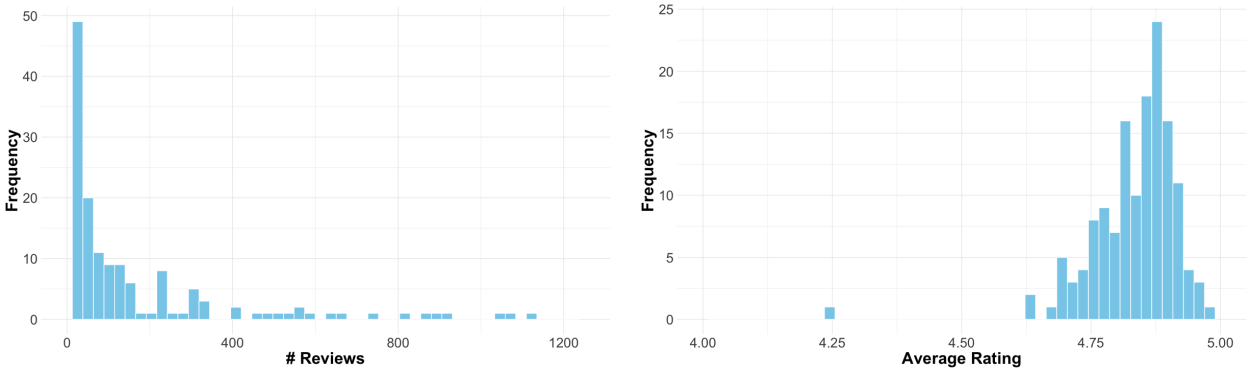
Review solicitations are sent for all orders placed from the beginning of each month until the free message quota is depleted. Figure 4 displays the daily number of review solicitations sent in 2019. Since review solicitations are triggered by actual orders, both the daily spikes and the days solicitations are sent reflect fluctuations in consumer demand. On average, about 2,600 solicitations are sent each month, typically for the first ten days.[5] 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. Of note, the retailer does not label whether a review is written following a solicitation, meaning buyers cannot distinguish solicited from unsolicited reviews on the product page.

[5]The monthly free quota varies slightly because the provider occasionally issues bonus messages, for instance, when system glitches occur during updates or when the review section experiences performance issues. The provider sends messages for all incoming orders starting at the beginning of the month, rather than spreading them randomly throughout the month, likely because many retailers purchase additional messages once the free quota runs out.

2.4 Data description

Table 1: Summary Statistics

	Mean	SD	Min	25%	50%	75%	Max
User-level activities in 2019							
Number of purchases	2.34	1.90	1	1	2	3	20
Number of reviews	0.21	0.76	0	0	0	0	15
Product-day level flow metrics							
Number of purchases	6.13	8.87	0	1	3	8	153
Percent returned	5.37	14.09	0.00	0.00	0.00	3.85	100.00
Number of reviews	0.56	1.13	0	0	0	1	19
Percent five-star ratings	89.63	27.00	0.00	100.00	100.00	100.00	100.00
Product-day level cumulative metrics to date							
Number of reviews	199.19	261.95	0	24	82	261	1,125
Average rating	4.86	0.10	3.50	4.81	4.86	4.90	5.00
Percent five-star rating	90.39	6.97	0.00	87.45	90.91	93.55	100.00
Variance of ratings	0.18	0.097	0.00	0.14	0.19	0.25	0.56



(a) Number of Reviews **(b)** Average Rating

Figure 5: Distribution of #reviews and average rating

Table 1 presents the summary statistics. The top panel reports user activity: each user makes about 2.3 purchases on average and writes 0.21 reviews over the sample period, corresponding to an approximately 9.0% probability of writing a review.[6] Given the limited

[6]Industry reports and recent academic studies estimate that only 2–19% of consumers leave a review after a purchase (Pocchiari et al., 2025).

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 6.13 purchases per day, of which 5.4% are returned, with considerable variation particularly in the return rate. Products also receive an average of 0.56 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 snapshot from December 31, 2019, is shown in Figure 5. The left panel displays the distribution of the number of reviews across products. While the mean is 199, 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.[7] Indeed, in our setting there is a strong negative correlation of -0.80 between average rating and variance. Accordingly, we focus on average rating as the primary outcome, and not both. Throughout our interpreta-

[7]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.

tion, we note that lower average ratings in our context are tightly associated with higher variance.

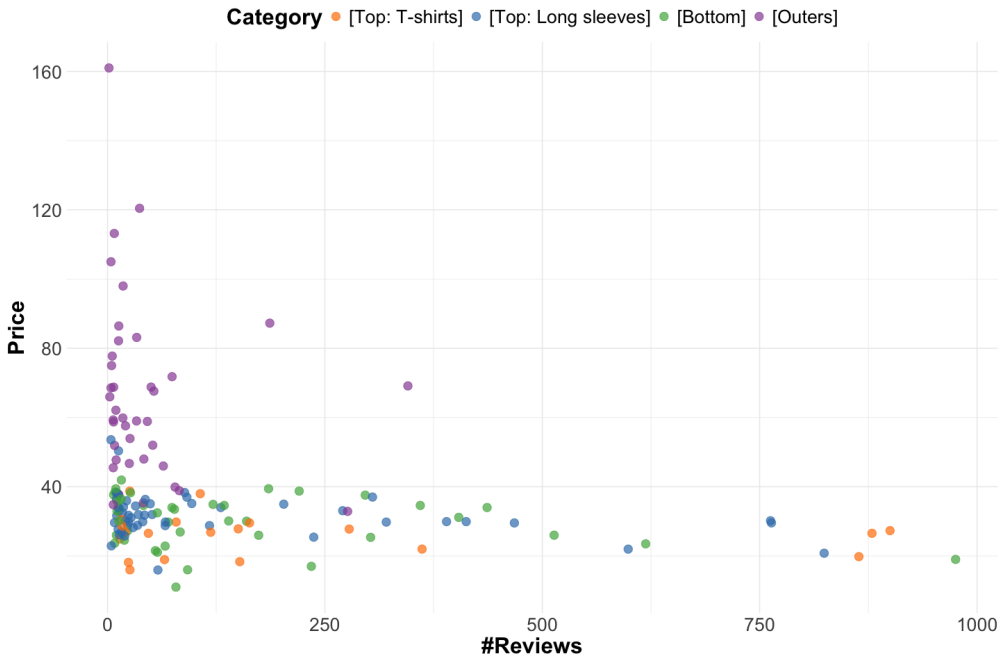


Figure 6: Reviews and product characteristics

To illustrate how review information varies across products, Figure 6 plots existing review volume and price by product category. Products in the Outers category, while being higher-priced, tend to accumulate fewer reviews, likely because they involve greater purchase uncertainty and require more complex fit assessments compared to basic items such as T-shirts. 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.

3 Empirical Strategy

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

3.1 Framework

This subsection exploits review solicitations as an exogenous lever (i) to measure reviewer responses across informational states (Stage 1) and (ii) to identify the state-contingent marginal effect of additional reviews on subsequent orders and return rates (Stage 2), enabling a comparison of reviewer responses and buyer value. Our empirical design separates review supply from demand-side value by first estimating how solicitations affect review incidence and rating composition, and then mapping these into market outcomes. We estimate both average and heterogeneous effects conditional on product's existing information environment.

We proxy each product's existing information environment using *review states* observed up to the day (Stage 1) or week (Stage 2) prior to the solicitation window. Review states are defined by the interactions of review–count and average rating bins. We index review state by $k \in \mathcal{K}$ and define $S_{jt-1}^{(k)} = 1$ if product j was in state k in (pre-solicitation) period $t-1$.

3.1.1 Stage 1: Reviewer response

Stage 1 estimates how solicitations affect individual reviewing behavior, allowing the effect to vary with the product's pre-existing review state. The unit of observation 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 $\mathsf{Solicited}_{ijt}$ is an indicator for receiving the post–purchase solicitation; $S^{(k)}_{j,t-1}$ denotes the product's review state in period $t-1$; X_{jt} includes price (with new–release status, product age, and product category considered in robustness checks); and $\lambda_{m(t)}$ are month fixed effects.

We interpret the heterogeneous solicitation effect in Stage 1, γ_k, as a conditional average treatment effect on the treated (CATT). In our setting, only those who purchase products are eligible to write reviews and receive solicitations. Accordingly, identification of γ_k relies on within-month variation in solicitation across days for a given set of product orders placed with existing review state $S^{(k)}_{jt-1}$, rather than comparisons across users/orders/products with different review states. Thus, γ_k represents the average solicitation effect for individuals who place orders for products that are in review states $S^{(k)}_{jt-1}$. We provide a detailed discussion of the identification strategy in Section 3.2.

3.1.2 Stage 2: Buyer value

Stage 2 identifies the state–contingent marginal effect of additional reviews on market outcomes of subsequent buyers. The unit of analysis is product j in week t. To avoid mechanical contemporaneous correlations (i.e., reserve causality where high sales lead to more reviews), we use non–overlapping windows: shifts in review supply pertain to week t, while outcomes are measured in week $t+1$. We study two outcomes,

$$Y_{j,t+1} \in \{\mathsf{Orders}_{j,t+1},\ \mathsf{ReturnRate}_{j,t+1}\}$$

and estimate the following Poisson Pseudo-Maximum Likelihood (PPML) model:

$$Y_{j,t+1} = \exp\Big(\tilde{\alpha} + \sum_{k\in\mathcal{K}} \delta_k\big(N_{jt}^{(k)} \times S_{j,t-1}^{(k)}\big) + \sum_{k\in\mathcal{K}} \tilde{\beta}_k S_{j,t-1}^{(k)} + X_{jt}'\tilde{\theta} + \tilde{\lambda}_{m(t)} + \sum_{k\in\mathcal{K}} \rho_k\big(\hat{v}_{jt}^{(k)} \times S_{j,t-1}^{(k)}\big) + \tilde{\varepsilon}_{jt}\Big) \tag{3.2}$$

where N_{jt} is the number of reviews for product j generated in week t. In other words, N_{jt} captures the shift in review information for product j presented to the buyers. Again, we allow the Stage 2 effects to vary by the existing review state k, by indexing $S_{jt-1}^{(k)} = 1$ if product j was in state k in week $t-1$. Thus, δ_k parameters capture the state–contingent marginal effects of one additional review during week t on next–week outcomes. δ_k, in combination with the results in Stage 1, tells us whether review supply in Stage 1 is allocated to states where buyers' values are highest. X_{jt} includes price (and other product attributes as in Stage 1 for robustness) and month fixed effects absorb seasonality in demand.

Here, N_{jt} denotes the number of reviews generated for product j in week t, capturing the week-t shift in review information presented to buyers. We again let effects vary by the product's existing review state: $S_{jt-1}^{(k)} = 1$ if product j was in state k in week $t-1$. The coefficients δ_k therefore capture the state-contingent marginal effect of one additional review on next-week outcomes. Together with the Stage 1 estimates, the δ_k parameters allow us to assess whether solicitations induce reviews to flow toward the review states where buyers derive the greatest value. The variable X_{jt} includes price (and, for robustness, additional product attributes as in Stage 1), and month fixed effects $\tilde{\lambda}_{m(t)}$ absorb seasonality in demand.

A central empirical challenge is that review generation is endogenous. For example, unobserved multi-week promotions for product j may influence both the reviews generated in week t and product demand in week $t+1$. To address this, we use review solicitations as an instrument for the generated reviews in Stage 2 (N_{jt}) and implement a control-function approach where the residuals $\hat{v}_{jt}^{(k)}$ are included to correct for endogeneity. Specifically, we first

predict review writing and rating behavior for each user-product–day from Stage 1 according to Equation (3.2), with solicitation serving as the instrument. Noting that N_{jt} captures the review information generated for product j in week t, we aggregate the Stage 1 residuals to the product-week level as

$$\hat{v}_{jt}^{(k)} = \sum_{d \in t} \sum_{i \in \mathcal{I}_{jt}} \widehat{\varepsilon}_{ijd} \, \mathbf{1}\{S_{jt-1}^{(k)} = 1\}, \tag{3.3}$$

where $\widehat{\varepsilon}_{ijd}$ is the user-product-day level residual from Stage 1, $d \in t$ indicates the days in week t, and $\mathcal{I}_{jt}$ denotes the set of individuals who purchased product j during that week. To examine heterogeneous treatment effects, we construct state-specific residuals $\hat{v}_{jt}^{(k)}$ by aggregating $\widehat{\varepsilon}_{ijd}$ for each pre-week review state k, and include these in Stage 2 to correct for the endogeneity of $N_{jt}^{(k)}$.

3.2 Identification

In this subsection, we outline the identification strategy for each stage of the analysis.

3.2.1 Stage 1 identification

For the heterogeneous solicitation effects in Stage 1, γ_k, to be interpreted as CATT (the Conditional Average Treatment Effect on the Treated), two standard assumptions are required: unconfoundedness and overlap. Unconfoundedness requires that the assignment of solicitation behaves as if random within each value of the review state $S_{jt-1}^{(k)}$ and that no unobserved confounders remain after conditioning on observed covariates $(X_{jt}, \lambda_{m(t)})$. The overlap assumption requires that, for every value of $S_{jt-1}^{(k)}$, we observe both treated and untreated units. As discussed in Section 2.3, our setting generates quasi-random variation in solicitation across days, particularly between the early-month "on" period (approximately days 1–10) and the later-month "off" period. In Appendix B, we provide empirical evidence

supporting the effective randomness of solicitation assignment - that treatment status (receipt of a solicitation) is independent of potential outcomes (whether a customer writes a review) - and document the presence of common support, with both treated and control observations within each review state $S^{(k)}_{jt-1}$.

3.2.2 Stage 2 identification

In Stage 2, we use review solicitations as an instrument for the generated reviews (N_{jt}). The causal interpretation of the IV estimator relies on the following conditions: relevance, exclusion restriction, and exogeneity.

Relevance requires that solicitations should significantly affect generation of reviews. This condition is satisfied in our case as Stage 1 results show review solicitations significantly impact whether customers write a review (see subsection 4.1 for Stage 1 results).

Additionally, identification of (3.2) hinges on the exclusion restriction that solicitations affect $Y_{j,t+1}$ only through their impact on review creation. This requires that merely receiving or viewing the solicitation, without writing a review, does not directly influence subsequent sales or return rates. Several institutional features of our setting support this assumption. First, solicitation messages are sent upon product delivery. Second, they contain no monetary incentives or promotional content. Third, users are not directed to the retailer's website through the solicitation; instead, the message links to a standalone review-submission interface (see Figure 1). Fourth, subsequent buyers cannot distinguish whether a posted review was solicited. A potential violation would arise if solicitations reminded customers of the retailer and thereby stimulated additional visits or purchases. In our context, however, both solicited and unsolicited customers receive product delivery, meaning the retailer is equally 'top of mind' for all customers regardless of solicitation status. Moreover, users in our sample make only about 2.3 purchases that are spaced months apart, suggesting that reminder effects are minimal. Consistent with this, we empirically show in Appendix B that purchase

behavior does not differ between solicited and unsolicited customers.

Lastly, the exogeneity condition requires that the instrument be uncorrelated with unobservables. To absorb product quality and demand shocks, we control for key product characteristics (price, new release status, product age, and product category) and include month fixed effects to capture seasonal tastes. Moreover, we separate the solicitation window (week t) from the outcome window (week $t+1$) to avoid potential contemporaneous feedback effects, such as high sales in week t simultaneously increasing both solicitation intensity and review volume.

We omit product fixed effects in our preferred specification because the review–state indicators $S_{j,t-1}^{(k)}$ (and their interactions with the instrumented change in reviews) are highly persistent and largely product–specific over weekly horizons. including product fixed effects would therefore absorb most cross–sectional variation in states and substantially reduce statistical power. Re-estimating (3.2) with product fixed effects yields the same qualitative patterns, larger effects at zero/low review counts and below-median ratings, but with attenuated magnitudes and lower precision, consistent with the limited within–product movement across states.

4 Results

This section presents the main empirical results from our two–stage framework. We first estimate how post–purchase review solicitations affect individual reviewing behavior across existing review states (Stage 1). We then identify how additional reviews causally affect subsequent orders and return rates (Stage 2). Finally, we compare where solicitations succeed in generating reviews with where those reviews deliver the greatest buyer value, documenting reviewer–buyer (mis)alignment and exploring mechanisms behind the first–review barrier using heterogeneity by review state, buyer type, and product price.

4.1 Stage 1: Reviewer response

Table 2 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 indicating that the order received the post-purchase SMS request to review. We also include indicators for the product's existing review state at the time of purchase, product characteristics such as whether the product is new (launched within the last 30 days), product age and price, and month fixed effects.

Table 2: 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	0.0112***	0.0112***	0.0112***	0.0112***	-0.0159***	-0.0160***	-0.0161***	-0.0165***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
Price			9.87×10^{-5}	6.57×10^{-6}			-0.0002	-2.17×10^{-5}
			(6.91×10^{-5})	(9.87×10^{-5})			(0.0002)	(0.0004)
Product attribute controls	No	No	No	Yes	No	No	No	Yes
Review state dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline outcome	0.0885	0.0885	0.0885	0.0885	0.9047	0.9047	0.9047	0.9047
N	142,079	142,079	142,079	142,079	13,031	13,031	13,031	13,031
R^2	0.00806	0.00826	0.00827	0.00830	0.00236	0.00422	0.00430	0.00641

All review states are those at the time of purchase.
Heteroskedasticity-robust standard errors in parentheses; in robustness we cluster by product.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

Models (3) and (7) with price and review state dummy controls are the baseline specifications; the remaining columns show robustness with respect to price, product attributes (product age, category dummies), and review state dummies. On average, a solicitation increases the probability that a buyer writes a review by about 1.1 percentage point ($p < 0.01$). Conditional on writing, the probability of giving a five-star rating declines by about 1.6 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 is not driven by correlated changes in new product status, product age, or existing review state.

Table 3 examines whether reviewer response varies with the existing review state, which represents product information already available to buyers at the time of purchase. We define review states by review counts and average rating. Review count is partitioned into terciles of the empirical distribution: $(0, 43]$, $[43, 201)$, and $[201, +)$, which we call T1 count, T2 count, and T3 count, respectively. Rating is split at the sample median, 4.85. In our data, the rating split also serves as a dispersion split: lower-than-median ratings coincide with much higher variance (e.g., mean variance 0.26 versus 0.11 in the high-rating group), so we interpret "low rating" as "higher-variance/noisier information" throughout. We use the same definition of review states in Stage 2.

Table 3: Effect of solicitation 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 reviews)	0.0040	0.0041	0.0039	-0.0375	-0.0382	-0.0399
	(0.0091)	(0.0091)	(0.0091)	(0.0313)	(0.0313)	(0.0313)
Solicited × I(T1 count & Low rating)	0.0125**	0.0126**	0.0126**	-0.0337*	-0.0333*	-0.0323*
	(0.0054)	(0.0054)	(0.0054)	(0.0197)	(0.0197)	(0.0196)
Solicited × I(T2 count & Low rating)	0.0155***	0.0155***	0.0154***	-0.0263*	-0.0265*	-0.0259
	(0.0043)	(0.0043)	(0.0043)	(0.0160)	(0.0160)	(0.0159)
Solicited × I(T3 count & Low rating)	0.0078*	0.0078*	0.0079*	0.0087	0.0086	0.0089
	(0.0041)	(0.0041)	(0.0041)	(0.0147)	(0.0147)	(0.0147)
Solicited × I(T1 count & High rating)	0.0116**	0.0116**	0.0116**	-0.0102	-0.0103	-0.0117
	(0.0045)	(0.0045)	(0.0045)	(0.0141)	(0.0141)	(0.0141)
Solicited × I(T2 count & High rating)	0.0110***	0.0111***	0.0112***	-0.0275*	-0.0275*	-0.0281*
	(0.0042)	(0.0042)	(0.0042)	(0.0148)	(0.0148)	(0.0148)
Solicited × I(T3 count & High rating)	0.0111***	0.0111***	0.0111***	-0.0125	-0.0124	-0.0137
	(0.0038)	(0.0038)	(0.0038)	(0.0109)	(0.0109)	(0.0109)
Price		9.81×10^{-5}	7.55×10^{-6}		-0.0002	-2.78×10^{-5}
		(6.91×10^{-5})	(9.87×10^{-5})		(0.0002)	(0.0004)
Product attribute controls	No	No	Yes	No	No	Yes
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	142,079	142,079	142,079	13,031	13,031	13,031
R^2	0.00827	0.00829	0.00835	0.00468	0.00476	0.00724

All review states are those at the time of purchase.
T1/T2/T3 count bins are $(0, 43]$, $[43, 201)$, and $[201, +)$.
Low rating is < 4.86; high rating is $[4.86, 5.00]$.
Heteroskedasticity-robust standard errors in parentheses; in robustness we cluster by product.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

We focus on models (2) and (5); the remaining columns 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: solicitations rarely generate the very first review even though early reviews have high informational value for buyers. Once a small base of information exists (e.g., T1 count in both rating groups), solicitations become markedly more effective. For low-rating products, the largest increases occur in the T2 count bin (43–201 reviews), where coefficients for review incidence are around 1.6 percentage points. For high-rating products, solicitation effects on review incidence are more even across count bins, suggesting that when reputation is already well established (T3 count), an additional reviews at broadly similar rates to lower-count states (T1 and T2).

These heterogeneous effects are not simply an artifact of sales velocity or product mix. As Figure 6 shows, products with few or no reviews tend to be higher-priced, belong disproportionately to categories such as outerwear and sets that require more complex fit assessments, and sell more slowly than basics such as T-shirts. One might therefore worry that the first-review barrier simply reflects that zero-review products are unusual, high-risk, or low-traffic items. Prior work indeed 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 that view, products with zero reviews simply have low sales velocity, so the first-review barrier would reflect a lack of chances to review. Our order-level empirical design allows us to separate this arrival channel from the decision to write. We compare solicited versus non-solicited buyers *conditional on* a purchase having already occurred, so the pool of potential reviewers and the arrival rate of buyers are held fixed. In robustness checks where we add or do not add price and product attributes (product age and category dummies), the negligible solicitation effect at the zero-review state persists, while effects at T1 and T2 remain positive and significant. Together, these patterns suggest that the first-review barrier is not merely picking up that zero-review products are expensive, slow-moving, or otherwise unusual.

Conditional on writing, in model (5), the decline in the five-star probability has a low statistical significance and is modest in the states where reviews are somewhat established (T1 and T2 in the low-rating group and T2 in the high-rating group). Given the existing review volume and average ratings for these products, such changes in composition are unlikely to shift product-level average ratings in economically meaningful ways.

Table 4: Summary statistics of changes in predicted review states

Product-level	Mean	SD	Min	Median	Max
Predicted # new reviews	4.06	5.61	0	2	53
Predicted change in avg. rating	-0.0015	0.077	-1.36	0.00	1.38
Predicted change in var. rating	0.0048	0.049	-0.41	0.00	0.56

Prediction for 7 days; N = 21,718
Baseline: #review 201.5, avg rating: 4.86, rating variance: 0.18

Table 4 summarizes how review states are predicted to change over a seven-day window. On average, about four additional reviews are written per product, from a baseline stock of roughly 202 reviews, while the predicted changes in average rating and rating variance are centered very close to zero and small relative to their baselines (4.86 and 0.18, respectively). Together with the state-contingent estimates above, this pattern confirms that changes in review states are driven primarily by additional volume, with only minor shifts in rating composition. Accordingly, in Stage 2 we focus on the marginal effect of additional reviews, treating rating levels and dispersion as approximately fixed over the short (week-level) horizon we study.

4.2 Stage 2: Buyer value

Table 5 reports the marginal effect of additional reviews on market outcomes, next-week orders and return rates using the PPML control-function specification described in

Section 3.1.2. Outcomes are measured at the product-week level, and the return rate is the number of returns divided by the number of orders over the corresponding 7 days. All specifications include review-state dummies and month fixed effects, and models (2)-(3) and (5)-(6) sequentially add price, product attributes (product age and category dummies) as controls. The key regressors are the number of reviews in week t, and the sum of first-stage residuals from the Stage 1 regression, which implements the control function correction for endogeneity.

Table 5: Average marginal effect of reviews on orders and return rate

Dependent Variables:	Orders_{t+1}			Return rate_{t+1}		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
# Reviews_t	0.0868***	0.0864***	0.0845***	-0.0188***	-0.0162***	-0.0171***
	(0.0013)	(0.0013)	(0.0013)	(0.0018)	(0.0017)	(0.0018)
Price		-0.0086***	-0.0017***		0.0088***	0.0079***
		(0.0005)	(0.0005)		(0.0006)	(0.0008)
Product attribute controls	No	No	Yes	No	No	Yes
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
CF residuals	Yes	Yes	Yes	Yes	Yes	Yes
Baseline outcome	26.11	26.11	26.11	5.69	5.69	5.69
N	20,276	20,276	20,276	19,709	19,709	19,709
Pseudo R^2	0.53588	0.54289	0.55062	0.01682	0.02867	0.06747
BIC	545,890.6	537,663.6	528,637.2	209,080.4	206,572.7	198,400.1

Heteroskedasticity-robust standard errors in parentheses; in robustness we cluster by product.
Return rate: (no. returns / no. orders) $\times$ 100% over the 7 days.
Signif.: $^{***}p < 0.01$, $^{**}p < 0.05$.

Our main specifications are models (2) and (5) as before. In model (2), the coefficient of the number of reviews implies that one additional review in the prior week increases expected orders in the subsequent week by roughly 9% ($exp(0.086) - 1 \simeq 0.09$). The coefficient in model (5) translates into reductions in returns by about 0.09 percentage point, given baseline weekly return rates of 5.69%. The negative return-rate effect is consistent with an information mechanism in which additional reviews reduce ex ante uncertainty about fit and quality, thereby lowering misfit-driven returns (Sahoo et al. (2018)).

Table 6 extends the same PPML control function framework to allow the marginal effect of reviews to vary with the existing review state. We interact the number of reviews in week

t and the first-stage residuals with the review-state indicators as of the beginning of week t. Models (1)–(3) report state-contingent effects on orders, and columns (4)–(6) on return rates. We focus on the baseline specifications in models (2) and (5). The specifications that do not add price or add product age and category controls yield very similar patterns.

Table 6: Marginal effect of reviews on orders and return rate: heterogeneity by existing review state

Dependent Variables:	Orders_{t+1}			Return rate_{t+1}		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
# Reviews_t × I(0 reviews)	0.1534***	0.1470***	0.1395***	-0.0945***	-0.0620***	-0.0631***
	(0.0062)	(0.0061)	(0.0061)	(0.0125)	(0.0113)	(0.0106)
# Reviews_t × I(T1 count & Low rating)	0.1589***	0.1559***	0.1516***	-0.0327***	-0.0196***	-0.0156**
	(0.0039)	(0.0038)	(0.0036)	(0.0065)	(0.0064)	(0.0063)
# Reviews_t × I(T2 count & Low rating)	0.1214***	0.1197***	0.1185***	-0.0112***	-0.0058	-0.0095**
	(0.0036)	(0.0036)	(0.0036)	(0.0043)	(0.0042)	(0.0040)
# Reviews_t × I(T3 count & Low rating)	0.0756***	0.0762***	0.0759***	0.0009	-0.0015	-0.0050**
	(0.0015)	(0.0015)	(0.0016)	(0.0022)	(0.0022)	(0.0024)
# Reviews_t × I(T1 count & High rating)	0.1312***	0.1290***	0.1235***	-0.0481***	-0.0374***	-0.0414***
	(0.0036)	(0.0036)	(0.0038)	(0.0061)	(0.0058)	(0.0057)
# Reviews_t × I(T2 count & High rating)	0.0962***	0.0961***	0.0918***	-0.0164***	-0.0171***	-0.0212***
	(0.0026)	(0.0026)	(0.0028)	(0.0039)	(0.0040)	(0.0044)
# Reviews_t × I(T3 count & High rating)	0.0825***	0.0823***	0.0799***	-0.0242***	-0.0240***	-0.0214***
	(0.0020)	(0.0020)	(0.0021)	(0.0035)	(0.0034)	(0.0036)
Price		-0.0056***	-0.0013**		0.0084***	0.0075***
		(0.0004)	(0.0005)		(0.0006)	(0.0008)
Product attribute controls	No	No	Yes	No	No	Yes
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
CF residuals	Yes	Yes	Yes	Yes	Yes	Yes
N	20,276	20,276	20,276	19,709	19,709	19,709
Pseudo R^2	0.57558	0.57849	0.57873	0.02051	0.03087	0.06928
BIC	499,335.9	495,925.0	495,702.4	208,415.7	206,223.4	198,133.9

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; $(0, 43]$, $[43, 201)$, and 201+.
Low rating is < 4.86; high rating is $[4.86, 5.00]$.
Heteroskedasticity-robust standard errors in parentheses; in robustness we cluster by product.
Significance: $^{***}p < 0.01$, $^{**}p < 0.05$.

The pattern connects directly to the logic that prior information determines buyer value of a review. The marginal effect of an additional review on 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. T3 count states have the lowest effect, consistent with diminishing returns to information. This pattern holds for return rates as well and is even more pronounced: the largest reductions occur at the zero review state, followed by T1 count bin, with a small exception at T3 for high

rating products. 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.

4.3 Reviewer–buyer (mis)alignment

We next examine whether reviewer response to solicitation occurs in the product states where buyers value additional information most. Figure 7 juxtaposes solicitation effects on review incidence and the five-star probability from Table 3, models (2) and (5) (top panels) with state-contingent PPML effects of one more review on next-week orders and return rates from Table 6, model (2) for orders and (5) for return rates (bottom panels). In each row, panels split observations by rating (low: < 4.86; high: $[4.86, 5]$) and by review count bins (0, T1, T2, T3).

The top panels show reviewer response. At zero prior reviews, the lift in review incidence is close to zero and statistically indistinguishable from zero, indicating a first-review barrier: solicitations rarely generate the very first review. Once a small base exists at the T1 and T2 states, solicitations are more effective, with review incidence increasing by about 1–1.5 percentage points and only modest declines in the five-star share. In other words, solicitations primarily convert marginal buyers into reviewers once some review information is already present.

The bottom panels plot buyer value. An additional review delivers the largest gains when information is scarce. Positive effects on orders and negative effects on return rates are strongest at zero and low counts. 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. Comparing the two rows in Figure 7, misalignment arises because reviewer response is weakest exactly where buyers derive the highest marginal value from another review. In Appendix C, we further show that this

reviewer–buyer misalignment is especially consequential when a product is attracting new customers: early reviews reduce return rates more strongly for new buyers, who lack product- and brand-specific experience, than for existing buyers.

4.4 Interpretation of the first-review barrier and misalignment

We now ask what it means that it is so hard to generate the first review despite its high buyer value. As discussed in Section 4.1, our order-level design and controls already address concerns about traffic and product mix for the first-review barrier. We do not claim to identify a single structural mechanism. Instead, we focus on what our results imply about reviewers' motives and what they reveal about psychological frictions in initiating the first review.

Not internalizing buyer value. Our misalignment results imply that reviewers are unlikely to internalize the value of their contribution for future buyers. If buyers rationally directed their reviewing effort to where their review is most helpful, we would expect the first-review barrier to be weakest in exactly those states where early reviews have the largest downstream impact on sales and returns. Our results point in the opposite direction. Stage 2 shows that an extra review has the largest marginal impact on orders and, especially, on return rates in zero- and low-review states with below-median ratings (Table 6). In those same states, Stage 1 shows that solicitations barely move review incidence at zero reviews and only modestly at low counts (Table 3). Early reviews are therefore both highly valuable to buyers and under-supplied by solicited reviewers, which is hard to reconcile with reviewers primarily internalizing buyer value.

The price split sharpens this point. Price is a natural dimension along which reviewer–buyer (mis)alignment may vary because high-priced products involve larger financial stakes and higher economic loss from mismatches, so the marginal value of pre-purchase information

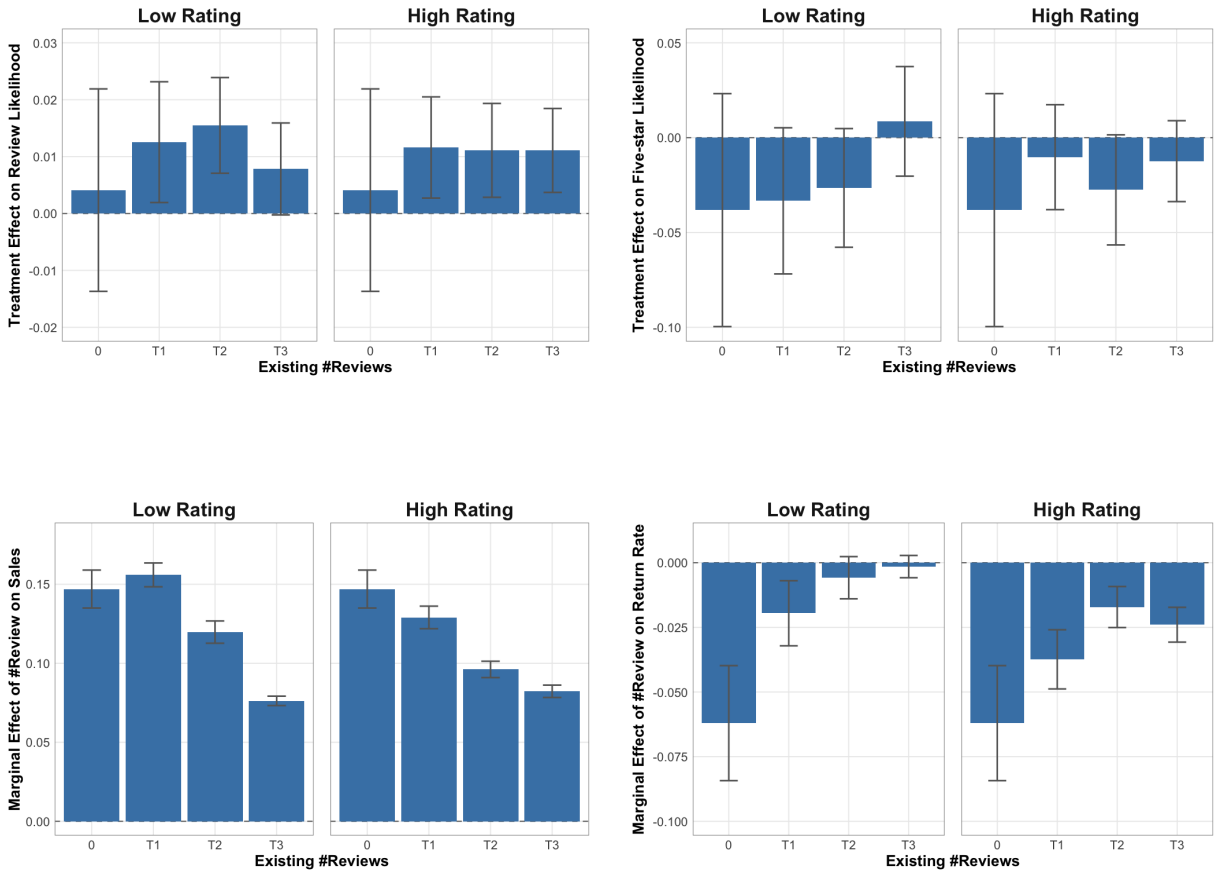


Figure 7: Reviewer response (top) and buyer value (bottom) by existing review state.

Note.
1) Top panels: solicitation effects on review incidence (left two subplots) and on the probability of a five-star rating conditional on reviewing (right two subplots), shown separately by rating group (Low: < 4.85, High: $[4.85, 5]$) and by review-count bins (T1: $(0, 43]$, T2: $[43, 201)$, and T3: $[201, +)$).

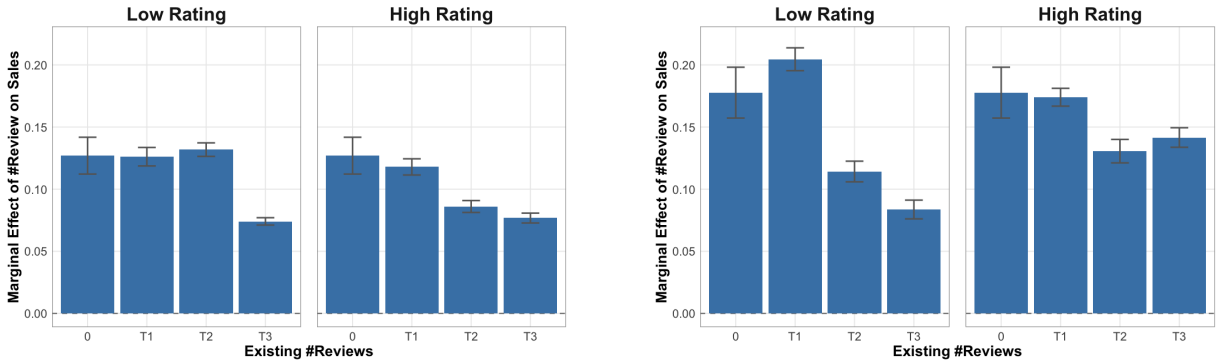
2) Bottom panels: instrumented, state-contingent marginal effects of additional reviews on orders (left two subplots) and on next-week return rate (right two subplots), shown separately by the same rating groups and review-count bins.

3) Error bars show 95% confidence intervals.

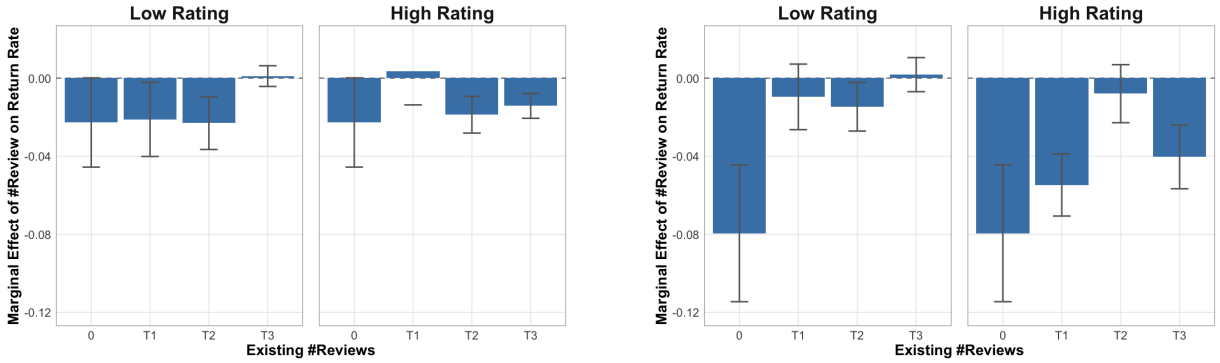
should be greater for expensive goods, both for conversion (orders) and for returns. At the same time, price co-varies with category and sales velocity (Figure 6), so a given review count can imply different informational richness across price points.

We split products at the median of the (time-invariant) product-level 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: $[43, 201)$, T3: $[201, +)$) and the same median rating split (4.86). On the buyer side, Figure 8 panels (a) and (b) show that within each price stratum the state-contingent shapes mirror our baseline pattern: effects are largest at zero and T1 counts and attenuate thereafter, but the magnitudes are consistently larger for high-priced products. When an expensive product with few reviews receives one more review, the gains in orders and, especially, the reductions in return rates are substantially larger than for low-priced items. From the buyer's perspective, high-priced products are therefore precisely where an additional early review is most valuable. On the reviewer side, panel (c) of Figure 8 shows that the first-review barrier is just as strong for high-priced products as for low-priced ones: in both strata, the estimated solicitation effect on writing the very first review is small and statistically indistinguishable from zero, while lifts become sizable only once a minimal base exists.

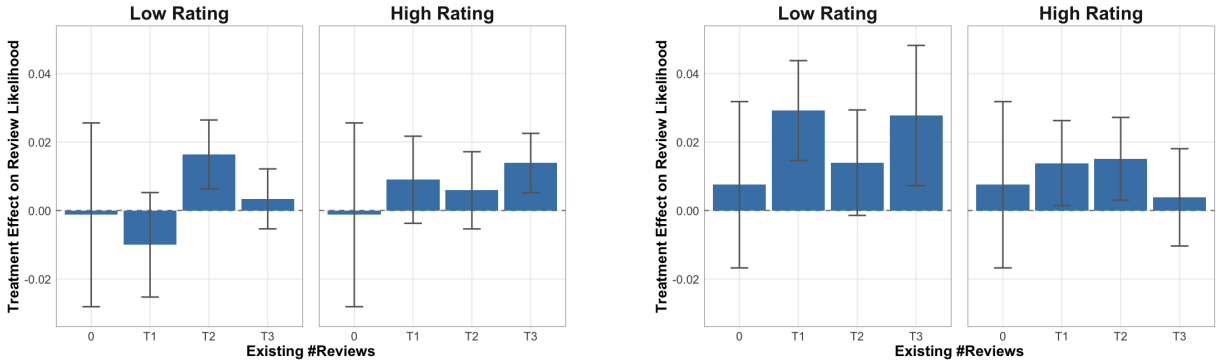
If altruistic or outcome-based motives were the dominant drivers of the first-review decision, we would expect the barrier to be attenuated for high-priced items, where the consequences of misinformation are greater and where we directly observe larger buyer gains from early reviews. Instead, the first-review barrier persists and the misalignment is *larger* for expensive products: early reviews deliver more buyer value there, yet they are no easier to elicit. Thus, the price-split analysis further suggests that reviewers are not especially willing to initiate reviews where buyer value is highest.



(a) Marginal effect of reviews on **sales** by price



(b) Marginal effect of reviews on **return rate** by price



(c) Solicitation effect on **review likelihood** by price

Figure 8: Price heterogeneity in buyer value and reviewer response (left: low price, right: high price)

Psychological barrier to initiating the first review. Our interpretation is that the first-review barrier reflects psychological frictions that make initiating the review history harder than contributing to an existing stream, much like the difficulty of starting a conversation rather than joining one. Prior work on online reviews typically models review writing as a decision that incurs effort and hassle costs, with participation increasing when these costs are reduced through reminders, financial incentives, or social norms (e.g., Hu et al., 2017; Moe and Schweidel, 2012; Burtch et al., 2018). Studies of review solicitations in particular show that post-purchase requests reliably increase review incidence and reduce extremity bias, primarily by lowering effort cost (e.g., Askalidis et al., 2017; Karaman, 2021; Brandes et al., 2022; Gao et al., 2025). Our average Stage 1 effect, an approximately one-percentage-point increase in review incidence, is consistent with this "effort reduction'' view.

However, the heterogeneous Stage 1 results indicate that this response is weakest exactly at the zero-review state. Conditional on a purchase, solicitations barely move the probability of writing the *first* review, but they do increase reviewing once a minimal base exists. Universal cost reduction alone is therefore insufficient: there appears to be an additional, state-dependent psychological cost associated with setting the first precedent.

Several mechanisms are consistent with this interpretation. First, being the first reviewer likely heightens evaluation apprehension. Without existing reviews as reference points, buyers may be uncertain about what aspects to comment on, what level of detail is appropriate, or how their rating will be interpreted, increasing the perceived effort and social risk of posting. This is consistent with evidence that the public and enduring nature of online feedback shapes both whether and how consumers contribute (e.g., Moe and Schweidel, 2012; Schoenmueller et al., 2020). Second, early reviews are unusually salient and persistent. Many products, especially high-priced, slow-moving items, accumulate reviews slowly, so the first few reviews can remain highly visible and influential for a long time, making early

reviewers disproportionately responsible for shaping the product's reputation. Park et al. (2021) show that initial reviews can cast a long shadow on subsequent review dynamics, underscoring that early contributions can have lasting effects on the product's reputation. Third, in the absence of existing reviews there is little social proof that reviewing is normal behavior for the product. Seeing prior reviews provides both a template for content and a norm that "people like me review this product," reinforcing the social and normative motives documented in work on user-generated content and review incentives (e.g., Khern-am nuai et al., 2018; Woolley and Sharif, 2021). Viewed through this lens, our misalignment patterns are consistent with solicitations primarily reducing generic effort costs but doing little to alleviate these first-review-specific frictions.

We cannot fully rule out all alternative interpretations. For example, it may be that buyers who purchase products in a no-review state are intrinsically less inclined to review under any lever. However, the combination of order-level identification that holds purchase opportunities fixed, robustness to controls for price, age, and category, and the price-split evidence that weakens altruistic explanations jointly point to psychological frictions in initiating review information as a potential dominant driver of the first-review barrier. This interpretation also clarifies why our counterfactual policies in Section 5 focus on the zero-to-one transition: overcoming the psychological cost of being first is key to aligning reviewer response with buyer value where it matters most.

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. Our two–stage model delivers the key parameters for simulation: state–specific solicitation

lifts in reviews $\widehat{\gamma}_k$ from Stage 1 and state–specific buyer–value $\widehat{\delta}_k$ from Stage 2.

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 solicitation message and aggregate them over 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 (0.0041), the sample average treatment effect (0.0112), and the upper tail of the HTEs we observe (0.0155), 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 detailed simulation procedure is provided in Appendix D.

Figure 9 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 (bottom) panel shows the change in net revenue (number of reviews written) per solicitation message that were sent to zero-review state products. Dashed vertical markers indicate three salient benchmarks from the in-sample estimation in Stage 2: the realized HTE for no review products (about 0.0041), the sample ATE (0.0112), and the highest HTE observed in our data (0.0155). The curves are monotonically increasing. Increasing the solicitation effectiveness for the first review from 0 to 0.02 increases per-solicitation net revenue by 4.88 USD on average. The distinctly

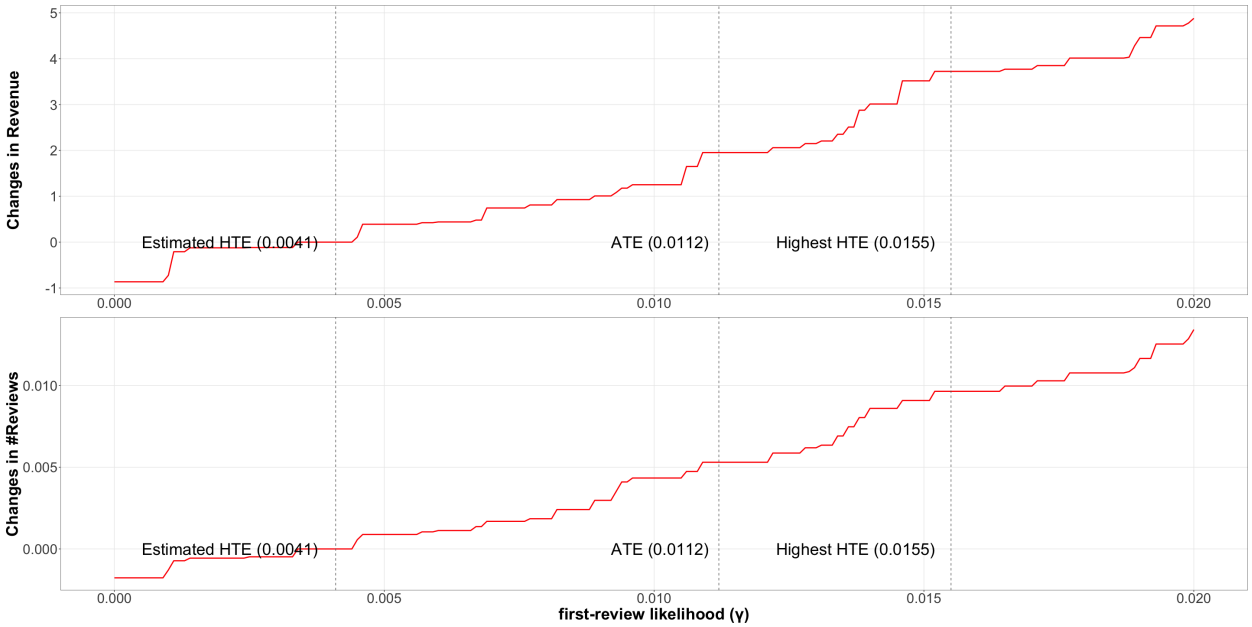


Figure 9: Changes in Net Revenue (top) and Reviews (bottom) per number of solicitation messages under Hypothetical Solicitation Effect for zero review state

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.

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 state by 0.013 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 0.0041) toward the sample average (about 0.0112) and then toward the best observed lift in the data (about 0.0155) 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.

Funding and Competing Interests

The data provider has a right to review the manuscript for the following conditions:

- The name of the data provider or its affiliates is not used;
- No data or any information from which the identity of the data provider or its affiliates could be inferred is used.
- No information relating to visitation frequencies, traffic levels or sales levels on the data provider affiliates' sites is disclosed, except for anonymous references to small portions or samples of the data provider affiliates, and from which overall traffic and sales patterns of the data provider affiliates' sites cannot be determined;
- No personal data is presented that identifies registrants or users of the data provider sites;

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.

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.

Karaman, H. (2025). The asymmetric effects of posting an online review on future spending and the dark side of solicitations. *Management Science*.

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.

Pocchiari, M., Proserpio, D., and Dover, Y. (2025). Online reviews: A literature review and roadmap for future research. *International journal of research in marketing*, 42(2):275–297.

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: Prevalence, 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 Sample criteria

Focal products Although more than 3,000 products appeared 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 analyses on economically meaningful and reasonably active products, we first rank products in descending order by (i) total 2019 revenue and (ii) daily revenue. We then select the top 300 products under each criterion and take the intersection of these two sets. The resulting products represent those most relevant to the retailer's managerial decisions. The retailer also sells certain products as bundles in addition to offering the items individually. Because the accumulated review information on the individual product pages and how it affects consumers' decisions between bundled versus individual purchases raise questions beyond the scope of this research, we exclude bundled products from our analysis.

Outlier handling Even after narrowing the sample to the focal products, many observations and outcomes of interest contain a large number of zero values and exhibit highly skewed distributions, posing substantial empirical challenges (see Table 1). To enhance the generalizability and robustness of our managerial insights, we apply outlier-handling criteria across several dimensions, including product sales, existing reviews, and user purchases. Specifically, we exclude one product with extreme day-level sales spikes (likely driven by unobserved promotional activity), three products with unusually large numbers of existing reviews ($\geq$1,200), and users with exceptionally high order volumes (z-score $\geq$ 3), who are likely resellers.

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 Additional Discussions on Identification

Table B.1 reports the balance between orders with and without review 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.

Table B.1: Balance Check

	Solicited	Control
Price ($)	31.46 (12.63)	31.47 (12.82)
Product age (days)	249.71 (222.03)	246.81 (222.05)
# Reviews	250.38 (283.40)	247.07 (282.88)
Average rating	4.87 (0.083)	4.86 (0.10)
Variance in ratings	0.17 (0.087)	0.18 (0.093)

Mean (SD) across orders: level of treatment

C Buyer Value: New Buyers vs. Existing Buyers

The reviewer–buyer misalignment is especially consequential when a product is attracting new customers. Table C.1 reports estimates from the Stage 2 PPML with control function approach 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 6 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 C.1 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

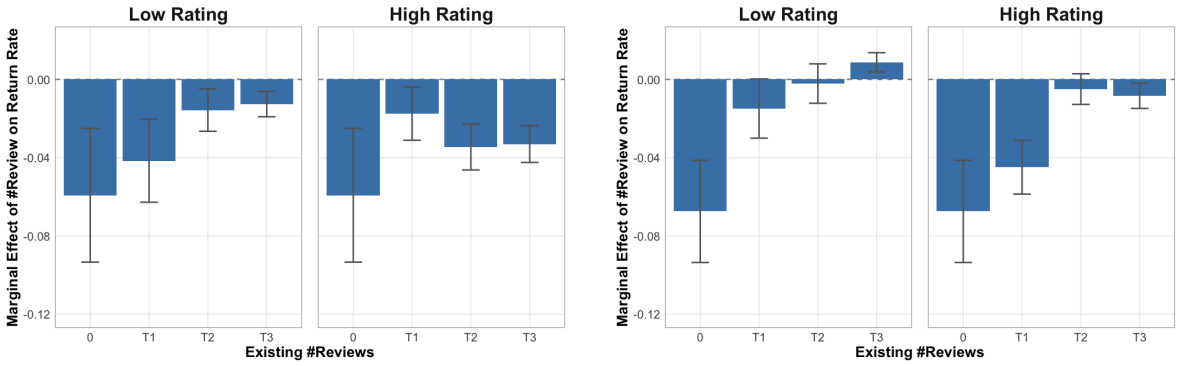
(high rating/low variance) looks favorable. Existing customers, by contrast, benefit most when the information environment is noisier (lower rating/higher variance) across review count bins, where additional reviews help them discriminate among close substitutes and reduce returns.

Table C.1: State-contingent effect of reviews on orders and return rate by buyer type

Dependent variable	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_t × I(0 reviews)	0.1470***	0.1513***	0.1414***	-0.0620***	-0.0592***	-0.0675***
	(0.0061)	(0.0064)	(0.0062)	(0.0113)	(0.0174)	(0.0133)
# Reviews_t × I(T1 count & Low rating)	0.1559***	0.1590***	0.1522***	-0.0196***	-0.0416***	-0.0149*
	(0.0038)	(0.0037)	(0.0041)	(0.0064)	(0.0108)	(0.0077)
# Reviews_t × I(T2 count & Low rating)	0.1197***	0.1194***	0.1199***	-0.0058	-0.0157***	-0.0021
	(0.0036)	(0.0036)	(0.0037)	(0.0042)	(0.0055)	(0.0051)
# Reviews_t × I(T3 count & Low rating)	0.0762***	0.0771***	0.0749***	-0.0015	-0.0126***	0.0087***
	(0.0015)	(0.0015)	(0.0015)	(0.0022)	(0.0033)	(0.0025)
# Reviews_t × I(T1 count & High rating)	0.1290***	0.1342***	0.1230***	-0.0374***	-0.0175**	-0.0449***
	(0.0036)	(0.0036)	(0.0038)	(0.0058)	(0.0069)	(0.0070)
# Reviews_t × I(T2 count & High rating)	0.0961***	0.0985***	0.0937***	-0.0171***	-0.0346***	-0.0050
	(0.0026)	(0.0026)	(0.0027)	(0.0040)	(0.0060)	(0.0040)
# Reviews_t × I(T3 count & High rating)	0.0823***	0.0809***	0.0834***	-0.0240***	-0.0331***	-0.0084**
	(0.0020)	(0.0019)	(0.0022)	(0.0034)	(0.0048)	(0.0033)
Price	-0.0056***	-0.0073***	-0.0029***	0.0084***	0.0079***	0.0068***
	(0.0004)	(0.0005)	(0.0004)	(0.0006)	(0.0009)	(0.0008)
Review state dummies	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
CF residuals	Yes	Yes	Yes	Yes	Yes	Yes
N	20,276	20,276	20,276	19,709	19,171	19,134
Pseudo R^2	0.57849	0.53651	0.53722	0.03087	0.02734	0.02858
BIC	495,925.0	321,229.6	285,310.0	206,223.4	250,038.2	252,302.9

Heteroskedasticity-robust standard errors in parentheses.
T1/T2/T3 count bins are $(0, 43]$, $[43, 201)$, and $[201, +)$.
Low rating is < 4.86; high rating is $[4.86, 5.00]$.
"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 C.1: Marginal effect of reviews on return rate by buyer type

Note.
1) For each buyer type, left: rating <4.86, higher variance; right: rating between $[4.86, 5]$, 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.

D 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, $\widehat{\text{Reviews}}^{\text{base}}_{j,t}$, from Stage 1; we then use Stage 2 to obtain baseline predictions for next–week orders and return rates:

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

Because the orders equation is estimated on the 7-day aggregation of Orders, we convert

them to a daily frequency:

$$\widehat{\text{Orders}}_{j,t}^{\text{base}} \;=\; \frac{1}{7}\Big\{\widehat{\text{Orders}}_{j,t+1}^{\text{base}})\Big\}, \qquad \widehat{\text{Returns}}_{j,t}^{\text{base}} \;=\; \widehat{\text{Orders}}_{j,t}^{\text{base}} \cdot \widehat{\text{ReturnRate}}_{j,t+1}^{\text{base}}.$$

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 $\widehat{\text{Reviews}}_{j,t}^{\text{cf}}(x)$, and feed them into Stage 2 to obtain

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

which we scale to daily values:

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

Counterfactual daily net revenue is

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

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

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

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