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

Hana Choi[*] Minkyung Kim[†] Jinsoul Seo[‡]

May 6, 2026

Abstract

Firms routinely solicit online reviews, yet it remains unclear whether solicitations generate reviews where future buyers value them most—and thereby increase sales and reduce product returns—because solicitations target reviewers while benefits accrue to buyers. Using individual-level panel data on purchases, solicitations, reviews, and returns from a large apparel e-commerce retailer and leveraging natural variation in solicitation exposure, we estimate a reviewer model measuring how solicitations affect review generation across informational states and a demand model measuring how additional reviews affect subsequent orders and returns. We find that an additional review improves downstream buyer outcomes, especially reducing return rates by up to 11.9% when prior review information is scarce. However, for reviewers, solicitations increase review incidence on average but are least effective in these low-information states. These patterns reveal a systematic reviewer–buyer misalignment: solicitations fail to elicit reviews where they matter most. We attribute this to a first-and-early review barrier driven by psychological frictions. A counterfactual exercise shows that increasing the solicitation effect in the zero-review state from its current near-zero level to 1.7%, the largest heterogeneous effect estimated across review-information states, raises net revenue by $0.72 per solicitation message, underscoring the value of overcoming early review barriers.

Keywords: online reviews; review solicitation; user-generated content; consumer information; product returns; review system design; natural experiments; e-commerce
JEL Codes: D12, D82, D83, L81, M31, C93

[*]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 Yufeng Huang, Minjae Kim, Kosuke Uetake, Sriraman Venkataraman and participants at Conference on Digital Experimentation (CODE), Conference on Artificial Intelligence, Machine Learning, and Business Analytics, University of Rochester Marketing Seminar, Marketing Science Conference, London Business School Summer Camp, Virtual Quant Marketing Seminar, Business and Marketing Analytics Seminar, and Frank M. Bass UTD FORMS Conference for helpful comments and suggestions.

1 Introduction

Online reviews play a central role in shaping consumer decisions by improving assessments of product quality and product–consumer fit. Recognizing their importance, firms do not passively wait for reviews; they actively solicit them, typically through post-purchase messages that invite buyers to rate and review their purchases. While firms intervene on past buyers, the value of review solicitation is realized if the resulting reviews improve future market outcomes.

This motivates our central question: do review solicitations generate reviews where buyers need them most, thereby increasing sales and reducing product returns? The answer is not straightforward. Solicitations target past purchasers, but their economic value accrues through subsequent buyers. Their effectiveness therefore depends on two forces: where reviewers choose to respond, and the extent to which the resulting reviews provide incremental informational value to future buyers. Prior research suggests that this informational value is greatest when existing product information is scarce (e.g., Anderson and Magruder, 2012; Zhao et al., 2013; Luca, 2016). For firms, the key question is whether solicitations generate reviews in these high-value states. Accordingly, the effectiveness of review solicitation hinges on the alignment between reviewer response and the informational value of reviews for buyers.

In addressing this question, we emphasize product returns alongside sales. Returns are theoretically informative because they provide a direct measure of mismatch between customer expectations—shaped by reviews—and realized product fit. They are also managerially important, as they impose substantial costs on both consumers and retailers. For consumers, returns involve hassle costs, including repacking, shipping, and time, which may reduce satisfaction and future purchase intent even when returns are nominally free. For retailers, returns require costly reverse logistics, customer service, inspection, restocking, and, in categories such as apparel, frequent liquidation. The aggregate stakes are considerable: U.S. retailers estimate that 16.9% of sales were returned in 2024 (approximately $890 billion), with online channels experiencing substantially higher return rates than brick-and-mortar stores.[1] Reviews speak directly to this problem by providing information about product quality and fit prior to purchase, thereby reducing uncertainty. If review solicitations successfully generate reviews where information is most needed, they can reduce

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

returns by mitigating post-purchase mismatch.

We study this question in partnership with a large apparel e-commerce retailer in South Korea that offers multiple products under a single brand and sends unincentivized SMS review solicitations upon delivery. Solicitation exposure varies due to the retailer's contract with a third-party review service provider: a fixed monthly allotment of solicitation messages is available at no additional cost, and messages are sent from the beginning of each month until the quota is exhausted. This institutional feature generates plausibly exogenous variation in solicitation exposure across delivered orders. Combined with rich individual-level panel data on purchases, solicitations, reviews, and returns, our dataset offers a valuable opportunity to examine how review solicitations shape review supply and, in turn, downstream market outcomes such as sales and returns.

Our empirical analysis links reviewer response to the informational value of reviews for future buyers. We allow this value to vary with the product's existing review information state, summarized by prior review count and average rating. We estimate a reviewer model that measures how solicitations affect review generation across these states, and a demand model that measures how newly generated reviews affect subsequent orders and return rates, again allowing those effects to vary with the existing information states. This design allows us to compare where solicitations successfully generate reviews with where an additional review creates the greatest value for future buyers.

On the buyer side, an additional review improves downstream outcomes, most notably by reducing return rates by 8.1% on average. Its value is also highly state dependent: an additional review increases next-week orders and reduces return rates by up to 11.9%, with the strongest effects occurring in zero- and low-review states. These patterns are consistent with an informational mechanism through improved product–consumer matching. On the reviewer side, solicitations increase review incidence on average, but the response is weakest precisely where the informational value to future buyers is highest. The effect of solicitation on review writing is negligible for products with no prior reviews and becomes strongest only once a minimal review base has already formed. Taken together, these patterns reveal a systematic reviewer–buyer misalignment: solicitations do not generate reviews where they matter most for subsequent buyers.

The observed misalignment is also informative about reviewer motivation. If reviewers fully internalized the downstream value their reviews provide to future buyers, reviews should be easiest to elicit in states where they generate the largest benefits, namely, when products have zero or few existing reviews. Instead, we find that reviewer response is weakest precisely in these states, suggesting that reviewer behavior is shaped by state-dependent psychological frictions, such as the difficulty of initiating the first review.

This interpretation carries important managerial implications. Simply reallocating or optimizing solicitation targeting—for example, by focusing on products with fewer reviews—is unlikely to yield substantial improvements in market outcomes, as the frictions that suppress reviewer response in low-information states would persist. By contrast, interventions that directly reduce first- and early-review barriers are likely to be more effective. In practice, such interventions include monetary or reputation-based incentives targeted at first reviews, as well as programs such as Home Depot's Seeds Program and Best Buy's Tech Insider Network, which subsidize early reviews for new or pre-release products.[2]

Building on this insight, we quantify the cost of the misalignment through a counterfactual exercise focused on the zero-review state. Increasing the solicitation effect in this state from the current estimate of 0.4% to 1.7%—the largest heterogeneous solicitation effect estimated across review-information states—raises net revenue by approximately $0.72 per solicitation message. Even modest improvements in eliciting first-and-early reviews translate into meaningful revenue gains, as they both expand demand and reduce costly returns, reflecting the high marginal value of information when existing reviews are scarce. More broadly, these findings underscore the managerial importance of overcoming the zero- and early-review barriers, where incremental alignment between reviewer response and the informational value to future buyers yields the largest economic payoff.

Our findings connect to and extend several strands of research on review generation, firms' interventions in review systems, and the role of reviews in shaping sales and returns. First, this paper contributes to the literature on review generation. Because reviews are generated by buyers, the accumulation of review information is shaped by buyer characteristics, sales velocity, and purchase timing (e.g., Li and Hitt, 2008; Godes and Silva, 2012; Park et al., 2021), as well as by selection into reviewing and the underlying motivations of reviewers (e.g., Hu et al., 2009; Moe and Schwei-

[2] Sources: https://www.homedepot.com/c/home_depot_seeds_program; https://www.bestbuy.com/site/misc/tech-insider-network/pcmcat1685559527403.c?id=pcmcat1685559527403.

del, 2012; Schoenmueller et al., 2020; Chakraborty et al., 2022; Sunada, 2025). Building on this literature, we examine how review generation depends on the product's existing information state. Even after accounting for purchase incidence, reviewer response remains highly state dependent. In particular, we document a pronounced first-and-early review barrier: solicitations have little effect on review incidence when a product has no prior reviews, even though an additional review is most valuable to future buyers precisely in that state. This pattern highlights frictions in early review formation, helping to explain why review solicitation may fail to improve downstream outcomes, and underscores the importance of interventions that directly address these barriers.

Second, we contribute to the literature on firms' interventions in review systems. A growing body of research examines firm interventions that shape review dynamics, including managerial responses to reviews (e.g., Proserpio and Zervas, 2017; Chevalier et al., 2018; Wang and Chaudhry, 2018), incentive schemes (e.g., Khern-am nuai et al., 2018; Sun et al., 2017; Burtch et al., 2018; Woolley and Sharif, 2021), and 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 solicitation messages, which have been shown to reliably increase review incidence and reduce extremity bias (e.g., Karaman, 2021; Brandes et al., 2022; Gao et al., 2025). More closely related to our setting, Fradkin and Holtz (2023) examine the downstream market effects of incentivized review solicitation for Airbnb stays with no prior reviews. They find that the treatment generates additional, more negative reviews but does not increase sales or revenue, which they attribute in part to supply constraints and platform design. Karaman (2025) shows that soliciting consumers to post reviews reduces those same consumers' subsequent spending on the solicited brand. We extend this literature in two ways. First, we examine solicitation messages in a setting largely free from supply or capacity constraints, showing that even unincentivized interventions can meaningfully affect the sales and returns of future buyers on average. Second, and more importantly, we show that these downstream effects are strongly state dependent: additional reviews are most valuable when prior review information is scarce, yet solicitation is least effective in those high-value states because of a pronounced first-and-early review barrier. This reviewer–buyer misalignment has not been documented in prior work and carries direct implications for the design of review solicitation strategies.

Finally, this paper contributes to the important yet relatively understudied literature on product returns and the firm-side levers that shape them. Prior work has examined how shipping and return policies influence return behavior and can reduce return costs (e.g., Wood, 2001; Anderson et al., 2009; Janakiraman et al., 2016; Shehu et al., 2020). We extend this literature by identifying review solicitation as a distinct and relatively light-touch firm intervention that operates through information. Compared with increasing shipping fees or tightening return policies, review solicitations are less likely to trigger negative consumer reactions, yet can meaningfully affect return rates by improving consumers' pre-purchase expectations. Related, Sahoo et al. (2018) show that richer review information reduces returns by mitigating uncertainty. Building on this insight, we study review solicitation as a firm-side lever that shifts the supply of review information and decompose its effects into reviewer response and the informational value to future buyers, conditional on the existing review environment. Our results reveal a reviewer–buyer misalignment that is particularly pronounced for returns when prior information is scarce. While firms may have incentives to distort reviews, for example through fake reviews, our findings highlight a countervailing force: firms can instead use marketing interventions to reduce informational bias and correct mismatches, especially when returns impose substantial costs on both consumers and firms.

The remainder of the paper proceeds as follows. Section 2 describes the data and the natural variation in review solicitation in our setting. Section 3 presents the empirical framework for estimating how review solicitations affect reviewer behavior and how the resulting reviews influence subsequent market outcomes. Section 4 reports the main results on reviewer response and the informational value to future buyers and documents the resulting reviewer–buyer misalignment. Section 5 presents counterfactual simulations that quantify the revenue implications of improving reviewer–buyer alignment. Section 6 concludes with implications for the design and management of review systems.

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, including purchases, returns, and review behaviors. The data are organized as follows:

1. Transaction Data: Each observation corresponds to an order and includes an order ID, user (buyer) ID, product ID, transaction price, and purchase timestamp.[3] For each transaction, we also observe a solicitation indicator (equal to one if a review was solicited), a review indicator (equal to one if a review was posted), and a return indicator (equal to one if the product was subsequently returned within the seven-day post-delivery eligibility period).
2. Review Data: For transactions with posted reviews, we observe the buyer's rating on a five-star discrete scale, the length of the review text, an indicator for photo inclusion, and the review timestamp.
3. Product Data: Regarding product attributes, we observe the name, category, launch date, and, importantly, the cumulative number of reviews posted at each of the five rating levels at a given product-day. These data allow us to construct key review metrics visible to buyers at the time of purchase, including the total number of reviews, average rating, and share of five-star ratings.
4. User Data: This dataset contains user-level characteristics, including recency, frequency, and monetary (RFM) metrics from the pre-sample period. Specifically, for 2018, we observe the number of orders, total purchase amount, and purchase dates for users who made purchases in the year prior to our analysis period.

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

Several unique features of our data make them particularly well suited to addressing our research questions. First, unlike most scraped datasets, our data include information on non-reviewed transactions (i.e., transactions that did not result in reviews). This information is critical for addressing self-selection in online reviews (Dellarocas and Wood, 2008; Hu et al., 2009), whereby transactions

[3] In our data, each order contains a single product. Purchases involving multiple products are recorded as separate orders.

associated with extremely positive or negative experiences are more likely to generate reviews than those with moderate experiences, a pattern commonly referred to as the J-curve. Observing both reviewed and non-reviewed transactions allows us to account for this selection and to examine how marketing interventions, such as review solicitations, influence review participation.

Second, we observe individual-level purchases as well as product returns, which is crucial for quantifying the informational value of reviews. When changes in review information affect not only sales but also returns, they reveal how reviews shape consumers' expectations about product match and quality uncertainty, and how those are realized after purchase. Returns therefore provide a direct and observable measure of how review-driven expectations influence consumer decisions and, in turn, retailer outcomes. Third, our data contain natural variation in review solicitation that exogenously shifts product review states (see Subsection 2.4 for details). Because reviews are inherently endogenous, for example, products experiencing positive demand shocks tend to generate both higher sales and more reviews, this exogenous variation allows us to credibly 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 orders, 13,031 reviews, and 6,856 returns by 60,614 users. Additional details on sample construction are provided in Appendix A.

2.2 Review Environment

2.2.1 Review Information Displayed to Buyers

Consumers considering a purchase can view a summary of review information prominently displayed on each product detail page. Figure 1 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 positive ratings (4 and above), and the cumulative number of reviews posted for each of the five rating levels (as shown on the right panel of Figure 1). Below this summary section, consumers can scroll down further to read individual reviews in detail.

Figure 1: Review summary displayed to buyers (translated from Korean-language screenshot)

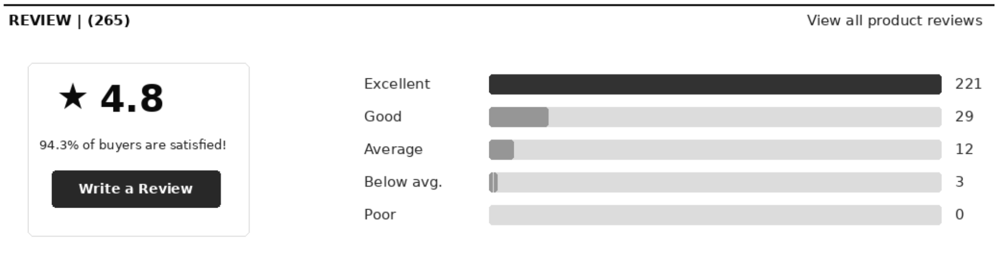
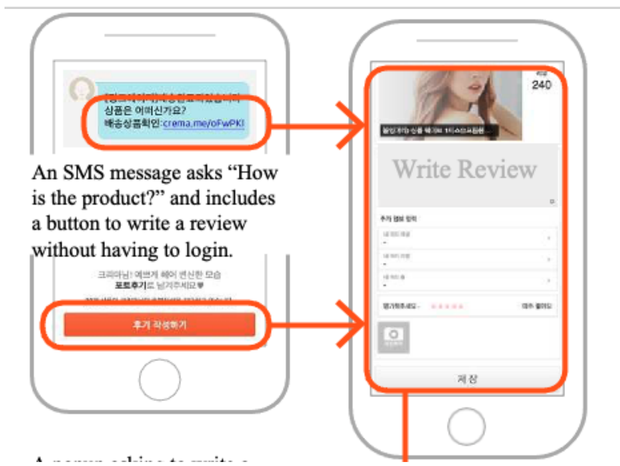


Figure 2: Review solicitation via SMS message



2.2.2 Review Writing Process

Consumers may post a review after purchasing and receiving a product. Upon delivery, the retailer sends an unincentivized SMS message, without monetary rewards or promotional content, to notify customers that the product has arrived and to facilitate review submission. As shown in Figure 2, the message typically includes a brief prompt such as “Your product has been delivered. How is the product?” together with a button that links directly to the review-submission page, without requiring the user to log in to the retailer’s website.[4] Consumers may also submit reviews by logging in directly to the retailer’s website after receiving their orders. When posting a review, consumers are required to provide both a star rating on a five-point scale and a written comment.

[4]When multiple products are delivered, the consumer receives a single SMS message. The linked review-submission interface then displays all delivered products, allowing the consumer to submit a separate review for each product.

They may also optionally upload photos and provide additional information, such as height, weight, and usual clothing size. Together, these elements constitute the individual reviews displayed on the product detail page.[5]

2.3 Data Description

Table 1: Summary statistics

	Mean	SD	Min	25%	50%	75%	Max
User-level activities							
Number of orders	2.34	1.90	1	1	2	3	20
Number of reviews	0.21	0.76	0	0	0	0	15
Product-week flow metrics							
Number of orders	45.34	57.65	0	9	25	59	632
Return rate (%)	5.47	8.02	0.00	0.00	3.51	7.69	100.00
Number of reviews	4.07	6.00	0	0	2	5	69
Five-star rating share	89.40	20.37	0.00	85.71	100.00	100.00	100.00
Product-day 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
Five-star rating share	90.39	6.97	0.00	87.45	90.91	93.55	100.00
Rating variance	0.18	0.097	0.00	0.14	0.19	0.25	0.56

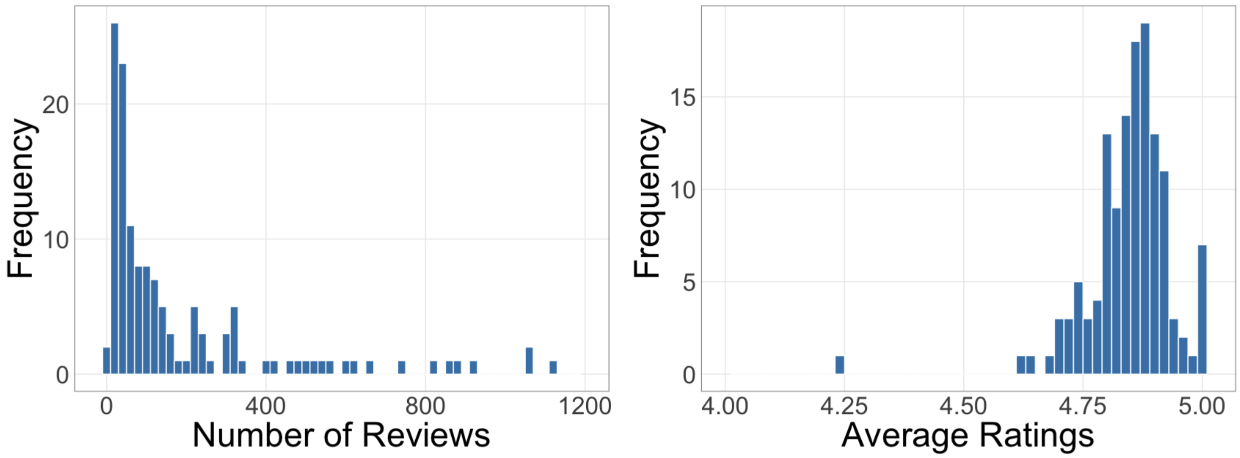
Table 1 reports summary statistics for user-level activity, product-week flow outcomes, and cumulative review information. The top panel shows that users place 2.34 orders on average during the sample period but write only 0.21 reviews, implying that only about 9% of purchases result in a review.[6] Given the small number of orders per user, our identification relies primarily on cross-sectional variation across users, leaving limited scope to analyze within-user effects. The second panel reports product-week flow metrics. On average, a product receives 45.34 orders per week, of which 5.47% are returned, with substantial heterogeneity across products.[7] Products also receive

[5] 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, transaction price, 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 marketplace platforms like Amazon, who face intense competition and may use inflated ratings to gain visibility or outrank rivals (Gandhi et al., 2025).

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

[7] The retailer in our study adopts a return policy that is typical among South Korean e-commerce platforms.

Figure 3: Distribution of number of reviews and average ratings (December 31, 2019)



(a) Number of reviews **(b)** Average ratings

4.07 reviews per week on average.

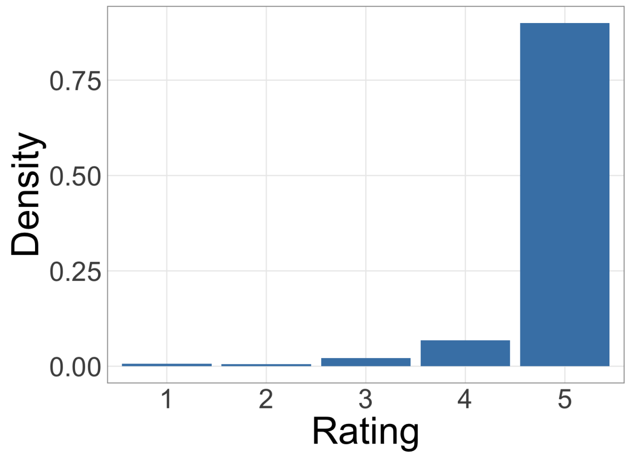
The third panel reports review information aggregated to date, corresponding to the review summary displayed to buyers on the product detail page (Figure 1). Figure 3 provides a closer view of the cross-sectional distribution of this information using a snapshot from December 31, 2019. The left panel shows the distribution of the number of reviews across products. While the mean is 199, the distribution exhibits substantial dispersion and is heavily right-skewed. Many products have very few reviews, highlighting the scope for marketing interventions, such as review solicitations, to encourage buyers to share feedback, while also raising the question of whether such interventions are effective.[8] The right panel shows the distribution of product-level average ratings. The mean average rating is 4.86, and the distribution is highly concentrated near the upper end of the scale. This pronounced negative skew suggests strong reviewer self-selection and underscores the importance of accounting for such selection when examining how review solicitations affect market outcomes.

Figure 4 shows the distribution of ratings across all individual reviews in our sample, 90% of which are five-star ratings. This empirical pattern has two important implications for our analysis. First, when examining individual-level rating behavior, we model whether a consumer gives a five-

Returns are permitted only within seven days of product delivery, and buyers are responsible for both the original shipping and return shipping costs unless the product is defective. These costs amount to approximately 23% of the average product price in our sample, which likely contributes to the relatively low average return rate.

[8]Similar to buyers, we cannot distinguish between solicited and unsolicited reviews in the cumulative review count, because the solicitation status of reviews written prior to the sample period is unobserved.

Figure 4: Distribution of individual review ratings



star rating, rather than modeling the choice among all discrete rating levels. This specification is appropriate because most of the variation in ratings arises from whether a review is five-star or not.

Second, unlike the well-known J-shaped distribution documented in prior work (Hu et al., 2009; Dellarocas and Wood, 2008), the individual ratings in our data follow an exponential pattern, as also observed in Sunada (2025). The highly negatively skewed, exponential-shaped distribution of individual ratings implies that, when aggregated to the product level, lower average ratings are associated with higher rating variance. Indeed, in our setting we observe a strong negative correlation (-0.80) between the average rating and rating variance. Accordingly, when operationalizing the information contained in existing reviews for a product, we focus on the average rating rather than considering both. Throughout our interpretation, we note that lower average ratings in our context are tightly associated with higher variance.

2.4 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.[9] 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. Review solicitations are sent for all orders delivered from the beginning of each month until the free message quota is depleted. This setup

[9]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.

creates natural, exogenous variation in which customers receive review solicitations.[10]

Figure 5: Natural experimental variation in review solicitation

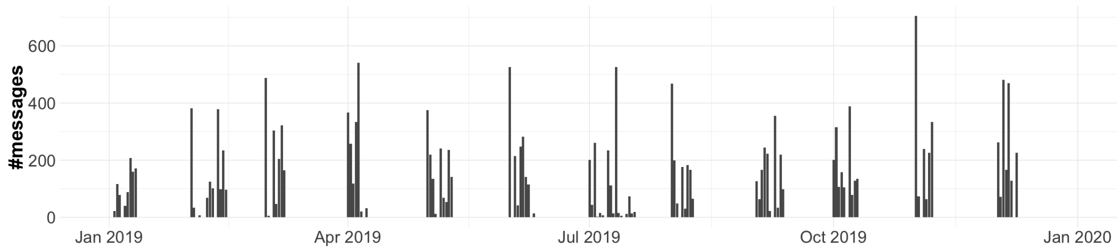


Figure 5 displays the daily number of review solicitations sent in 2019. Since review solicitations are triggered by actual orders, both the daily spikes and the days on which solicitations are sent reflect fluctuations in consumer demand. On average, about 2,600 solicitations are sent each month, typically for the first ten days.[11] Overall, 28.6% of orders receive a solicitation. At the user level, approximately 38% of buyers receive at least one review solicitation, whereas the remaining 62% receive none. 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.

3 Empirical Strategy

In this section, we present the empirical framework used to estimate the effect of review solicitations on reviewers and the subsequent demand responses generated by these reviews. In doing so, we emphasize the role of the product's existing information environment, which may give rise to heterogeneous effects. Figure B.1 in the Online Appendix summarizes the timing and structure of the framework.

[10] The review service provider sends messages for all delivered orders beginning at the start of the month, rather than spreading them randomly over the month or targeting them strategically. Based on our discussions with the provider, this practice is largely motivated by the fact that many retailers purchase additional messages once the free quota is exhausted, generating additional revenue for the provider. Although the focal retailer could allocate additional marketing budget beyond the monthly subscription fee to send more SMS review solicitations once the free allotment is exhausted, it did not do so.

[11] The monthly free quota varies slightly across months because the review service provider occasionally provides bonus messages, for instance, to compensate for system glitches during updates or performance issues in the review section.

3.1 Reviewer Model

3.1.1 Framework

We first model how post-purchase solicitations affect reviewing behavior, allowing the effect to vary with the product's existing review information. Consider a consumer i who purchases product j on day t. Prior to purchase, the consumer observes the product's existing reviews, which shape expectations about product quality and match. We summarize this information environment faced by the consumer through *review states*, defined by the interaction of review-count bins and average-rating bins. Let $S^{(k)}_{j,t-1} = 1$ indicate that product j is in review state $k \in \mathcal{K}$ during the pre-purchase period $t-1$. Specifically, review-count bins consist of a zero-review category and quintiles of the (positive) review-count distribution, while average-rating bins are defined by a median split.

The unit of analysis is user-product-day, indexed by (i, j, t). We examine four outcome variables Y_{ijt}: (i) whether a review is written, (ii) whether the rating is five stars, (iii) review text length, and (iv) whether a photo review is written. Our baseline specification is:

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

where $\mathsf{Solicited}_{ijt}$ is an order-level indicator equal to 1 if the order for product j placed by consumer i on day t receives a post-purchase review solicitation;[12] $S^{(k)}_{j,t-1}$ denotes the product's review state in period $t-1$;[13] the vector X_{jt} includes both time-varying and time-invariant product characteristics, including product-category indicators, price, product age, and an indicator for new-release status (equal to one if the product age is less than 30 days, capturing potential promotional effects for newly launched products). Finally, $\lambda_{m(t)}$ denotes month fixed effects. The coefficients of interest are γ_k, which capture how the effect of receiving a solicitation varies with the product's review state.[14] In what follows, we refer to Equation (1) as the reviewer model.

[12] The time subscript t indexes the order date, not the date on which the solicitation message is sent.

[13] In our model, the review state $S^{(k)}_{j,t-1}$ is measured prior to purchase to reflect the information available to consumers during their pre-purchase search. In our empirical setting, as is typical in e-commerce in South Korea, products are usually delivered within a few days after an order is placed. As a result, the review information observed at the time of purchase is likely to remain salient when consumers subsequently decide whether to write a review.

[14] We impose no prior restrictions on the sign or magnitude of γ_k. We interpret the estimates as capturing heterogeneity in review-writing behavior associated with the review information encountered during consumers' pre-purchase search.

3.1.2 Identification

We interpret the heterogeneous solicitation effect, γ_k, as the conditional average treatment effect (CATE) of receiving a review solicitation for consumers purchasing products in review state $S^{(k)}_{j,t-1}$.[15] Identification relies on quasi-random variation in solicitation induced by the retailer's monthly message quota. As discussed in Section 2.4, orders placed while the monthly message quota remains available are exposed to post-purchase solicitation upon delivery, whereas similar orders placed after the quota is exhausted are not. Accordingly, γ_k is identified from within-month variation in solicitation exposure among orders for products in the same pre-purchase review state, conditional on the included controls. Because the quota exhausts based on retailer-wide order flow rather than on characteristics of any specific product or purchase, this variation is plausibly orthogonal to unobserved determinants of review-writing behavior.

Formally, this interpretation requires two standard assumptions (e.g., Rubin, 1974; Rosenbaum and Rubin, 1983): conditional independence and overlap. Conditional independence requires that, given the pre-purchase review state $S^{(k)}_{j,t-1}$ and observed covariates $(X_{jt}, \lambda_{m(t)})$, solicitation assignment is independent of potential review outcomes. The overlap assumption requires that, for every review state $k \in \mathcal{K}$, the probability of receiving a solicitation is strictly between zero and one, ensuring that both solicited and non-solicited observations exist within each review state.

Because the quota threshold is determined by the retailer's subscription plan with the review service provider, rather than by targeting based on product or consumer characteristics, there is limited scope for strategic selection into the solicitation window. Nevertheless, conditional independence could be threatened if consumers who purchase earlier in the month differ systematically from those who purchase later — for instance, through payday or shopping-cycle effects that concentrate higher-engagement consumers in the early part of the month, or through promotional timing if the retailer disproportionately launches products or runs promotions during solicitation-active periods.

We provide empirical support for both assumptions in Online Appendix B. To address conditional independence, we show that solicitation assignment is uncorrelated with pre-treatment observables, including product characteristics and pre-sample user measures (Online Appendix B.1). For overlap,

[15] Because $S^{(k)}_{j,t-1}$ is a pre-treatment characteristic, γ_k captures treatment effect heterogeneity conditional on an observed moderator. This does not require the moderator itself to be exogenously assigned; rather, it requires solicitation assignment to be conditionally independent of potential outcomes given the moderator and the included controls.

because the solicitation regime applies uniformly across all products, purchases in any review state can occur in either the solicitation-active or solicitation-inactive period. We confirm this empirically in Table B.5 (Online Appendix B.2), where both solicited and non-solicited orders are observed within each review state.

3.2 Demand Responses to Newly Generated Reviews

3.2.1 Framework

We next study how reviews generated by a focal cohort of orders affect subsequent product demand and returns. The key idea is that consumers enter the focal period with a pre-existing stock of review information, summarized by the product's review state. Reviews generated by orders placed during the focal week may subsequently shift the product page's public information environment, which may in turn affect orders and return rates in the following week.[16]

To make the timing explicit, we anchor the analysis at a focal day t and define two consecutive seven-day windows:

$$\tau(t) \equiv [t,\, t+6], \qquad \tau(t)+1 \equiv [t+7,\, t+13]$$

We refer to $\tau(t)$ as the *focal week* and to $\tau(t)+1$ as the *outcome week*. For notational convenience, once the focal day t is fixed, we write τ in place of $\tau(t)$. The unit of analysis is product j at focal day t.

Let $\mathcal{O}_{j\tau}$ denote the set of orders for product j placed during the focal week τ, and let

$$O_{j\tau} \equiv |\mathcal{O}_{j\tau}|$$

denote the size of this focal-week order cohort. We define

$$N_{j\tau} \equiv N_j(\mathcal{O}_{j\tau})$$

as the number of reviews generated by orders in $\mathcal{O}_{j\tau}$, and

$$Z_{j\tau} \equiv Z_j(\mathcal{O}_{j\tau})$$

[16] The reviewer model examines how solicitation affects the focal purchaser's own review-writing behavior. The demand model, by contrast, examines how the resulting reviews affect *subsequent* buyers through changes in the product page's public information environment. Our demand analysis is thus distinct from Karaman (2025), who studies how soliciting a consumer to post a review affects that same consumer's subsequent spending on the solicited brand.

as the number of orders in that same focal-week cohort that receive a post-purchase review solicitation.[17]

We study two outcome variables measured over the outcome week $\tau + 1$:

$$Y^{(O)}_{j,\tau+1} \equiv \text{number of orders placed for product } j \text{ during } \tau + 1,$$
$$Y^{(R)}_{j,\tau+1} \equiv \text{return rate for orders placed for product } j \text{ during } \tau + 1.$$

The return rate is defined as the percentage of orders placed during $\tau + 1$ that are subsequently returned, conditional on positive order volume.[18]

For each outcome $Y_{j,\tau+1} \in \{Y^{(O)}_{j,\tau+1},\, Y^{(R)}_{j,\tau+1}\}$, we estimate the following Poisson pseudo-maximum likelihood (PPML) specification (Silva and Tenreyro, 2006):

$$\mathbb{E}[Y_{j,\tau+1}] = \exp\Bigg(\tilde{\alpha} + \sum_{k\in\mathcal{K}} \delta_k\big(N_{j\tau} \times S^{(k)}_{j,t-1}\big) + \sum_{k\in\mathcal{K}} \tilde{\beta}_k S^{(k)}_{j,t-1} + X'_{jt}\tilde{\theta} + \tilde{\lambda}_{m(t)} + \eta O_{j\tau} + \sum_{k\in\mathcal{K}} \rho_k\big(\hat{v}_{j\tau} \times S^{(k)}_{j,t-1}\big)\Bigg) \tag{2}$$

PPML naturally accommodates zero outcomes, is robust to heteroskedasticity, and is well suited to the skewed, nonnegative outcomes in our setting.

The key regressor is $N_{j\tau}$, the number of reviews generated by orders placed for product j during the focal week τ. It captures review generation by the focal-week order cohort, which may alter the product page's public information environment for subsequent consumers. Its effect is allowed to vary with the pre-existing review state $S^{(k)}_{j,t-1}$, measured on day $t-1$, immediately before the focal week begins. This timing ensures that the moderator captures the stock of review information available before the focal order cohort is realized. The vector X_{jt} includes product characteristics (category indicators, price, product age, and a new-release indicator), and $\tilde{\lambda}_{m(t)}$ denotes month fixed effects to absorb demand seasonality. We also include $O_{j\tau}$, the number of orders placed during the focal week, to control for persistence in demand across adjacent weeks. Finally, $\hat{v}_{j\tau}$ denotes the first-stage residual used in the control-function correction for the endogeneity of review generation,

[17] Because reviews are submitted after purchase and delivery, reviews generated by orders in $\mathcal{O}_{j\tau}$ may not all be visible at the start of $\tau + 1$. Accordingly, the demand model should be interpreted as estimating the effect of review generation by the focal-week order cohort as those reviews enter the public information environment over the subsequent product-week horizon, rather than the effect of reviews that are necessarily observed by all consumers at the beginning of $\tau + 1$.

[18] Since returns are permitted within seven days of delivery, realized returns for these orders may occur either during $\tau + 1$ or in the following week, depending on delivery timing.

as described in Subsection 3.2.2.

The coefficients of interest are δ_k, which capture the state-contingent effect of an additional review—identified using solicitation-driven variation in review generation by the focal-week order cohort—on next-week outcomes for products in review state k. Controlling for $O_{j\tau}$ ensures that the variation in $N_{j\tau}$ is not mechanically driven by the fact that product-weeks with more orders have more opportunities to generate reviews. In what follows, we refer to Equation (2) as the demand model.[19]

3.2.2 Control-Function Approach

A central empirical challenge is that review generation by the focal-week cohort, $N_{j\tau}$, is endogenous. For example, persistent product-level demand shocks or multi-week promotions may increase the number of orders placed during τ, the number of reviews generated by those orders, and product outcomes in $\tau + 1$. To address this concern, we instrument review generation using review solicitations and implement a control-function approach for a nonlinear outcome model (Wooldridge, 2014).

Specifically, in the first stage, review generation by the focal-week order cohort is modeled as

$$N_{j\tau} =\bar{\alpha} + \sum_{k\in\mathcal{K}} \pi_k\big(Z_{j\tau} \times S^{(k)}_{j,t-1}\big) + \sum_{k\in\mathcal{K}} \bar{\beta}_k S^{(k)}_{j,t-1} + X'_{jt}\bar{\theta} + \bar{\lambda}_{m(t)} + \bar{\eta} O_{j\tau} + v_{j\tau}, \tag{3}$$

where $Z_{j\tau}$ denotes the number of post-purchase review solicitation messages sent to orders in the focal-week cohort. The interaction terms allow the effect of solicitations on review generation to vary with the product's pre-existing review state. Including $O_{j\tau}$ in the first stage is critical because it ensures that identification comes from variation in solicitation exposure among focal-week orders, rather than from variation in the number of focal-week order opportunities themselves.

Let $\widehat{N}_{j\tau}$ denote the fitted value from Equation (3). The first-stage residual is then constructed as

$$\hat{v}_{j\tau} = N_{j\tau} - \widehat{N}_{j\tau}. \tag{4}$$

[19] We model demand at the product level rather than the individual level because newly generated review information alters the public information environment of a product page, thereby affecting prospective buyers collectively. The treatment variation—review generation by the focal-week order cohort—is product-time specific, and the corresponding managerial outcomes, orders and returns, are measured at the same level.

Thus, $\hat{v}_{j\tau}$ captures the component of review generation by the focal-week order cohort unexplained by solicitation exposure among that cohort, the pre-period review state, focal-week orders, and the other included controls. In the second-stage demand model, we include

$$\sum_{k\in\mathcal{K}} \rho_k \big(\hat{v}_{j\tau} \times S^{(k)}_{j,t-1}\big)$$

which mirrors the heterogeneous structure of the second-stage treatment effect and allows the control-function correction to vary with the product's pre-existing review state.

3.2.3 Validity of the Instrument

We use solicitation exposure among orders placed during the focal week τ as an instrument for the number of reviews generated by that order cohort. Consistent estimation of the state-contingent causal effect δ_k via the control-function approach requires three conditions: relevance, exclusion, and conditional exogeneity of the instrument, each conditional on focal-week order volume $O_{j\tau}$, the pre-period review state, product characteristics, and month fixed effects.

Relevance. The first-stage results, reported in Subsection 4.1.1 and Online Appendix C.1, confirm that solicitation exposure among focal-week orders significantly increases review generation by those orders.

Exclusion restriction. The exclusion restriction requires that solicitation exposure among orders placed during τ affect next-week outcomes $Y_{j,\tau+1}$ only through their effect on review generation by those orders. In other words, conditional on focal-week order volume and the other controls, receiving a solicitation without writing a review should not directly affect subsequent product orders or return rates.

The main concern is that solicitation messages may serve as reminders of the retailer, thereby stimulating additional visits or purchases independent of review generation. Several institutional features mitigate this concern. First, all consumers, whether solicited or not, receive product delivery, so any reminder effect associated with delivery itself should be common across the two groups. Relative to this common delivery experience, the solicitation message constitutes only a brief prompt to submit a review. Second, the solicitation message contains neither monetary incentives nor promotional content. Third, the message directs users to a standalone review-submission interface rather than back to the retailer's website (Figure 2), limiting the scope for solicitation-induced browsing or purchasing. Fourth, customers in our sample make only about 2.3 purchases on average over the year, with purchases typically spaced months apart, suggesting that any residual reminder effect on repurchase is likely negligible over a one-week outcome horizon.

A separate concern is that solicited reviews may carry a different signal value for subsequent buyers. In our setting, however, buyers cannot distinguish whether a posted review was solicited or unsolicited, which rules out differential signaling effects of this kind on the demand side.

Finally, solicitation could affect not only the number of reviews generated but also their composition, for example by shifting rating valence or review richness. In that case, the instrument could affect subsequent outcomes through review attributes other than count. Our reviewer model results indicate that these compositional effects are economically small, especially after aggregation to the product-level information environment observed by future buyers: solicitation primarily increases review volume, while its effects on product-level average rating, text length, and photo volume are modest or negligible. Thus, the dominant channel through which solicitation changes the public information environment in our setting is review generation itself.

Conditional exogeneity. Conditional exogeneity requires that, after conditioning on the controls in the first-stage equation, the instrument be uncorrelated with unobserved determinants of next-week demand and returns. As discussed in Section 2.4, solicitation assignment is governed by the retailer's monthly message quota rather than by product- or consumer-level targeting, which provides the institutional basis for this assumption.

A key complication is that the number of solicitations sent for product j for orders placed in week τ is mechanically related to the number of orders placed during the same period, because solicitations are triggered by orders. This creates a potential threat to conditional exogeneity: if demand is serially correlated, products with more orders in τ may also have both more solicitations in τ and higher demand in $\tau + 1$. We address this issue by conditioning directly on focal-week orders $O_{j\tau}$ in both the first and second stages. Conditional on $O_{j\tau}$, the identifying variation comes from variation in solicitation exposure among focal-week orders, rather than from differences in the number of order opportunities or the level of product demand itself.

After further conditioning on product characteristics, month fixed effects, and the pre-period review state, identification comes from within-month variation in solicitation intensity among comparable focal-week order cohorts. Because this variation is generated by the retailer-wide quota rather than by characteristics of any specific product or purchase, the residual variation is plausibly orthogonal to unobserved demand shocks affecting next-week demand and returns. We provide empirical support through balance tests (Online Appendix B.1) and a series of robustness checks discussed in Subsection 4.4.

4 Results

We now present the empirical results. We begin with the demand model, which identifies where an additional review has the greatest impact on orders and return rates. This provides a benchmark for where review information is most valuable to future buyers. We then turn to the reviewer model, which identifies where solicitations are most effective at generating such reviews. Comparing the two allows us to characterize the resulting reviewer–buyer (mis)alignment.

4.1 Demand Model

4.1.1 First Stage

For the demand model, we implement a control-function approach in which solicitation exposure serves as an instrument for review supply. The first stage models weekly review generation ($\#\text{Reviews}_{j\tau}$) as a function of weekly solicitation exposure ($\#\text{Solicitations}_{j\tau}$). Because solicitations are triggered by orders, we control for contemporaneous order volume, $O_{j\tau}$, so that identification comes from variation in solicitation exposure among focal-week orders rather than from variation in order volume itself. For notational simplicity, we omit the product subscript j hereafter.

Column (1) in Table 2 reports the first-stage estimates. The coefficient on $\#\text{Solicitations}_{\tau}$ is positive and highly significant, and the excluded-instrument F-statistic is 377, indicating strong instrument relevance. In magnitude terms, 100 additional solicitation messages generate about 1.63 additional reviews for a given product within the focal week, conditional on contemporaneous orders, product attributes (price, product category, product age, and new-release status), review state indicators, and month fixed effects. Thus, solicitation exposure provides a strong and economically meaningful source of variation in review supply.

Table 2: Demand model: first-stage and average second-stage estimates

Dependent variable:	#Reviews$_{\tau}$	#Orders$_{\tau+1}$	Return rate$_{\tau+1}$
Column:	(1)	(2)	(3)
Panel A. First stage			
#Solicitations$_{\tau}$	0.0163***		
	(0.0043)		
Panel B. Second stage			
#Reviews$_{\tau}$		-0.0091	-0.0845***
		(0.0227)	(0.0300)
First-stage residual (control function)		0.0059	0.0744**
		(0.0262)	(0.0295)
Product attributes	Yes	Yes	Yes
Review states	Yes	Yes	Yes
#Orders$_{\tau}$	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Dep. var. mean	4.07	45.34	5.47
N	20,276	20,276	19,709
R^2	0.82205		
Pseudo R^2		0.57974	0.07176
First-stage F-statistic (excluded IV)	377.0		

Notes: Column (1) reports the OLS first-stage estimates using #Solicitations$_{\tau}$ as the excluded instrument. Columns (2) and (3) report PPML second-stage estimates with a control-function correction based on the first-stage residual. Observations with zero orders in week $\tau + 1$ are excluded in column (3). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.1.2 Second Stage

We next examine how review supply affects downstream product orders and returns. Columns (2) and (3) in Table 2 report the second-stage PPML control-function estimates for subsequent orders and return rates. These specifications correct for endogenous review generation using the first-stage residual, while controlling for product attributes, review state indicators, contemporaneous orders, and month fixed effects.

The average effect of an additional review on subsequent orders is small and statistically indistinguishable from zero. By contrast, additional reviews significantly reduce return rates. The coefficient in Column (3) is -0.0845, which implies an 8.1% reduction in the next week's return rate ($\exp(-0.0845) - 1 \approx -0.081$), corresponding to a 0.44 percentage-point decline from the baseline mean of 5.47%. These average-effect results suggest that reviews create substantial value by improving buyers' information about product fit, which in turn reduces ex post corrections through returns. The implied savings can be economically substantial given the shipping and handling costs, customer service expenses, and potential reputational costs associated with product returns.

The control-function residual further indicates that the endogeneity correction is quantitatively important. In the return-rate specification, the residual is positive and statistically significant, implying that unobserved shocks that increase review generation are also associated with higher subsequent returns. Absent this correction, the estimated return-reducing effect of reviews would be biased toward zero.

Heterogeneity by product's existing review state. The marginal effect of an additional review is likely to depend on how much review information is already available. When existing information is scarce, an additional review can meaningfully reduce uncertainty about the product, whereas its incremental information value may diminish as reviews become abundant.

To examine such heterogeneity, we interact both the review count $N_{j\tau}$ and the control-function residual $\hat{v}_{j\tau}$ with indicators for the product's pre-existing review state $S_{j,t-1}^{(k)}$, as specified in Equation (2). Review states are defined by a zero-review category and quintiles of the positive review-count distribution, where Q1–Q5 count bins correspond to $(0, 21]$, $(21, 58]$, $(58, 136]$, $(136, 329]$, and $(329, +)$, respectively. In the first stage, the excluded instrument is similarly interacted with the review state indicators (i.e., $Z_{j\tau} \times S_{j,t-1}^{(k)}$) in constructing the control-function residual. The corresponding first-stage estimates and joint instrument-relevance statistics are reported in Table C.2 (Online Appendix C.1).

In richer specifications, we further split each count bin into low- versus high-rating groups at the median rating of 4.86. This additional dimension is empirically informative because, in our setting, review volume and rating capture distinct dimensions of the information environment.[20]

Table 3 reports the resulting second-stage estimates. Columns (1) and (3) define review states by review-count bins only, whereas Columns (2) and (4) further split those bins by rating.[21] For orders, an additional review has the largest effect when existing review information is sparse. In Column (1), one more review increases next-week orders by about 4.5% in the zero-review state, 3.1% in Q1, 4.9% in Q2, and 2.8% in Q3.[22] The effect becomes statistically indistinguishable from zero in Q4 and turns negative in Q5. The rating-split specification in Column (2) reveals that this negative effect is concentrated in the low-rating group. This pattern is consistent with reviews functioning

[20]The correlation between average rating and review count is negligible (0.011), confirming that the rating split is not redundant with the count bins and also that low-rated products are not simply products with few reviews.

[21]The results are qualitatively similar when information states are defined using text length and photo volume as well. We therefore adopt review count and rating as parsimonious summary state variables.

[22]Because the models are estimated by PPML, a coefficient β implies an approximate $100 \times (exp(\beta) - 1)\%$ change in the expected outcomes from one additional review.

Table 3: Demand model: effect of an additional review on orders and return rates by existing review state

Dependent variable:	$\#\text{Orders}_{\tau+1}$		Return $\text{rate}_{\tau+1}$	
Column:	(1)	(2)	(3)	(4)
Marginal effect of reviews				
$\#\text{Reviews}_\tau \times I$(zero reviews)	0.0438***	0.0476***	−0.1271***	−0.1227***
	(0.0159)	(0.0174)	(0.0387)	(0.0382)
$\#\text{Reviews}_\tau \times I$(Q1 count)	0.0301**		−0.1132***	
	(0.0143)		(0.0343)	
$\times I$(Low rating)		0.0333*		−0.0944***
		(0.0191)		(0.0362)
$\times I$(High rating)		0.0360**		−0.1205***
		(0.0172)		(0.0364)
$\#\text{Reviews}_\tau \times I$(Q2 count)	0.0478***		−0.0893***	
	(0.0146)		(0.0294)	
$\times I$(Low rating)		0.0570***		−0.0831***
		(0.0180)		(0.0301)
$\times I$(High rating)		0.0507***		−0.0871***
		(0.0182)		(0.0311)
$\#\text{Reviews}_\tau \times I$(Q3 count)	0.0274**		−0.0656**	
	(0.0129)		(0.0256)	
$\times I$(Low rating)		0.0419**		−0.0515*
		(0.0164)		(0.0265)
$\times I$(High rating)		0.0266*		−0.0707***
		(0.0154)		(0.0268)
$\#\text{Reviews}_\tau \times I$(Q4 count)	−0.0119		−0.0897***	
	(0.0088)		(0.0285)	
$\times I$(Low rating)		−0.0202		−0.0767***
		(0.0137)		(0.0281)
$\times I$(High rating)		0.0097		−0.0951***
		(0.0158)		(0.0296)
$\#\text{Reviews}_\tau \times I$(Q5 count)	−0.0270**		−0.0658**	
	(0.0108)		(0.0307)	
$\times I$(Low rating)		−0.0351**		−0.0557*
		(0.0157)		(0.0309)
$\times I$(High rating)		−0.0059		−0.0689**
		(0.0142)		(0.0305)
First-stage residual (control function)				
Residual $\times I$(zero reviews)	−0.0222	−0.0240	0.1506***	0.1463***
	(0.0249)	(0.0264)	(0.0555)	(0.0551)
Residual $\times I$(Q1 count)	−0.0357**		0.0929**	
	(0.0168)		(0.0411)	
$\times I$(Low rating)		−0.0291		0.1000**
		(0.0229)		(0.0487)
$\times I$(High rating)		−0.0440**		0.0829*
		(0.0211)		(0.0462)
Residual $\times I$(Q2 count)	−0.0412**		0.0897***	
	(0.0191)		(0.0337)	
$\times I$(Low rating)		−0.0288		0.0779**
		(0.0253)		(0.0382)
$\times I$(High rating)		−0.0557**		0.0938**
		(0.0220)		(0.0380)
Residual $\times I$(Q3 count)	−0.0527***		0.0429	
	(0.0181)		(0.0288)	
$\times I$(Low rating)		−0.0616***		0.0540
		(0.0207)		(0.0337)
$\times I$(High rating)		−0.0561***		0.0224
		(0.0217)		(0.0303)
Residual $\times I$(Q4 count)	−0.0073		0.0893***	
	(0.0129)		(0.0288)	
$\times I$(Low rating)		0.0209		0.0677**
		(0.0187)		(0.0317)
$\times I$(High rating)		−0.0576**		0.0993***
		(0.0224)		(0.0326)
Residual $\times I$(Q5 count)	0.0430***		0.0567*	
	(0.0130)		(0.0336)	
$\times I$(Low rating)		0.0408**		0.0468
		(0.0170)		(0.0364)
$\times I$(High rating)		0.0123		0.0646*
		(0.0169)		(0.0340)
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
$\#\text{Orders}_\tau$	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	45.34	45.34	5.47	5.47
N	20,276	20,276	19,709	19,709
Pseudo R^2	0.61894	0.62932	0.07541	0.07687
First-stage F-statistic (excluded IV)	126.5	79.7	126.5	79.7

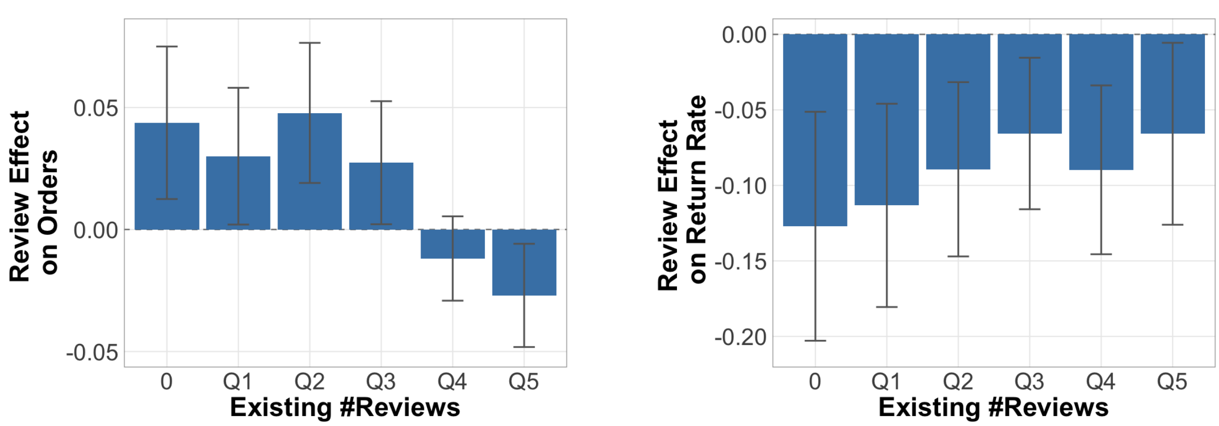
Notes: All columns use $\#\text{Solicitations}_\tau$ as the excluded instrument in the first stage. Columns (1) and (3) define review states using review-count bins only. Columns (2) and (4) further split each review-count bin by low versus high rating. Observations with zero orders in week $\tau + 1$ are excluded in columns (3)–(4). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

as information signals: rating captures whether the signal is favorable or unfavorable, while review count captures the precision of that signal. When little information is available, early reviews materially update buyers' beliefs and raise demand. Once many reviews have accumulated, however, one more review adds relatively little new information. When the existing signal is unfavorable, a further review can even reduce demand by making an adverse posterior more precise.

For returns, the pattern is stronger, and the heterogeneity is even more pronounced. In Column (3), one additional review reduces the next week's return rate by about 11.9% in the zero-review state and 10.7% in Q1. The effect attenuates thereafter, but remains economically meaningful throughout the remaining states, ranging from about 6.3% to 8.6% in Q2-Q5. In the rating-split specification in Column (4), the coefficients remain negative in every rating cell, again with the largest magnitudes concentrated in the zero- and early-review states. These results reinforce the interpretation that reviews create value primarily by improving information about product fit and thereby reducing ex post mismatch.

The interacted residual terms are statistically significant in several states as well, especially in the return specifications, indicating that endogeneity correction matters not only on average but also for the state-dependent effects that are central to our interpretation.

Figure 6: Heterogeneous effect of an additional review on orders and return rates



(a) Effect on orders **(b)** Effect on return rates

Notes: Error bars represent 95% confidence intervals.

Figure 6 visualizes these heterogeneous effects reported in Table 3. The pattern highlights that the economic value of an additional review is highly state dependent, with first and early reviews carrying disproportionately greater value than later ones, especially by improving buyers' information about product fit and reducing costly mismatches that later appear as product returns.

Mechanism test: new versus existing buyers. This interpretation regarding the informational value of reviews is further sharpened when we distinguish between new and existing buyers. If reviews primarily serve to inform buyers about product fit, their effects should be larger for those with less private information. We test this by re-estimating the heterogeneous demand model separately for new buyers—defined as customers with no prior purchase history with the retailer and thus unable to draw on past experience to assess fit—and existing buyers, who have previously purchased. The full results are reported in Appendix C.4.

The findings strongly support an informational mechanism. The effect of additional reviews on product returns is substantially larger for new buyers than for existing buyers, especially in zero- and low-review states. In the zero-review state, one additional review reduces return rates by about 15.6% for new buyers, whereas the effect for existing buyers is statistically indistinguishable from zero. This contrast persists in Q1. This pattern is exactly what we would expect under an informational channel: buyers who lack a prior basis for evaluating fit benefit more from incremental review information.

Moreover, the same state-dependent heterogeneity pattern documented earlier in the main specification is preserved within each buyer type. For both new and existing buyers, the marginal effect of an additional review is greatest when prior review information is scarce and attenuates as the review base thickens. This rules out a compositional explanation in which the main results are driven by different buyer types sorting into different review states.

4.2 Reviewer Model

We now turn to the reviewer side. The central managerial question is whether solicitations generate reviews where they are most valuable to future buyers, thereby improving downstream market outcomes. We begin with review incidence, which captures whether solicitations induce customers to post a review at all (extensive margin), with a particular focus on zero- and early-review states where existing information is limited. We then examine review rating and two measures of information richness—text length and photo inclusion—to assess whether solicitations also change the valence and richness of the resulting reviews. If solicitations primarily affect review participation, the linkage between the reviewer and demand models operates through the volume of available information. If solicitations also materially alter ratings or review richness, then they influence not only the quantity of reviews but also their informational content.

Table 4 reports the average effects of solicitation on the four reviewer outcomes: review incidence (whether a buyer writes a review), five-star indicator (whether the posted review is rated five stars), text length, and photo review incidence (whether a buyer writes a review with photos). The treatment indicator, *Solicited*, equals one if the order received a post-purchase review request. Each column presents the preferred specification, which includes product attribute controls, review state indicators, and month fixed effects; the estimates are stable both in magnitude and significance across alternative specifications that omit subsets of these controls.

Table 4: Reviewer model: average effects of solicitation on reviewer outcomes

Dependent variable:	Review incidence	Five-star indicator	Text length	Photo review incidence
Column:	(1)	(2)	(3)	(4)
Solicited	0.0114***	-0.0156**	-1.400**	9.19×10^{-5}
	(0.0022)	(0.0062)	(0.6195)	(0.0009)
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	0.09	0.90	45.95	0.02
N	142,079	13,031	13,031	142,079
R^2	0.00830	0.00713	0.05206	0.00111

Notes: Columns (2) and (3) are conditional on writing a review. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

On average, solicitations operate primarily on the extensive margin of reviewing. Receiving a solicitation increases the probability of posting a review by 1.14 percentage points, representing a 12.4% increase relative to the 9.17% baseline. Conditional on a review being posted, solicitations slightly reduce the likelihood that the review is rated five stars by 1.56 percentage points from a baseline of 90%, suggesting that they draw in marginal reviewers who are less likely to leave highly positive ratings. They also reduce text length by only about 1.4 characters relative to a mean of 45.95, while having no statistically detectable effect on the probability that an order results in a photo review.

To interpret these reviewer side effects, it is useful to adopt the perspective of future buyers. From this standpoint, what matters is how solicitation alters the product-level information environment they observe. Two findings have first-order implications. First, solicitations primarily affect this environment by increasing review volume. This channel is likely to influence buyer decisions because review counts are prominently displayed in review summaries on product detail pages, and the 12.4% increase in review volume induced by solicitation is economically meaningful. Second, the 1.56 percentage point reduction (1.7% relative to the baseline) in the probability of a five-star rating appears small at the individual-review level, and its impact on product-level average ratings over short horizons (weekly in our demand analysis) is even smaller, because such averages are computed over the full stock of accumulated reviews. These patterns motivate the demand model's focus on review count as the primary measure of review supply in the focal week ($N_{j\tau}$), while retaining attention to cumulative review volume and average rating—both prominently displayed in the review summary—as key dimensions of the buyers' pre-existing information environment ($S^{(k)}_{j,t-1}$).[23]

Heterogeneity by product's existing review state. Table 5 examines how reviewer response varies with the existing review state, which reflects the product information the reviewer observed at the time of purchase. We employ the same state definition as in the demand model. Columns (1), (3), (5), and (7) allow the solicitation effect to vary with review-count bins only, whereas Columns (2), (4), (6), and (8) further split each count bin by low versus high average rating to verify that the patterns are not artifacts of pooling across rating regimes. We organize the discussion around review incidence, the primary review supply outcome linking the reviewer model to the demand model's value of information. The remaining outcomes help rule out alternative interpretations of the first-and-early review barrier we document.

The key takeaway is a pronounced *first-and-early review barrier*. In Column (1), when a product has *zero* prior reviews, the estimated lift in review incidence from a solicitation is minimal (0.40 percentage points) and statistically indistinguishable from zero. Even in Q1, the effect remains modest, at about 0.83 percentage points, and is only marginally significant ($p < 0.1$). Once a product

[23] The small reduction in text length and the null effect on photo review incidence are unlikely to materially alter the information richness perceived by buyers within a given week. Many buyers do not process all posted reviews in detail, as reading full text is costly; consequently, a 1.4 character change relative to a 46 character mean is unlikely to affect the information they extract or their downstream decisions. Likewise, solicitations do not affect the unconditional probability that an order results in a photo review, implying little change in the total volume of photos available for a given product. These considerations motivate our focus on review count as the primary measure of review supply in the focal week. For the buyers' pre-existing information environment (i.e., the review state to date, $S^{(k)}_{j,t-1}$), results remain qualitatively similar when states are also discretized using text length or photo volume, suggesting that these dimensions provide limited additional independent information. We therefore adopt review count and rating as parsimonious summary state variables.

Table 5: Reviewer model: effect of solicitation on reviewer outcomes by existing review state

Dependent variable:	Review incidence		Five-star indicator		Text length		Photo review incidence	
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solicited $\times I$(zero reviews)	0.0040	0.0040	−0.0408	−0.0408	−4.172	−4.156	−0.0037	−0.0037
	(0.0092)	(0.0093)	(0.0347)	(0.0347)	(3.281)	(3.282)	(0.0040)	(0.0040)
Solicited $\times I$(Q1 count)	0.0083*		−0.0189		−1.748		0.0029	
	(0.0049)		(0.0138)		(1.608)		(0.0021)	
$\times I$(Low rating)		0.0076		−0.0225		−0.4556		0.0023
		(0.0077)		(0.0251)		(2.339)		(0.0029)
$\times I$(High rating)		0.0086		−0.0170		−2.444		0.0032
		(0.0061)		(0.0168)		(2.072)		(0.0028)
Solicited $\times I$(Q2 count)	0.0153***		−0.0247*		−1.232		0.0000	
	(0.0048)		(0.0146)		(1.774)		(0.0017)	
$\times I$(Low rating)		0.0189***		−0.0283		−1.392		0.0002
		(0.0066)		(0.0215)		(2.134)		(0.0025)
$\times I$(High rating)		0.0118*		−0.0206		−1.043		−0.0002
		(0.0065)		(0.0183)		(2.931)		(0.0021)
Solicited $\times I$(Q3 count)	0.0173***		−0.0132		−0.5386		0.0003	
	(0.0035)		(0.0160)		(1.843)		(0.0016)	
$\times I$(Low rating)		0.0166***		−0.0179		2.789		0.0015
		(0.0052)		(0.0251)		(2.500)		(0.0023)
$\times I$(High rating)		0.0181***		−0.0081		−4.069		−0.0011
		(0.0047)		(0.0182)		(2.601)		(0.0022)
Solicited $\times I$(Q4 count)	0.0119***		−0.0396***		−3.404***		−0.0008	
	(0.0040)		(0.0099)		(1.205)		(0.0016)	
$\times I$(Low rating)		0.0112*		−0.0357**		−1.743		−0.0023
		(0.0060)		(0.0151)		(2.059)		(0.0025)
$\times I$(High rating)		0.0123**		−0.0420***		−4.476***		0.0001
		(0.0055)		(0.0143)		(1.446)		(0.0021)
Solicited $\times I$(Q5 count)	0.0076**		0.0107		0.0031		−0.0006	
	(0.0036)		(0.0115)		(0.8415)		(0.0016)	
$\times I$(Low rating)		0.0060		0.0227		0.6726		−0.0010
		(0.0065)		(0.0197)		(1.695)		(0.0027)
$\times I$(High rating)		0.0089**		0.0025		−0.4616		−0.0002
		(0.0039)		(0.0117)		(0.7612)		(0.0019)
Product attributes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	0.09	0.09	0.90	0.90	45.95	45.95	0.02	0.02
N	142,079	142,079	13,031	13,031	13,031	13,031	142,079	142,079
R^2	0.00834	0.00835	0.00796	0.00805	0.05232	0.05270	0.00115	0.00116

Notes: Columns (1), (3), (5), and (7) define review states using review-count bins only. Columns (2), (4), (6), and (8) further split each review-count bin by low versus high rating. Columns (3)–(6) are conditional on writing a review. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

moves beyond this early region, reviewer response rises substantially, peaking at 1.53 percentage points in Q2 and 1.73 percentage points in Q3, before attenuating to 1.19 and 0.76 percentage points in Q4 and Q5, respectively. The rating-split specification in Column (2) preserves this inverted-U shape across both low- and high-rating cells, indicating that the first-and-early review barrier reflects a fundamental feature of reviewing behavior common to both rating regimes.

Conditional on review submission, solicitations shift rating composition only in select states, with the largest declines in the five-star probability occurring in higher-review count bins. For example, the five-star probability falls by about 3.9 percentage points in Q4. Given the accumulated review volume in Q4, these compositional shifts are unlikely to translate into economically meaningful changes in product-level average ratings over the short horizon we study. Column (4) remains broadly consistent with Column (3), indicating that the observed pattern in five-star review propensity is also not an artifact of pooling across rating regimes.

The effects on review richness are less systematic and less precisely estimated than the effects on review participation. Several text-length coefficients are negative, but only the Q4 estimate is statistically significant, and for photo inclusion, none of the state-specific coefficients is statistically significant. Taken together, these results suggest that the first-and-early review barrier is primarily a participation barrier: solicitations are least effective at generating reviews where information is scarcest.[24]

4.3 Reviewer–Buyer Misalignment

We now bring the demand and reviewer sides together. Our central question is whether solicitations generate reviews where additional reviews are most valuable for product orders and returns. To answer this question, we juxtapose two state-contingent objects defined over the same existing-review bins (zero reviews and review-count quintiles): (i) the buyer-side value of an additional review, measured by its impact on downstream orders and returns in Table 3, and (ii) the reviewer-side response, measured by the effect of solicitation on review incidence in Table 5. For the demand effect, we translate the estimated order and return effects into the implied marginal change in net revenue from one additional review. Figure 7 visualizes these two objects side-by-side.

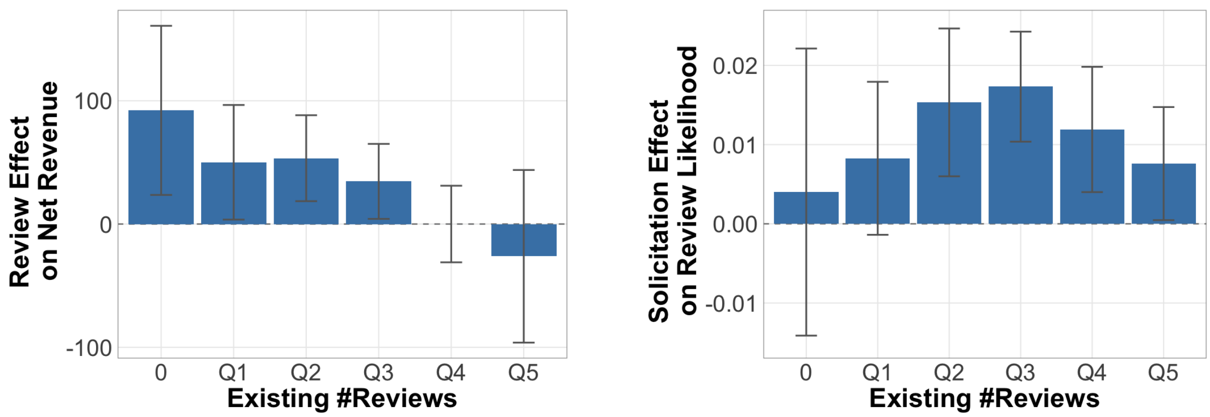
Before we interpret the comparison, it is useful to clarify the scope of the demand model estimates. We do not separately identify the effects of solicited versus unsolicited reviews on orders and return rates. Rather, we use solicitation-driven variation in *total* review supply to identify the downstream effect of an additional review.[25] Two features of the setting make this pooled interpretation plausible. First, the reviewer model results indicate that solicitations primarily affect whether a review is written, whereas observable differences in rating composition and richness are modest—so solicitation-induced reviews are not fundamentally different in kind from organic re-

[24]The largest negative point estimate for text length occurs in the zero-review state, suggesting that induced reviews may be shorter at the cold-start stage. Although this estimate is imprecise, it leaves open the possibility that successful solicitations in the earliest states generate somewhat less rich reviews. If so, the cold-start friction may be even more severe than the participation results alone suggest.

[25]Identifying a separate causal effect of organically generated reviews would require an additional source of exogenous variation in unsolicited review generation, which we do not observe.

views over the short horizon (weekly) we analyze. Second, buyers in our setting cannot distinguish solicited from unsolicited reviews when making purchase and return decisions.

Figure 7: Buyer-side value and reviewer response by existing review state



(a) Marginal net revenue of an additional review **(b)** Solicitation effect on review incidence

Note. Error bars represent 95% confidence intervals.

Figure 7 Panel (a) illustrates that buyer-side value is concentrated in low-information states. The implied marginal net revenue gain from one additional review is highest in the first-and-early region (zero reviews, Q1, and Q2), and declines sharply once products have accumulated thicker review bases (Q4 and Q5), reflecting diminishing marginal informativeness as similar signals accumulate. For example, one more review brings in about \$92.19 in implied net revenue at the zero-review state, \$50.08 at Q1, \$53.40 at Q2, falling to \$34.60 at Q3, and becoming indistinguishable from zero at Q4 and Q5.[26] This pattern is consistent with an informational mechanism: early reviews reduce ex ante uncertainty about product fit, thereby improving conversion and preventing costly mismatches that surface as returns.

Panel (b) shows a strikingly different pattern for reviewer response: a pronounced first-and-early review barrier. Solicitation effects on review incidence are weakest in zero- and early-review states and become sizable only once a minimal review base has formed. The firm's primary tool for generating reviews underperforms precisely where the marginal review information is most valuable. The juxtaposition of Panels (a) and (b) therefore reveals a systematic *reviewer–buyer misalignment*:

[26] These marginal impacts are evaluated at the mean $\#\text{Reviews}_\tau$ and corresponding mean price within each review state (\$40.52, \$42.73, \$37.65, \$32.41, \$31.64, and \$27.00), respectively.

solicitations do not generate reviews where they would most benefit future buyers.

This misalignment is unlikely to be explained solely by product traffic or by the lower purchase opportunities of cold-start products. Classic explanations emphasize that early reviews are scarce because low-traffic products generate fewer opportunities to review and that visibility dynamics create rich-get-richer patterns (Moe and Trusov, 2011; Moe and Schweidel, 2012). Our reviewer-model evidence points to something beyond those mechanisms. In the reviewer model, we compare solicited and non-solicited buyers conditional on purchase, thereby holding fixed the pool of potential reviewers. The specification also includes rich controls for product attributes, review states, and month fixed effects, which absorb observable differences and the volume of review information present across products. The first-and-early review barrier therefore appears not to be merely a byproduct of traffic dynamics, but rather to reflect frictions in review generation itself that are concentrated in low-information states.

The misalignment pattern is also informative about reviewer motivation. If reviewers were primarily motivated by the downstream value their reviews provide to future buyers, solicitations should be most effective where the demand effect is highest. Instead, reviewer response is weakest in that region. A more consistent interpretation is that, although solicitations reduce the general hassle cost of reviewing and increase participation at the margin (Karaman, 2021; Brandes et al., 2022; Gao et al., 2025), they do not eliminate state-dependent psychological frictions that make the first review especially costly to produce. When a product has very few or no prior reviews, prospective reviewers lack reference points for what constitutes a useful review for a given product or how to calibrate their own opinion, which makes writing more uncertain. Being among the first reviewers also makes a review unusually visible, heightening self-presentation and evaluation-apprehension concerns (Leary and Kowalski, 1990). Once a few reviews are already in place, these reference points and social cues become clearer, making contribution easier and more normatively natural. This interpretation is consistent with threshold models of collective behavior, in which participation rises once a minimal base of prior contributors is observed (Granovetter, 1978). It matches the pattern we observe: a pronounced barrier at zero and very low counts, followed by much stronger solicitation response once a product has crossed an initial review threshold.[27]

[27] A simple descriptive pattern in our data points in the same direction. On average, reviewers take 7.3 days after purchase to write a product's first review, compared with 6.9 days for subsequent reviews. Although modest in magnitude, this longer lag is consistent with greater hesitation or effort when buyers have no prior review benchmark

4.4 Sensitivity and Robustness

Our empirical approach relies on several assumptions related to identification and model specification. We examine these in detail in Appendix C and D and summarize them here.

4.4.1 Demand Model

Appendix C presents several robustness checks for the demand model. First, in Appendix C.1, we replace our baseline instrument—the number of solicitations sent to the focal-week order cohort—with the solicitation share, defined as the number of solicitations divided by the number of focal-week orders. The resulting estimates are qualitatively similar to those obtained under the baseline instrument and preserve the same state-contingent pattern of demand effects, especially for product returns. Second, in Appendix C.2, we compare our preferred control-function specification with an otherwise identical model that omits the residual term used to correct for endogeneity. The comparison indicates that the endogeneity correction is quantitatively important, particularly for return rates. Finally, in Appendix C.3, we report additional sensitivity analyses for the early-information states and show that the main directional pattern remains robust.

4.4.2 Reviewer Model

As a robustness check for the reviewer model estimation, we implement a local design that restricts the sample to orders placed shortly before and after the quota start and exhaustion dates within each month, as described in Appendix D.1. The results remain qualitatively similar when focusing on narrow windows around the treatment transitions (on $\rightarrow$ off and off $\rightarrow$ on). We also estimate heterogeneous treatment effects using Generalized Random Forests (Athey et al., 2019) in Appendix D.2. Allowing for flexible interactions among covariates and moderators yields estimates and first-and-early review barrier patterns consistent with our baseline results.

4.4.3 Product Fixed Effect

Our preferred specifications omit product fixed effects in both the reviewer and demand models. Including product fixed effects does not preclude identification of the state-contingent treatment effects, γ_k and δ_k. These interaction coefficients can still be identified from within-product variation while a product is observed in a given review state: in the reviewer model, from within-product

to anchor their own contribution.

variation in solicitation exposure, and in the demand model, from within-product variation in review generation, with the endogenous component addressed through the first stage of the control-function approach. Product fixed effects absorb time-invariant product-level heterogeneity, leaving identification to rely solely on within-product variation over time. In our setting, review states are highly persistent and products rarely transition across review-count bins over the horizons relevant for identification. As a result, specifications with product fixed effects draw on a much narrower set of identifying variation and therefore suffer a substantial loss of statistical power. As a robustness check, specifications with product fixed effects yield similar qualitative patterns: small solicitation effects in zero- and low-review counts in the reviewer model and large buyer-side effects in zero- or low-review states in the demand model, but with attenuated magnitudes and lower precision, consistent with the reduction in usable identifying variation. Our decision to omit product fixed effects is therefore driven primarily by statistical efficiency rather than identification.

5 Quantifying Missed Revenue

This section quantifies the economic cost of the reviewer–buyer misalignment documented above. The purpose of this exercise is therefore not to ask whether the firm should simply reallocate solicitations across products. Rather, it is to ask how much the firm would gain if it could directly reduce the first-and-early review barrier and make solicitations more effective in the zero-review state, where the marginal informational value of a new review is highest.

To answer this question, we conduct a counterfactual simulation focused on the zero-review state. We vary only one object: the effect of solicitation on review generation when a product has zero prior reviews. Everything else is held fixed, including solicitation exposure, product characteristics, prices, and the estimated buyer-side effect of reviews on orders and returns. This design isolates the value of improving review generation exactly where reviewer response is currently weakest and buyer-side value is currently strongest.

Formally, let γ denote the effect of a solicitation on review incidence in the zero-review state. In the reviewer model, this state-specific effect is estimated to be about 0.004, meaning that a solicitation increases review incidence by 0.4 percentage points when the product has no prior reviews (Table 5, column (1)). In the simulation, we vary γ over the range from 0 to 0.02 and trace out the implied market outcomes to compute the net revenue. Along this range, we highlight three benchmark values. The first is the estimated zero-review state effect, 0.004, taken from the coefficient of Solicited $\times I$(zero reviews) in Table 5, column (1). The second is the sample-average solicitation effect, 0.0114, taken from Table 4, column (1). The third is 0.0173, which corresponds to the largest heterogeneous solicitation effect reported across review-information states in Table 5, column (1), which is the coefficient of Solicited $\times I$(Q3 count).

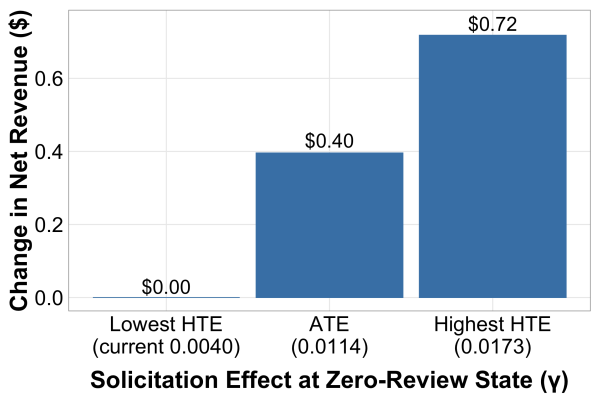
The implementation proceeds in three steps. First, for orders associated with products in the zero-review state, we replace the estimated solicitation effect from the reviewer model, $\hat{\gamma}_0$, with a counterfactual value, γ_0^{cf}. This yields a counterfactual number of additional reviews generated by solicitations in that state. Second, we feed those counterfactual reviews into the estimated demand model to map the change in review supply into subsequent orders and return rates. This step allows us to predict how more effective first-review generation would affect downstream orders and returns. Third, we convert those predicted changes into net revenue and express the result on a per-solicitation basis. Thus, the counterfactual quantifies how much additional net revenue the firm would expect from one solicitation message if solicitations were more effective at eliciting first reviews. Details of the simulation procedure and the conversion to net revenue are provided in Online Appendix E.

This exercise is intentionally static. We do not model dynamic feedback whereby additional first reviews accelerate future review accumulation or alter the rating distribution over time. As a result, the simulation should be interpreted as a conservative, local calculation: holding the broader review system fixed, what is the immediate economic value of improving solicitation effectiveness at the zero-to-one review margin? This design matches the managerial question we want to answer, whether reducing the first-review barrier is economically meaningful even before accounting for any longer-run amplification effects.

Figure 8 summarizes the counterfactual results at the three benchmark values of γ. The horizontal axis reports the solicitation effect in the zero-review state, and the vertical axis reports the corresponding change in net revenue per solicitation message. All values are normalized relative to the baseline case, $\gamma = 0.004$, so that the first bar is set to zero. The figure shows that expected net revenue increases as solicitations become more effective in the zero-review state.

Moving from the estimated zero-review effect of 0.004 to the strongest benchmark value of 0.0173 increases expected net revenue by about \$0.72 per solicitation message. We interpret this as the incremental value of overcoming the first-review barrier: it is the additional net revenue generated by one solicitation when that solicitation becomes more effective at eliciting first reviews. The gain reflects both higher downstream demand and fewer costly returns, because early reviews improve the information available to subsequent buyers. The fact that a seemingly small increase in review generation translates into meaningful revenue gains underscores how valuable an additional review is when prior information is absent.

Figure 8: Counterfactual net-revenue gains from improving solicitation effectiveness in the zero-review state



Note. The figure reports the change in net revenue per solicitation message at three benchmark values of the zero-review solicitation effect, γ. All values are normalized to zero at the baseline case, $\gamma = 0.004$.

Overall, the simulation converts the reviewer–buyer misalignment into economically interpretable units. The result does not imply that every solicitation is worth $0.72 in absolute terms. Rather, it means that improving solicitation effectiveness specifically in the zero-review state—from the currently estimated level to a substantially stronger one—would increase expected net revenue by about $0.72 per solicitation message. More broadly, the exercise highlights that the economic payoff from review generation depends not just on generating more reviews overall, but on generating them earlier, when they are most informative to future buyers. This is why interventions that reduce first-and-early review frictions can be substantially more valuable than interventions that simply increase review volume in already well-reviewed states.

6 Conclusion

This paper studies whether review solicitations generate reviews where they create the greatest value for future buyers. We show that additional reviews improve downstream market outcomes, primarily by reducing product returns. These effects are strongest when prior review information is scarce, consistent with an informational mechanism through reduced uncertainty and improved product–consumer matching. At the same time, solicitations increase review incidence on average but are least effective precisely in those low-information states. Taken together, these findings reveal a systematic reviewer–buyer misalignment: firms intervene on reviewers, yet reviews are least likely to be generated where they are most valuable to buyers.

Our evidence suggests that this misalignment reflects a first-and-early review barrier. If reviewers fully internalized the downstream value of their reviews, solicitation should be most effective when products have no or few prior reviews, since those are the states in which an additional review has the largest impact on orders and returns. Instead, reviewer response is weakest in these states, indicating that standard, unincentivized solicitation messages do not overcome the psychological frictions associated with initiating early reviews. This insight has clear managerial implications: simply reallocating solicitation efforts toward low-information products is unlikely to resolve the problem if the underlying reviewer-side frictions persist. By contrast, interventions that directly reduce first-and-early review barriers are likely to be substantially more effective.

Our counterfactual exercise quantifies the cost of this misalignment. Increasing the solicitation effect in the zero-review state to the sample-average level would raise net revenue by about \$0.40 per solicitation message, with the gain increasing to approximately \$0.72 at the largest heterogeneous solicitation effect estimated across review-information states. Even modest improvements in early review generation translate into economically meaningful gains, as early reviews carry high informational value: they expand demand and, importantly, reduce costly mismatches that manifest as product returns. Several firms have adopted interventions that directly address these frictions. For example, Home Depot's Seeds Program and Best Buy's Tech Insider Network subsidize early reviews for new or pre-release products, effectively compensating reviewers for the higher costs of initiating reviews. Beyond highlighting the importance of overcoming early-review barriers, our findings also underscore the value of improving information quality to mitigate mismatch, rather than distorting information, for example through fake reviews, particularly when returns impose substantial costs on both consumers and firms.

Our results also point to several promising directions for future research. Our study analyzes one standalone apparel retailer with unincentivized SMS solicitations. Whether the alignment patterns we document generalize to other categories, solicitation formats, or perhaps most interestingly, to platform settings where multiple sellers compete and review dynamics interact across sellers, remains an open question. Another is to investigate which interventions are most effective at overcoming first-and-early review barriers, whether through monetary incentives, reputation-based rewards, structured review prompts, or design changes to the review interface itself. More broadly, an important agenda for future work is to understand how firms can better align reviewer response with buyers' informational value over time, and how review system design interacts with consumer learning, platform policies, and other levers for managing demand and returns.

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.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

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.

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.

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.

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

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443.

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.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical* *sciences*. Cambridge university press.

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.

Leary, M. R. and Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107(1):34–47.

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.

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.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

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.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, pages 641–658.

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.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182(1):226–234.

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.

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 bundled products raise separate questions about how accumulated review information on individual product pages affects consumers' choices between bundled and individual purchases, we exclude bundled products from the 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, number of 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 ≥ 3), who are likely resellers.

A.2 Price Construction

In our analyses, we construct a time-invariant product-level price by averaging the prices paid by consumers across all orders in our sample. Our data record the transaction price paid by consumers after accounting for retailer promotions (e.g., discounts, coupons) as well as the use of store credits or rewards, which reflect consumers' own resources rather than retailer-provided price reductions. Consequently, leveraging product-day price variation (especially in specifications with product fixed effects) would make the interpretation of price coefficients or elasticities ambiguous and potentially misleading. In addition, descriptive statistics indicate minimal within-product price variation over time. Finally, we examine heterogeneity in the effect of solicitations with respect to price by interacting price with the review solicitation treatment. In this context, treating price as a time-invariant product characteristic captures the financial risk associated with potential product mismatch. Given how transaction prices (prices paid) are recorded in our data, a product-level price provides a more stable and conceptually appropriate measure for analyzing price-related heterogeneity.

B Empirical Strategy: Timeline, Balance, and Overlap

This section begins with Figure B.1, which summarizes the timing and structure of the empirical models. It then presents evidence on covariate balance between solicited and non-solicited orders and documents overlap in solicitation assignment across review states.

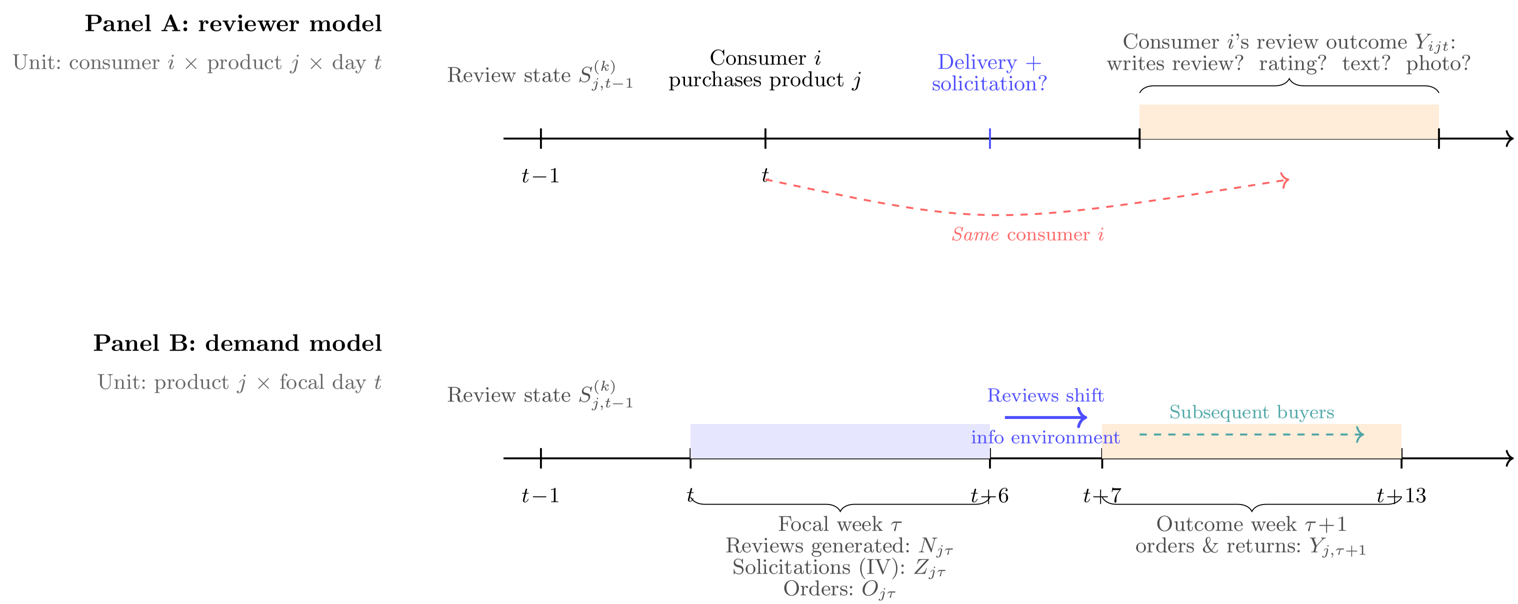
B.1 Balance on Observables

If solicitation assignment is conditionally quasi-random, solicited and non-solicited orders should be similar along pre-determined observable dimensions. We assess this in two steps. First, we compare the two groups unconditionally on product and user characteristics. Second, we examine balance within each review state using standardized mean differences, which provide the more relevant notion of balance for our heterogeneous-effect specifications.

Unconditional balance. Table B.1 compares solicited and non-solicited orders along product characteristics observed at the time of purchase: price paid, new-release status, product age, and the review information displayed on the product page (number of reviews, average rating, and rating variance). The two groups are highly similar across all dimensions. Two comparisons are particularly informative for the threats discussed in Section 3.2.3. First, the near-identical new-release shares and prices across groups suggest that the retailer does not disproportionately launch products or run promotions during solicitation-active periods. Second, the similarity in prices provides little evidence of payday-driven shopping patterns that would systematically shift the composition of products purchased across the solicitation-active and -inactive periods.

Table B.2 reports the analogous balance comparison on the user side. We compare solicited and non-solicited orders using pre-sample user characteristics constructed from 2018 purchase histories, including recency, frequency, monetary value, and return behavior, restricting attention to users for

Figure B.1: Timeline of the reviewer and demand models



Notes: Panel A illustrates the timing of the reviewer model. Consumer i purchases product j on day t after observing the pre-existing review state $S^{(k)}_{j,t-1}$. After delivery, consumer i may receive a solicitation; the outcomes are whether and how consumer i writes a review. Panel B illustrates the timing of the demand model. Reviews generated during the focal week $\tau = [t, t+6]$ change the product's public information environment, and the outcomes (orders and returns of subsequent buyers) are measured in the following week, $\tau + 1 = [t+7, t+13]$.

Table B.1: Balance on product characteristics

Group:	Solicited	Non-solicited
Price ($)	31.46 (12.63)	31.47 (12.82)
Launched within 30 days	0.13 (0.33)	0.14 (0.35)
Product age (days)	249.71 (222.03)	246.81 (222.05)
Number of reviews	250.38 (283.40)	247.07 (282.88)
Average rating	4.87 (0.083)	4.86 (0.10)
Rating variance	0.17 (0.087)	0.18 (0.093)
N	41,178	100,901

Notes: Entries are mean (SD) computed across orders.

Table B.2: Balance on pre-sample user characteristics

Group:	Solicited	Non-solicited
Recency (days)	213.46 (132.95)	219.69 (130.09)
Frequency	1.78 (1.15)	1.80 (1.18)
Monetary ($)	71.45 (59.54)	71.33 (60.02)
Return rate in 2018 (%)	3.63 (15.34)	4.48 (17.33)
N	9,620	23,860

Notes: Entries are mean (SD) computed across orders. The sample is restricted to users with at least one purchase in 2018. Recency is the number of days between the user's last 2018 order and first 2019 order. Frequency is the total number of 2018 orders. Monetary is total 2018 spending ($). Return rate in 2018 is the share of the user's 2018 orders that were returned.

whom these measures are observed. The two groups remain closely balanced across these dimensions, suggesting that solicitation assignment is not systematically related to prior user engagement, spending, or return tendencies.

Within-state balance. Tables B.3 and B.4 extend the balance analysis by comparing solicited and non-solicited orders within each existing review state using standardized mean differences (SMDs). This conditional comparison is particularly important because the heterogeneous treatment effects are identified conditional on the product's pre-period review state. Following Imbens and Rubin (2015), we compute the SMD as the difference in group means divided by the square root of the average of the two group variances. For both product characteristics and pre-sample user characteristics, the SMDs are uniformly small in magnitude and exhibit no systematic pattern across review states. Most are below 0.10 in absolute value, and all are well below the conventional 0.25 threshold for meaningful imbalance (Imbens and Rubin, 2015). These results provide additional evidence that, conditional on review state, solicitation assignment is not systematically related to observable product or user characteristics.

Taken together, the unconditional and within-state balance comparisons support the identifying

Table B.3: Standardized mean differences in product characteristics by existing review state

Review state:	Price	Product age	New release	Number of reviews	Average rating	Rating variance
Zero reviews	−0.06	0.08	0.08	—	—	—
Q1 count & low rating	−0.03	0.07	−0.16	0.13	0.19	−0.07
Q1 count & high rating	0.03	0.07	−0.13	−0.05	−0.05	0.05
Q2 count & low rating	−0.01	−0.13	0.09	0.04	0.16	−0.15
Q2 count & high rating	0.03	0.04	−0.07	0.17	0.01	−0.01
Q3 count & low rating	−0.01	0.09	—	0.09	−0.12	0.12
Q3 count & high rating	0.03	−0.11	—	0.11	−0.03	0.03
Q4 count & low rating	−0.04	−0.03	—	−0.02	0.13	−0.13
Q4 count & high rating	−0.03	0.03	—	0.13	0.04	−0.04
Q5 count & low rating	−0.00	0.04	—	−0.00	−0.02	0.02
Q5 count & high rating	0.02	0.05	—	0.01	0.02	−0.02

Notes: Each cell reports the standardized mean difference (SMD) between solicited and non-solicited orders within the indicated review state, computed as the difference in group means divided by the square root of the average of the two group variances. Values below 0.25 in absolute value indicate acceptable balance. "—" indicates that the variable is either undefined (e.g., average rating for products with zero reviews) or has no within-state variation (e.g., no new-release products in upper review-count bins).

Table B.4: Standardized mean differences in pre-sample user characteristics by existing review state

Review state:	Recency	Frequency	Monetary	Return rate
Zero reviews	0.04	−0.07	0.01	−0.14
Q1 count & low rating	−0.03	0.07	0.06	−0.02
Q1 count & high rating	0.00	−0.09	−0.07	−0.03
Q2 count & low rating	−0.14	−0.04	0.01	−0.01
Q2 count & high rating	0.07	−0.04	0.00	−0.08
Q3 count & low rating	−0.05	−0.02	0.01	−0.05
Q3 count & high rating	−0.12	0.06	0.02	−0.04
Q4 count & low rating	−0.04	0.02	0.03	−0.10
Q4 count & high rating	−0.02	−0.05	−0.01	−0.04
Q5 count & low rating	−0.00	−0.05	−0.05	−0.09
Q5 count & high rating	−0.04	0.01	0.01	−0.03

Notes: Each cell reports the standardized mean difference (SMD) between solicited and non-solicited orders within the indicated review state, computed as the difference in group means divided by the square root of the average of the two group variances. The sample is restricted to users with observed 2018 purchase histories ($n_{\text{solicited}} = 9{,}620$; $n_{\text{non-solicited}} = 23{,}860$).

assumption that solicitation assignment is conditionally quasi-random with respect to observable product and user characteristics.

B.2 Overlap Across Review States

The overlap condition requires that, within each review state $k \in \mathcal{K}$, both solicited and non-solicited observations are present, so that the state-contingent treatment effects γ_k are identified. Table B.5 reports the solicitation share, the fraction of orders receiving a solicitation, within each review state. In every state, the share lies strictly between zero and one, confirming that both solicited and non-solicited observations are present throughout the support of the review-state space. It is also reassuring that solicitation shares are fairly stable across review states, which is consistent with the absence of targeting based on review volume or ratings.

Table B.5: Solicitation shares by existing review state

Review state:	N	Solicited	Solicitation share
Zero reviews	4,285	1,625	0.38
Q1 count & low rating	8,616	2,176	0.25
Q1 count & high rating	14,437	3,821	0.26
Q2 count & low rating	10,208	3,145	0.31
Q2 count & high rating	10,634	3,039	0.29
Q3 count & low rating	14,146	3,861	0.27
Q3 count & high rating	11,338	3,579	0.32
Q4 count & low rating	11,699	3,175	0.27
Q4 count & high rating	17,098	5,213	0.30
Q5 count & low rating	17,407	5,130	0.29
Q5 count & high rating	22,211	6,414	0.29
All	142,079	41,178	0.29

C Demand Model: Sensitivity and Robustness

This appendix section presents additional sensitivity analyses for the demand model. We first examine instrument choice by comparing our preferred solicitation-volume instrument with the alternative solicitation-share instrument. We then compare how the control-function correction affects the estimates, assess robustness of the early-information state pattern, and finally study heterogeneity by buyer type to further illuminate the informational mechanism.

Throughout this section, we use the notation $\#\text{Reviews}_{j\tau}$, $\#\text{Solicitations}_{j\tau}$, and $\#\text{Orders}_{j\tau}$ to refer to the corresponding order-cohort variables defined in the main text. Specifically, let $\mathcal{O}_{j\tau}$ denote the set of orders for product j placed during the focal week τ. Then

$$\#\text{Orders}_{j\tau} \equiv O_{j\tau} = |\mathcal{O}_{j\tau}|,$$

$$\#\text{Reviews}_{j\tau} \equiv N_{j\tau} = N_j(\mathcal{O}_{j\tau}),$$

and

$$\#\text{Solicitations}_{j\tau} \equiv Z_{j\tau} = Z_j(\mathcal{O}_{j\tau}).$$

Thus, τ indexes the timing of order placement. Reviews counted in $\#\text{Reviews}_{j\tau}$ are reviews generated by orders placed during τ, and $\#\text{Solicitations}_{j\tau}$ denotes the number of focal-week orders that receive a post-purchase solicitation.

C.1 Instrument Choice and Alternative Instrument Robustness

As discussed in Section 3.2.3, a key challenge in estimating the causal effect of additional reviews on subsequent outcomes is that review generation may be correlated with unobserved product-week shocks. Promotions, stockouts, temporary quality issues, or demand shifts can affect both the number of reviews generated by a focal-week order cohort and downstream outcomes. Our control-function approach addresses this concern by using solicitation exposure among focal-week orders as an excluded source of variation in review generation.

C.1.1 Average Specification

Our baseline excluded instrument is the weekly number of solicitations sent to orders placed for product j during focal week τ, $\#\text{Solicitations}_{j\tau}$. Because solicitation exposure is mechanically related to the number of focal-week orders, we control for $\#\text{Orders}_{j\tau}$ in the first stage. Identification therefore comes from variation in solicitation exposure conditional on the size of the focal-week order cohort, rather than from variation in focal-week order volume itself.

As a robustness check, we replace the excluded instrument with the *solicitation share*,

$$\text{SolicitShare}_{j\tau} \equiv \frac{\#\text{Solicitations}_{j\tau}}{\#\text{Orders}_{j\tau}}.$$

This alternative instrument captures intensive-margin variation in solicitation exposure by measuring the fraction of focal-week orders that receive a solicitation. Both the baseline and alternative instruments address the mechanical relationship between solicitations and order volume, but they do so differently. In the baseline specification, we use solicitation volume while conditioning on $\#\text{Orders}_{j\tau}$, so identification comes from variation in solicitation exposure among focal-week order cohorts of similar size. By contrast, the solicitation-share specification normalizes solicitation exposure by $\#\text{Orders}_{j\tau}$ directly.

Baseline instrument: solicitation volume. The average first-stage results for the baseline instrument are reported in Table 2 in the main text. The coefficient on $\#\text{Solicitations}_{\tau}$ is positive and statistically significant, and the excluded-instrument F-statistic is 377.0, indicating strong instrument relevance.

Alternative instrument: solicitation share. For all specifications that use solicitation share as the excluded instrument—reported in Tables C.1, C.2 columns (3)–(4), and C.3—we exclude observations with zero focal-week orders, since the solicitation share is undefined when $\#\text{Orders}_{j\tau} = 0$. As a result, the estimation sample differs from the one used with solicitation volume as the excluded instrument.

Column (1) in Table C.1 reports the average first-stage results using solicitation share as the excluded instrument. The first stage remains positive and statistically significant. The coefficient on $\text{SolicitShare}_{\tau}$ is 0.6563, and the excluded-instrument F-statistic is 116.3, indicating that solicitation share remains a relevant instrument, albeit a notably weaker one than the baseline solicitation volume instrument. In magnitude terms, a 10-percentage-point increase in solicitation share predicts about 0.066 additional reviews generated by the focal-week order cohort, conditional on focal-week orders and the same controls used in the main specification.

Columns (2) and (3) in Table C.1 report the corresponding average second-stage estimates. The qualitative conclusions from the main specification are unchanged. The coefficient on $\#\text{Reviews}_{\tau}$ remains statistically indistinguishable from zero for next-week orders, while the return-rate effect remains negative and economically meaningful. Under the alternative instrument, the coefficient on reviews in the return-rate equation is -0.1896, which implies a 17.3% reduction in next-week return rates ($\exp(-0.1896) - 1 \approx -0.173$), corresponding to roughly a 0.95 percentage-point decline

Table C.1: Demand model using the alternative instrument (solicitation share): first-stage and average second-stage estimates

Dependent variable:	#Reviews$_\tau$	#Orders$_{\tau+1}$	Return rate$_{\tau+1}$
Column:	(1)	(2)	(3)
Panel A. First stage			
SolicitShare$_\tau$	0.6563***		
	(0.1241)		
Panel B. Second stage			
#Reviews$_\tau$		0.0527	−0.1896*
		(0.0485)	(0.1098)
First-stage residual (control function)		−0.0556	0.1800
		(0.0516)	(0.1100)
Product attributes	Yes	Yes	Yes
Review states	Yes	Yes	Yes
#Orders$_\tau$	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Dep. var. mean	4.40	46.22	5.49
N	19,832	19,832	19,439
R^2	0.81855		
Pseudo R^2		0.58130	0.07349
First-stage F-statistic (excluded IV)	116.3		

Notes: Column (1) reports the OLS first-stage estimate using SolicitShare$_\tau$ as the excluded instrument. Columns (2) and (3) report PPML second-stage estimates with a control-function correction based on the first-stage residual. All columns exclude observations with zero orders in week τ because SolicitShare$_\tau$ is undefined when #Orders$_\tau = 0$. Observations with zero orders in the outcome week $\tau + 1$ are further excluded in column (3). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

relative to the sample mean of 5.49.

This estimated return-reducing effect is larger in magnitude than the corresponding estimate under the baseline instrument in Table 2 in the main text, where the implied relative reduction is about 8.1%. We interpret this difference cautiously. The two specifications use different sources of identifying variation and, in the case of the alternative instrument, are estimated on a different sample because solicitation share is undefined when focal-week orders equal zero. Thus, the estimates may correspond to different local average treatment effects: the baseline instrument identifies the effect of review generation shifted by solicitation volume conditional on focal-week orders, while the alternative instrument identifies the effect among product-weeks with substantial intensive-margin variation in solicitation share. For the purposes of this robustness exercise, the main takeaway is that both specifications support the same qualitative conclusion: additional reviews reduce subsequent return rates on average, with the effect identified from solicitation-driven variation in review generation.

C.1.2 State-Contingent Specification

In the heterogeneous demand model, the endogenous review variable enters as the set of state-specific interactions $\#\text{Reviews}_\tau \times S^{(k)}_{j,t-1}$. Accordingly, we estimate state-contingent first stages by interacting the excluded instrument with the same review-state indicators in each specification. This ensures that the first stage mirrors the heterogeneous structure of the second stage.

Table C.2: Demand model: first-stage estimates by existing review state

Dependent variable:	$\#\text{Reviews}_\tau$			
Excluded instrument:	$\#\text{Solicitations}_\tau$		SolicitShare_τ	
Column:	(1)	(2)	(3)	(4)
IV $\times I$(zero reviews)	0.0187	0.0185	0.6034	0.6028
	(0.0135)	(0.0134)	(0.5972)	(0.5973)
IV $\times I$(Q1 count)	0.0161**		0.5001**	
	(0.0067)		(0.2411)	
$\times I$(Low rating)		0.0048		0.3701
		(0.0099)		(0.3675)
$\times I$(High rating)		0.0224**		0.6081*
		(0.0098)		(0.3110)
IV $\times I$(Q2 count)	0.0160**		0.6995***	
	(0.0075)		(0.2337)	
$\times I$(Low rating)		0.0168		0.5601**
		(0.0155)		(0.2791)
$\times I$(High rating)		0.0154**		0.8990**
		(0.0075)		(0.4019)
IV $\times I$(Q3 count)	0.0417***		1.141***	
	(0.0115)		(0.3270)	
$\times I$(Low rating)		0.0298*		0.8208**
		(0.0180)		(0.3682)
$\times I$(High rating)		0.0533***		1.566***
		(0.0156)		(0.5912)
IV $\times I$(Q4 count)	0.0241**		0.2967	
	(0.0106)		(0.2380)	
$\times I$(Low rating)		0.0238		0.2957
		(0.0182)		(0.3850)
$\times I$(High rating)		0.0241		0.2984
		(0.0165)		(0.3098)
IV $\times I$(Q5 count)	0.0026		0.6951**	
	(0.0066)		(0.2737)	
$\times I$(Low rating)		−0.0063		0.3492
		(0.0085)		(0.2772)
$\times I$(High rating)		0.0104		0.9914**
		(0.0083)		(0.4655)
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
$\#\text{Orders}_\tau$	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	4.07	4.07	4.40	4.40
N	20,276	20,276	19,832	19,832
R^2	0.82529	0.82626	0.81875	0.81889
First-stage F-statistic (excluded IV)	126.5	79.7	23.1	14.0

Notes: Columns (1)–(2) use $\#\text{Solicitations}_\tau$ as the excluded instrument. Columns (3)–(4) use SolicitShare_τ as the excluded instrument and exclude observations with zero orders in week τ. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Baseline instrument: solicitation volume. Table C.2 reports the resulting state-contingent first-stage estimates. Columns (1)–(2) use the baseline instrument, $\#\text{Solicitations}_\tau$, and correspond to the first-stage equations underlying the main heterogeneous demand estimates. Column (1) interacts solicitation volume with review-count bins only, whereas Column (2) further splits each count bin by rating. Under the baseline instrument, the first stage remains strong, with excluded-instrument F-statistics of 126.5 and 79.7. In Column (1), the coefficients are positive and statistically significant in most low- and middle-review states: 100 additional solicitations predict about 1.61 additional reviews in Q1, 1.60 in Q2, 4.17 in Q3, and 2.41 in Q4. The coefficients in the zero-review state and Q5 are smaller than those in the middle-review states and are less precisely estimated. In Column (2), the first stage is especially strong in the high-rating cells of Q1–Q3, where 100 additional solicitations predict about 2.24, 1.54, and 5.33 additional reviews, respectively. Overall, these estimates indicate that the baseline instrument remains highly relevant for review generation by the focal-week order cohort after allowing the first stage to vary across review state.

Alternative instrument: solicitation share. Columns (3)–(4) report the analogous state-contingent first stages using the alternative instrument, $\text{SolicitShare}_{\tau}$. Once the instrument is interacted with review-state indicators, the first stage becomes weaker: the joint excluded-instrument F-statistics fall to 23.1 and 14.0.[28] Nevertheless, the coefficients remain generally positive and are statistically significant in several low- and middle-review states, indicating that the alternative instrument still captures meaningful variation in review generation.

Table C.3 reports the corresponding heterogeneous second-stage estimates with the solicitation share as the alternative excluded instrument. The qualitative pattern remains broadly consistent with the main heterogeneous results. On the orders margin, the estimated effect of an additional review is positive in the zero-review and low-to-middle review states, and then attenuates toward zero in the high-count states. On the returns margin, the coefficients remain negative across all review states and are generally largest in magnitude when prior review information is sparse. In the specification with review-count bins only, for example, one additional review raises next-week orders by about 8.6% in the zero-review state, 7.3% in Q1, and 9.4% in Q2, while reducing next-week return rates by about 17.0%, 15.3%, and 13.5% in those same states. The rating-split

[28] Because the heterogeneous specification contains multiple state-interacted endogenous regressors and excluded instruments, the reported first-stage F-statistics should be viewed as summary diagnostics of instrument relevance. Accordingly, we treat the heterogeneous specifications using solicitation share as a qualitative robustness check, rather than as a basis for precise magnitudes or inference.

Table C.3: Demand model using the alternative instrument (solicitation share): second-stage estimates by existing review state

Dependent variable:	$\#\text{Orders}_{\tau+1}$		Return $\text{rate}_{\tau+1}$	
Column:	(1)	(2)	(3)	(4)
Marginal effect of reviews				
$\#\text{Reviews}_\tau \times I(\text{zero reviews})$	0.0827*	0.1111**	−0.1867*	−0.2075**
	(0.0470)	(0.0461)	(0.1021)	(0.0943)
$\#\text{Reviews}_\tau \times I(\text{Q1 count})$	0.0705		−0.1666*	
	(0.0452)		(0.0995)	
$\times I(\text{Low rating})$		0.0987**		−0.1653*
		(0.0457)		(0.0914)
$\times I(\text{High rating})$		0.1002**		−0.2039**
		(0.0444)		(0.0931)
$\#\text{Reviews}_\tau \times I(\text{Q2 count})$	0.0901*		−0.1452	
	(0.0463)		(0.0962)	
$\times I(\text{Low rating})$		0.1225***		−0.1655*
		(0.0455)		(0.0897)
$\times I(\text{High rating})$		0.1172**		−0.1664*
		(0.0456)		(0.0894)
$\#\text{Reviews}_\tau \times I(\text{Q3 count})$	0.0745		−0.1215	
	(0.0456)		(0.0970)	
$\times I(\text{Low rating})$		0.1124**		−0.1334
		(0.0458)		(0.0913)
$\times I(\text{High rating})$		0.1007**		−0.1483*
		(0.0445)		(0.0895)
$\#\text{Reviews}_\tau \times I(\text{Q4 count})$	0.0304		−0.1456	
	(0.0485)		(0.0967)	
$\times I(\text{Low rating})$		0.0461		−0.1570*
		(0.0467)		(0.0893)
$\times I(\text{High rating})$		0.0794*		−0.1769**
		(0.0421)		(0.0887)
$\#\text{Reviews}_\tau \times I(\text{Q5 count})$	0.0173		−0.1175	
	(0.0476)		(0.0968)	
$\times I(\text{Low rating})$		0.0351		−0.1312
		(0.0456)		(0.0886)
$\times I(\text{High rating})$		0.0608		−0.1479
		(0.0434)		(0.0905)
First-stage residual (control function)				
Residual $\times I(\text{zero reviews})$	−0.0570	−0.0836*	0.2122*	0.2331**
	(0.0498)	(0.0491)	(0.1085)	(0.1013)
Residual $\times I(\text{Q1 count})$	−0.0722		0.1483	
	(0.0462)		(0.1014)	
$\times I(\text{Low rating})$		−0.0930*		0.1712*
		(0.0475)		(0.1000)
$\times I(\text{High rating})$		−0.1031**		0.1695*
		(0.0464)		(0.0957)
Residual $\times I(\text{Q2 count})$	−0.0833*		0.1474	
	(0.0467)		(0.0963)	
$\times I(\text{Low rating})$		−0.0968**		0.1609*
		(0.0469)		(0.0886)
$\times I(\text{High rating})$		−0.1208***		0.1785*
		(0.0458)		(0.0928)
Residual $\times I(\text{Q3 count})$	−0.0981**		0.0946	
	(0.0490)		(0.1020)	
$\times I(\text{Low rating})$		−0.1306***		0.1331
		(0.0471)		(0.0927)
$\times I(\text{High rating})$		−0.1303***		0.0924
		(0.0489)		(0.1025)
Residual $\times I(\text{Q4 count})$	−0.0472		0.1459	
	(0.0534)		(0.0969)	
$\times I(\text{Low rating})$		−0.0421		0.1503*
		(0.0483)		(0.0912)
$\times I(\text{High rating})$		−0.1274***		0.1803**
		(0.0436)		(0.0893)
Residual $\times I(\text{Q5 count})$	−0.0012		0.1084	
	(0.0499)		(0.0976)	
$\times I(\text{Low rating})$		−0.0237		0.1220
		(0.0469)		(0.0902)
$\times I(\text{High rating})$		−0.0525		0.1463
		(0.0460)		(0.0912)
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
$\#\text{Orders}_\tau$	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	46.22	46.22	5.49	5.49
N	19,832	19,832	19,439	19,439
Pseudo R^2	0.62016	0.63176	0.07710	0.07905
First-stage F-statistic (excluded IV)	23.1	14.0	23.1	14.0

Notes: All columns use SolicitShare_τ as the excluded instrument in the first stage. Columns (1)–(2) define review states using review-count bins only. Columns (2) and (4) further split each review-count bin by low versus high rating. Columns (1)–(4) exclude observations with zero orders in week τ. Columns (3)–(4) further exclude observations with zero orders in the outcome week $\tau + 1$. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

specification preserves the same broad pattern, with many low- and middle-review coefficients remaining statistically significant.

As an additional sensitivity check on the solicitation-share specification, we re-estimate the model after restricting the sample to product-weeks with $\text{SolicitShare}_{j\tau} \in [0.1, 0.9]$. This restriction removes observations with extreme shares near 0 or 1, which are likely to arise in low-volume product-weeks (e.g., a product-week with one order can take only two share values, 0 or 1) and can exert disproportionate leverage in estimation. The results remain qualitatively similar under this restricted sample.

Taken together, these results indicate that our main substantive conclusions do not depend on whether the excluded instrument is specified as solicitation volume or solicitation share.

Preferred instrument specification We treat solicitation volume as the preferred excluded instrument for three reasons. First, relative to solicitation share, the volume instrument retains more identifying variation and yields a substantially stronger first stage. Second, the institutional source of variation is the retailer's quota on the number of solicitation messages, which directly maps to variation in $\#\text{Solicitations}_{j\tau}$. The share formulation imposes an additional functional-form restriction that the relevant variation is the specific ratio. Third, the volume instrument avoids the sample restrictions required for the share specification, which is undefined for product-weeks with zero focal-week orders. We therefore use solicitation volume in the main specification and treat solicitation share as a robustness check.

C.2 Comparison With and Without Control-Function Correction

This subsection examines how accounting for endogeneity through the control-function correction affects the demand model estimates. Using the baseline solicitation-volume instrument, we compare otherwise identical specifications with and without the first-stage residual. The comparison shows whether correcting for endogenous review generation materially changes the estimated effects of additional reviews on downstream orders and return rates.

Intuitively, the first-stage residual, $\hat{v}_{j\tau}$, captures the component of review generation by the focal-week order cohort that remains unexplained by solicitation exposure, focal-week order volume, and the other included controls. Including $\hat{v}_{j\tau}$ in the second stage implements the control-function correction by conditioning on this residual component, thereby accounting for the endogeneity of

$\#\text{Reviews}_\tau$. Under the maintained identifying assumptions, the remaining variation in $\#\text{Reviews}_\tau$ is identified from solicitation-driven variation in review generation. Statistically significant coefficients on the residual terms are consistent with review generation being correlated with unobserved determinants of next-week outcomes and indicate that omitting the control-function residual would yield biased estimates of the effect of reviews.

Tables C.4 and C.5 compare estimates with and without the control-function correction. In each table, the "Without CF" columns report the same PPML specification *excluding* the first stage residual, while the "With CF" columns include the residual terms. The first stage is strong for the baseline instrument used for these comparisons, as the excluded-instrument F-statistic is 377.0 in the average specification and 126.5 in the state-interacted specification.

Table C.4: Demand model: average second-stage estimates with and without control-function correction

Dependent variable:	$\#\text{Orders}_{\tau+1}$		Return $\text{rate}_{\tau+1}$	
Excluded instrument:	$\#\text{Solicitations}_\tau$	–	$\#\text{Solicitations}_\tau$	–
Column:	With CF (1)	Without CF (2)	With CF (3)	Without CF (4)
$\#\text{Reviews}_\tau$	-0.0091	-0.0033	-0.0845^{***}	-0.0114^{*}
	(0.0227)	(0.0064)	(0.0300)	(0.0066)
First-stage residual (control function)				
Residual	0.0059		0.0744^{**}	
	(0.0262)		(0.0295)	
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
$\#\text{Orders}_\tau$	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	45.34	45.34	5.47	5.47
N	20,276	20,276	19,709	19,709
Pseudo R^2	0.57974	0.57973	0.07176	0.07149
First-stage F-statistic (excluded IV)	377.0		377.0	

Notes: Columns (1) and (3) include the first-stage residual from instrumenting weekly reviews; columns (2) and (4) omit this term. Observations with zero orders in week $\tau + 1$ are excluded in columns (3)–(4). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table C.4 shows that the endogeneity correction matters differently for orders and returns. For orders, the estimated effect of an additional solicitation-induced review remains close to zero with or without the control function (-0.0091 with CF versus -0.0033 without CF). Both estimates are statistically indistinguishable from zero, and the residual coefficient is itself small and insignificant. Thus, at the average level, correcting for endogeneity does not materially change the conclusion that an additional review has little detectable effect on next-week orders.

Table C.5: Demand model: heterogeneous second-stage estimates with and without control-function correction

Dependent variable:	$\#\text{Orders}_{\tau+1}$		Return rate$_{\tau+1}$	
Excluded instrument:	$\#\text{Solicitations}_\tau$	–	$\#\text{Solicitations}_\tau$	–
Column:	With CF (1)	Without CF (2)	With CF (3)	Without CF (4)
Marginal effect of reviews				
$\#\text{Reviews}_\tau \times I(\text{zero reviews})$	0.0438***	0.0336***	−0.1271***	−0.0301*
	(0.0159)	(0.0099)	(0.0387)	(0.0174)
$\#\text{Reviews}_\tau \times I(\text{Q1 count})$	0.0301**	0.0177*	−0.1132***	−0.0344***
	(0.0143)	(0.0096)	(0.0343)	(0.0129)
$\#\text{Reviews}_\tau \times I(\text{Q2 count})$	0.0478***	0.0371***	−0.0893***	−0.0145
	(0.0146)	(0.0101)	(0.0294)	(0.0120)
$\#\text{Reviews}_\tau \times I(\text{Q3 count})$	0.0274**	0.0095	−0.0656**	−0.0043
	(0.0129)	(0.0080)	(0.0256)	(0.0084)
$\#\text{Reviews}_\tau \times I(\text{Q4 count})$	−0.0119	−0.0151***	−0.0897***	−0.0138*
	(0.0088)	(0.0058)	(0.0285)	(0.0081)
$\#\text{Reviews}_\tau \times I(\text{Q5 count})$	−0.0270**	−0.0277***	−0.0658**	0.0073
	(0.0108)	(0.0085)	(0.0307)	(0.0096)
First-stage residual (control function)				
Residual $\times I(\text{zero reviews})$	−0.0222		0.1506***	
	(0.0249)		(0.0555)	
Residual $\times I(\text{Q1 count})$	−0.0357**		0.0929**	
	(0.0168)		(0.0411)	
Residual $\times I(\text{Q2 count})$	−0.0412**		0.0897***	
	(0.0191)		(0.0337)	
Residual $\times I(\text{Q3 count})$	−0.0527***		0.0429	
	(0.0181)		(0.0288)	
Residual $\times I(\text{Q4 count})$	−0.0073		0.0893***	
	(0.0129)		(0.0288)	
Residual $\times I(\text{Q5 count})$	0.0430***		0.0567*	
	(0.0130)		(0.0336)	
Product attributes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes
$\#\text{Orders}_\tau$	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Dep. var. mean	45.34	45.34	5.47	5.47
N	20,276	20,276	19,709	19,709
Pseudo R^2	0.61894	0.60686	0.07541	0.07364
First-stage F-statistic (excluded IV)	126.5		126.5	

Notes: Columns (1) and (3) include the first-stage residual from instrumenting weekly reviews; columns (2) and (4) omit this term. Observations with zero orders in week $\tau + 1$ are excluded in columns (3)–(4). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

For return rates, however, the correction is both economically and statistically important. Once the residual is included, the coefficient on reviews becomes much more negative (-0.0845 with CF versus -0.0114 without CF), and the residual coefficient is positive and statistically significant. This pattern indicates that unobserved shocks associated with greater review generation are also associated with higher subsequent returns. Because the review coefficient in the return equation is negative, omitting the residual therefore biases the estimated return effect toward zero. In other words, the naive specification substantially understates the return-reducing effect of additional reviews.

Table C.5 shows that the same conclusion carries over to the heterogeneous specification. For orders, including the control-function residual terms strengthens the positive effects of reviews in low-information states (e.g., from 0.0336 to 0.0438 in the zero-review state, and from 0.0177 to 0.0301 in Q1). The residual interactions are significant in several review states, indicating that endogenous review inflow matters for the state-contingent order estimates even though the average order effect remains null.

For returns, the correction is even more consequential. Without the control function, the estimated return-rate reductions are small and often statistically weak or absent outside the earliest states. With the correction, the coefficients become substantially more negative across review states. The residual interactions are positive and statistically significant in most states, again implying that unobserved shocks raising review generation are also associated with higher subsequent returns. The naive specification without the control-function correction therefore masks part of the information channel by biasing the return effect toward zero.

Taken together, these comparisons support the use of the control-function approach in the main demand model. Correcting for endogeneity is empirically consequential, especially for return rates, and strengthens the causal interpretation of the impact of solicitation-driven variation in review generation on downstream market outcomes.

C.3 Robustness of the Early-Information State Results

Because solicitations are least effective at eliciting reviews in the zero-review state, the corresponding heterogeneous demand estimates for that state should be interpreted with more caution than those for neighboring review states. Importantly, however, our substantive conclusion does not rely on a knife-edge distinction at exactly zero reviews. The economically relevant pattern is that buyer-side value is concentrated in the earliest information states, whereas reviewer response is weakest in that same early-information region.

We assess this interpretation in several ways. First, we combine the zero-review and Q1 states into a single early-information category, which avoids relying on a separately estimated zero-review coefficient. Second, we compare specifications with and without the control-function correction (see Subsection C.2). Third, we re-estimate the demand model using a linear specification instead of PPML to verify that the results are not driven by the nonlinear functional form. In all three cases, the same directional pattern emerges: additional reviews are most valuable when prior review information is sparse, while solicitations remain least effective at generating reviews in those same states. These checks therefore suggest that our main misalignment result is robust, even if the exact zero-review state coefficient is not the most precisely identified component of the heterogeneous demand model.

C.4 Demand Model Heterogeneity: New Buyers vs. Existing Buyers

This subsection extends the heterogeneous demand model by splitting outcomes by buyer type. We distinguish between existing buyers, who have prior purchase history with the retailer before the focal week τ, and new buyers, who do not. Table C.6 reports PPML control-function estimates separately for all buyers, existing buyers, and new buyers. Columns (1)–(3) report effects on next-week orders, and Columns (4)–(6) report effects on next-week return rates. Columns (1) and (4) replicate the review-count-only specifications from Table 3 in the main text for the full sample.

Because the endogenous regressor $\#\text{Reviews}_\tau$ and the excluded instrument $\#\text{Solicitations}_\tau$ are both defined at the product-week order-cohort level, the first stage is common across buyer types; the buyer-type split occurs only in the second-stage outcomes. The identical first-stage F-statistics across columns in Table C.6 reflect this.

On the orders margin, the coefficients are broadly similar across existing and new buyers. In the zero-review state, one additional review increases next-week orders by 0.0434 for existing buyers and 0.0445 for new buyers, corresponding to roughly 4.4% and 4.6% increases, respectively. The same similarity appears in the low- and middle-review states. At Q2, for example, the coefficients are 0.0464 for existing buyers and 0.0496 for new buyers, implying increases of about 4.8% and

Table C.6: Demand model: second-stage estimates by customer type

Dependent variable:	#Orders$_{\tau+1}$			Return rate$_{\tau+1}$		
Column:	All customers (1)	Existing customers (2)	New customers (3)	All customers (4)	Existing customers (5)	New customers (6)
Marginal effect of reviews						
#Reviews$_\tau$ × I(zero reviews)	0.0438***	0.0434***	0.0445***	−0.1271***	−0.0558	−0.1691***
	(0.0159)	(0.0162)	(0.0171)	(0.0387)	(0.0479)	(0.0450)
#Reviews$_\tau$ × I(Q1 count)	0.0301**	0.0328**	0.0273*	−0.1132***	−0.0213	−0.1693***
	(0.0143)	(0.0141)	(0.0153)	(0.0343)	(0.0410)	(0.0416)
#Reviews$_\tau$ × I(Q2 count)	0.0478***	0.0464***	0.0496***	−0.0893***	−0.0670*	−0.1076***
	(0.0146)	(0.0149)	(0.0151)	(0.0294)	(0.0405)	(0.0345)
#Reviews$_\tau$ × I(Q3 count)	0.0274**	0.0278**	0.0271*	−0.0656**	−0.0266	−0.0900***
	(0.0129)	(0.0126)	(0.0139)	(0.0256)	(0.0344)	(0.0311)
#Reviews$_\tau$ × I(Q4 count)	−0.0119	−0.0117	−0.0124	−0.0897***	−0.0463	−0.1115***
	(0.0088)	(0.0085)	(0.0100)	(0.0285)	(0.0362)	(0.0358)
#Reviews$_\tau$ × I(Q5 count)	−0.0270**	−0.0286***	−0.0253**	−0.0658**	−0.0068	−0.0976***
	(0.0108)	(0.0106)	(0.0118)	(0.0307)	(0.0386)	(0.0363)
First-stage residual (control function)						
Residual ×I(zero reviews)	−0.0222	−0.0173	−0.0261	0.1506***	0.1001	0.1744***
	(0.0249)	(0.0272)	(0.0247)	(0.0555)	(0.0665)	(0.0612)
Residual ×I(Q1 count)	−0.0357**	−0.0326*	−0.0387**	0.0929**	−0.0130	0.1696***
	(0.0168)	(0.0168)	(0.0179)	(0.0411)	(0.0501)	(0.0477)
Residual ×I(Q2 count)	−0.0412**	−0.0373*	−0.0466**	0.0897***	0.0338	0.1227***
	(0.0191)	(0.0202)	(0.0190)	(0.0337)	(0.0447)	(0.0432)
Residual ×I(Q3 count)	−0.0527***	−0.0506***	−0.0549***	0.0429	0.0141	0.0678**
	(0.0181)	(0.0179)	(0.0194)	(0.0288)	(0.0408)	(0.0328)
Residual ×I(Q4 count)	−0.0073	−0.0070	−0.0074	0.0893***	0.0702*	0.0959***
	(0.0129)	(0.0131)	(0.0131)	(0.0288)	(0.0414)	(0.0356)
Residual ×I(Q5 count)	0.0430***	0.0430***	0.0430***	0.0567*	−0.0100	0.0945**
	(0.0130)	(0.0122)	(0.0152)	(0.0336)	(0.0421)	(0.0384)
Product attributes	Yes	Yes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes	Yes	Yes
#Orders$_\tau$	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	45.34	24.70	21.86	5.473	5.078	5.771
N	20,276	20,276	20,276	19,709	19,171	19,134
Pseudo R^2	0.61894	0.57412	0.57431	0.07541	0.07491	0.06124
First-stage F-statistic (excluded IV)	126.5	126.5	126.5	126.5	126.5	126.5

Notes: Observations with zero orders in week $\tau + 1$ are excluded in columns (4)–(6). Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.1%, respectively. For both groups, the order effect attenuates as review volume accumulates and becomes slightly negative in the highest-count states. Thus, the order response to additional reviews is not meaningfully different across buyer types.

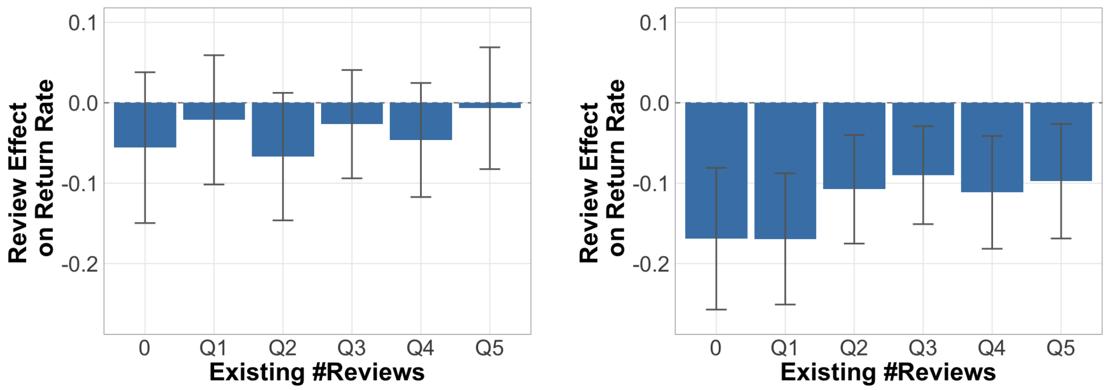
The returns margin shows a sharper difference. Additional reviews reduce return rates much more for new buyers than for existing buyers, especially when prior review information is scarce. In the zero-review state, the coefficient is -0.1691 for new buyers, corresponding to about a 15.6% reduction in return rates. The corresponding coefficient for existing buyers is smaller and statistically indistinguishable from zero. The same contrast persists in Q1: the coefficient is -0.1693 (15.6% reduction) for new buyers, whereas the estimate for existing buyers is again small and not distinguishable from zero. At Q2, both groups exhibit negative effects, but the magnitude remains larger for new buyers: -0.1076 for new buyers versus -0.0670 for existing buyers, corresponding to reductions of about 10.2% and 6.5%, respectively. More generally, the return coefficients for new buyers are negative and statistically significant across all review states, whereas the estimates for existing buyers are smaller and often less precisely estimated.

This pattern sharpens the informational interpretation developed in the main text. New buyers have less private information about the retailer's products and fit, so the first and early reviews are especially valuable for clarifying match quality and avoiding mismatches that later result in returns. Existing buyers, by contrast, can rely partly on their prior experience, making them less dependent on incremental review information when assessing whether a product is a good match. The stronger effects on returns among new buyers therefore suggest that reviews create value largely through improved information about product fit.

Figure C.1 visualizes the return-side coefficients by buyer type. Compared with Panel (a), which plots the estimates for existing buyers, Panel (b) shows consistently larger negative effects for new buyers, especially in the zero-review and Q1 states. The figure highlights that the return-reducing value of early reviews is concentrated among first-time customers.

From a managerial perspective, the misalignment between buyer-side value and reviewer response is especially costly for products that primarily attract new customers. The products for which early reviews are hardest to generate are also the ones whose first reviews most reduce avoidable returns among first-time buyers. Encouraging early-review generation for products with a high share of new customers, particularly while they remain in zero- or low-review states, should therefore yield

Figure C.1: Marginal effect of reviews on return rate by buyer type



(a) Existing buyers **(b)** New buyers

Note. Error bars show 95% confidence intervals.

disproportionate payoffs. Because standard solicitation underperforms exactly at those early states, complementary interventions that help products overcome the first-and-early review barrier may be necessary to improve onboarding for new customers.

D Reviewer Model: Sensitivity and Robustness

D.1 An Alternative Empirical Design: Local Design

Our identification of the heterogeneous solicitation effects, γ_k, in Equation (1) relies on within-month variation in solicitation exposure generated by the retailer's monthly SMS quota. As described in Section 2.4, solicitations are sent for all orders from the beginning of each month until the quota is exhausted, creating quasi-random variation in treatment assignment across purchase days within a month.

A potential concern with this design is that the solicitation-active period always precedes the solicitation-inactive period within a month. If solicitation-induced reviews generated early in the month alter the product's information environment—for example, by increasing review counts or shifting average ratings—then buyers in the later, non-solicitation period may face a different review state from otherwise comparable buyers in the earlier, solicitation period. Such within-month spillovers could affect both the composition of buyers who purchase and their review-writing behavior, potentially biasing the estimated solicitation effects.

To address this concern, we implement a local design that restricts the sample to narrow windows around the within-month treatment transitions, where the product's review information environment is unlikely to have shifted materially. The intuition follows a regression discontinuity approach: within a sufficiently narrow bandwidth around the transition, treated and untreated buyers are likely to face nearly identical informational conditions. The comparison therefore isolates a local, quasi-random change in solicitation probability, so any remaining difference in review-writing behavior can be attributed to solicitation rather than to a shifted review state.

We define two transition points for each month: the off-to-on cutoff, defined as the first day of the on-period, and the on-to-off cutoff, defined as the first day of the off-period. We then retain only observations whose purchase dates fall within a bandwidth of h days on either side of a cutoff, with $h \in \{3, 4, 5, 6\}$. Narrower bandwidths strengthen local comparability between treated and untreated observations but reduce sample size, whereas wider bandwidths provide greater statistical power at the cost of allowing more scope for the review environment to evolve within the estimation window. For each bandwidth, we re-estimate Equation (1) on the restricted sample. If spillovers in the review environment were materially biasing the baseline estimates, we would expect these local estimates to differ systematically from the full-sample results.

Table D.1: Reviewer model: average effect of solicitation across local bandwidths

Dependent variable:	Review incidence					Five-star indicator				
Bandwidth (days):	Full	3	4	5	6	Full	3	4	5	6
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Solicited	0.0114***	0.0101***	0.0080***	0.0093***	0.0108***	−0.0156**	−0.0204**	−0.0192**	−0.0166**	−0.0162**
	(0.0022)	(0.0031)	(0.0028)	(0.0027)	(0.0025)	(0.0062)	(0.0084)	(0.0077)	(0.0073)	(0.0069)
Product attributes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	0.09	0.09	0.09	0.09	0.09	0.90	0.90	0.90	0.90	0.90
N	142,079	64,717	83,746	100,491	114,682	13,031	6,032	7,785	9,265	10,637
R^2	0.00831	0.00933	0.00922	0.00873	0.00907	0.00713	0.01202	0.01101	0.00982	0.00887

Notes: Columns (1) and (6) report the baseline full-sample specification. The remaining columns restrict the sample to observations within the indicated bandwidth (in days) around the cutoff dates. Columns (6)–(10) are conditional on writing a review. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Tables D.1 and D.2 report the results from the local-design robustness checks. Table D.1 presents the average solicitation effects on review incidence and the probability of a five-star review across bandwidths. For review incidence, the estimated effect ranges from 0.0080 to 0.0108 across bandwidths of $h = 3$ to $h = 6$, compared with 0.0114 in the full sample, with all estimates significant at the 1% level. For the five-star indicator, the restricted-window estimates range from −0.0162

Table D.2: Reviewer model: heterogeneous effect of solicitation across local bandwidths

Dependent variable:	Review incidence					Five-star indicator				
Bandwidth (days):	Full	3	4	5	6	Full	3	4	5	6
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Solicited $\times I$(Zero reviews)	0.0040	0.0013	0.0045	0.0016	−0.0003	−0.0408	−0.0412	−0.0427	−0.0161	−0.0364
	(0.0092)	(0.0136)	(0.0118)	(0.0111)	(0.0099)	(0.0347)	(0.0393)	(0.0396)	(0.0340)	(0.0339)
Solicited $\times I$(Q1 count)	0.0083*	0.0074	0.0051	0.0073	0.0081	−0.0189	−0.0283	−0.0167	−0.0163	−0.0260*
	(0.0049)	(0.0069)	(0.0065)	(0.0056)	(0.0053)	(0.0138)	(0.0249)	(0.0195)	(0.0169)	(0.0150)
Solicited $\times I$(Q2 count)	0.0153***	0.0052	0.0038	0.0057	0.0106**	−0.0247*	−0.0304	−0.0217	−0.0228	−0.0177
	(0.0048)	(0.0067)	(0.0058)	(0.0053)	(0.0052)	(0.0146)	(0.0271)	(0.0219)	(0.0193)	(0.0168)
Solicited $\times I$(Q3 count)	0.0173***	0.0196***	0.0168***	0.0189***	0.0205***	−0.0132	−0.0096	−0.0055	−0.0075	−0.0085
	(0.0035)	(0.0065)	(0.0048)	(0.0047)	(0.0045)	(0.0160)	(0.0235)	(0.0229)	(0.0211)	(0.0188)
Solicited $\times I$(Q4 count)	0.0119***	0.0176**	0.0112*	0.0136**	0.0138***	−0.0396***	−0.0480***	−0.0542***	−0.0385***	−0.0363***
	(0.0040)	(0.0078)	(0.0058)	(0.0054)	(0.0045)	(0.0099)	(0.0158)	(0.0094)	(0.0110)	(0.0098)
Solicited $\times I$(Q5 count)	0.0076**	0.0032	0.0040	0.0039	0.0053	0.0107	0.0088	0.0037	−0.0016	0.0040
	(0.0036)	(0.0050)	(0.0047)	(0.0045)	(0.0041)	(0.0115)	(0.0150)	(0.0123)	(0.0128)	(0.0118)
Product attributes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review states	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	0.09	0.09	0.09	0.09	0.09	0.90	0.90	0.90	0.90	0.90
N	142,079	64,717	83,746	100,491	114,682	13,031	6,032	7,785	9,265	10,637
R^2	0.00834	0.00944	0.00928	0.00882	0.00915	0.00796	0.01296	0.01207	0.01026	0.00944

Notes: Columns (1) and (6) report the baseline full-sample specification. The remaining columns restrict the sample to observations within the indicated bandwidth around the treatment cutoff dates. Column (6)–(10) are conditional on writing a review. Standard errors clustered at the product level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

to −0.0204, closely tracking the full-sample estimate of −0.0156, with all estimates significant at the 5% level. The bandwidth-restricted estimates closely bracket the full-sample values, rather than exhibiting the systematic attenuation one would expect if within-month spillovers were materially biasing the baseline results. Thus, the local-design evidence provides little indication that the average solicitation effects are driven by spillovers from the solicitation-active period to the solicitation-inactive period.

Table D.2 reports the heterogeneous solicitation effects by review-count bin, which are the estimates underlying our reviewer–buyer misalignment finding. The main pattern from the full-sample specification is preserved across bandwidths. Most importantly, the first-and-early review barrier persists: in the zero-review state, the effect of solicitation on review incidence is close to zero and statistically insignificant in all restricted-window specifications, and the effect in Q1 remains small and statistically insignificant as well. By contrast, the effects in Q3 and Q4 remain positive, sizable, and statistically significant across all bandwidths. At bandwidth $h = 3$, for example, the estimated effects are 0.0196 ($p < 0.01$) for Q3 and 0.0176 ($p < 0.05$) for Q4, compared with 0.0173 and 0.0119, respectively, in the full sample. Some coefficients in Q1, Q2, and Q5 lose statistical significance in the narrower windows, which is unsurprising given the reduction in sample size—the tightest bandwidth ($h = 3$) retains only about 46% of the full sample—but their point estimates remain similar in sign and order of magnitude. A similar conclusion holds for the five-star indicator. The most pronounced negative effect continues to appear in Q4 across all bandwidths, and it remains precisely estimated throughout.

These results indicate that restricting the sample to narrow windows around the treatment transitions does not alter the substantive conclusions of the reviewer model. The first-and-early review barrier—and therefore the reviewer–buyer misalignment that follows from it—is robust to addressing potential spillover concerns through local comparisons around the treatment cutoffs.

D.2 Generalized Random Forests

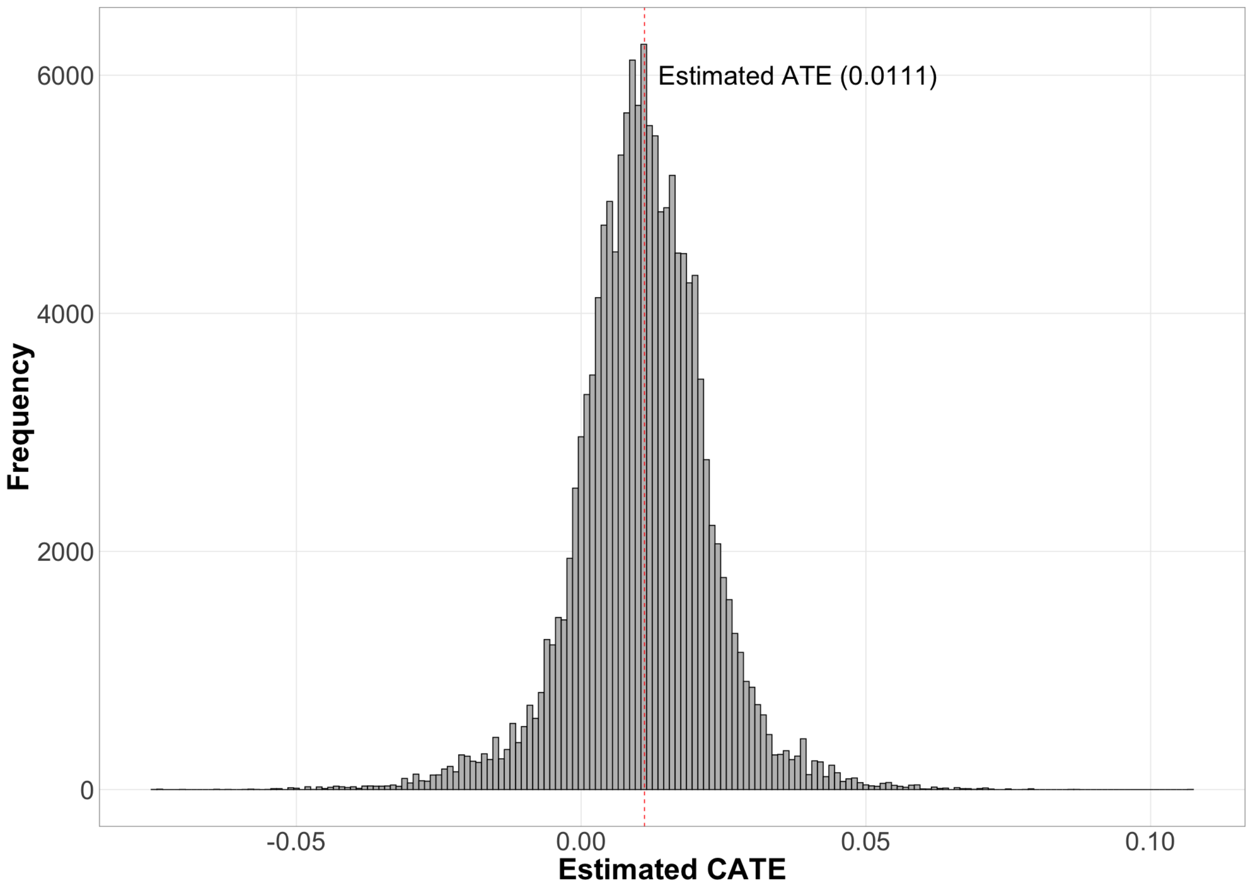
As a further robustness check, we re-estimate the reviewer-model heterogeneous effects using Generalized Random Forests (GRF) (Athey et al., 2019), a machine learning method for heterogeneous treatment effect estimation. This exercise uses the same quasi-experimental variation and identifying assumptions as the baseline reviewer model, while relaxing its functional-form restrictions. Relative to our baseline specification in Equation (1), GRF relaxes two restrictions. First, it allows covariates to affect outcomes through flexible, nonlinear functions rather than a linear functional form. Second, it allows treatment effects to vary smoothly with the underlying covariates, rather than being constant within the pre-specified review-state bins. We therefore use GRF to assess whether the first-and-early review barrier documented in Section 4.2 is robust to a substantially more flexible estimator.

Following the setup in Athey et al. (2019), we estimate the heterogeneous partially linear model

$$Y_{ijt} = \gamma(H_{jt})\,\mathsf{Solicited}_{ijt} + m(H_{jt}) + \varepsilon_{ijt}, \tag{5}$$
$$\text{with } \mathbb{E}\left[\varepsilon_{ijt} \mid \mathsf{Solicited}_{ijt}, H_{jt}\right] = 0,$$

where $\gamma(h) = \mathbb{E}\left[Y_{ijt}(1) - Y_{ijt}(0) \mid H_{jt} = h\right]$ is the conditional average treatment effect (CATE). The covariates H_{jt} include continuous measures of the product's pre-purchase review environment—existing review count and average rating observed on day $t-1$—along with month indicators and the same product characteristics X_{jt} used in Equation (1). We implement the estimator using the R package `grf` (Athey et al., 2019), which estimates heterogeneous treatment effects using orthogonalized nuisance adjustment. Intuitively, the causal forest learns $\gamma(\cdot)$ nonparametrically by recursively partitioning the covariate space to isolate regions with different treatment effects. For average effects, we report the corresponding doubly robust augmented inverse probability weighted (AIPW) estimate, which remains consistent if either the outcome model or the propensity score model is correctly specified (Robins et al., 1994).

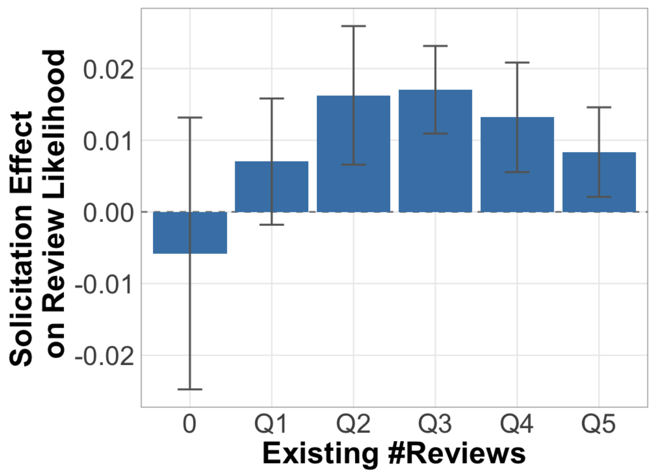
Figure D.1: Distribution of estimated conditional treatment effects from generalized random forests



Note. The figure plots the distribution of the estimated conditional average treatment effects (CATEs) of solicitation on review incidence. The implied average treatment effect is 0.0111, with a product-level clustered standard error of 0.0019.

Figure D.1 shows the distribution of the estimated CATEs from the GRF model. The implied average treatment effect is 0.0111 (with a product-level clustered standard error of 0.0019), very close to the baseline OLS estimate of 0.0114 in Table 4. The average effect is therefore not sensitive to replacing the linear specification with a more flexible estimator. At the same time, the estimated CATE distribution is unimodal and centered on a positive effect, but with substantial dispersion across observations. This heterogeneity suggests that the effect of solicitation depends meaningfully

Figure D.2: Estimated treatment effect on review incidence by existing review count



Note. The figure reports average estimated conditional treatment effects from the generalized random forest, grouped into the same existing-review-count bins used in Table 5. Error bars show 95% confidence intervals.

on the product's pre-purchase information environment.

To connect more directly to the main heterogeneity analysis, Figure D.2 presents the estimated CATEs averaged within the same review-count bins used in Table 5. The same qualitative pattern emerges: the treatment effect is near zero in the zero-review state, remains modest in the earliest review states, and becomes substantially larger once a product has accumulated an initial review base. This close correspondence indicates that the first-and-early review barrier documented in Section 4.2 is not an artifact of the linear functional form or the specific interactions imposed by the OLS specification. A flexible causal machine-learning estimator delivers the same substantive pattern, reinforcing the reviewer-side evidence underlying our reviewer–buyer misalignment result.

E Simulation Procedure

Our counterfactual in Section 5 quantifies how much additional net revenue the firm would expect from one solicitation message if solicitations were more effective at eliciting first reviews. This appendix section provides the details of the simulation procedure and the conversion to net revenue. The simulation proceeds in three steps: (i) predicting counterfactual review generation under an alternative zero-review-state solicitation effect, (ii) propagating these counterfactual reviews through the demand model to obtain counterfactual orders and return rates, and (iii) converting these outcomes into net revenue per solicitation message.

E.1 Step 1: Counterfactual Review Generation

For orders associated with products in the zero-review state, we replace the estimated solicitation effect in the reviewer model with a counterfactual value. This yields a counterfactual number of additional reviews generated by solicitations in that state.

Let $\mathbf{\Theta}_{\text{Reviewer}}$ and $\mathbf{\Theta}_{\text{Demand}}$ denote the estimated parameter vectors from the reviewer and demand models, respectively. Let $\widehat{\gamma}_0 \in \mathbf{\Theta}_{\text{Reviewer}}$ denote the estimated solicitation effect in the zero-review state.

Using $\mathbf{\Theta}_{\text{Reviewer}}$, we obtain the baseline predicted review-incidence outcome $\widehat{Y}_{ijt}^{\text{base}}$ for each order (i, j, t). For each product j and focal day t, the corresponding baseline predicted number of reviews generated by the focal-week order cohort $\mathcal{O}_{j\tau(t)}$ is

$$\widehat{N}_{j\tau(t)}^{\text{base}} := \sum_{(i,j,s)\in\mathcal{O}_{j\tau(t)}} \widehat{Y}_{ijs}^{\text{base}}.$$

Next, we construct a counterfactual reviewer-model parameter vector by replacing $\widehat{\gamma}_0$ with a counterfactual value γ_0^{cf}, while holding all other reviewer-model parameters fixed:

$$\mathbf{\Theta}_{\text{Reviewer}}^{\text{cf}} := \mathbf{\Theta}_{\text{Reviewer}} \text{ with } \widehat{\gamma}_0 \text{ replaced by } \gamma_0^{\text{cf}}.$$

Using $\mathbf{\Theta}_{\text{Reviewer}}^{\text{cf}}$, we similarly obtain the counterfactual predicted review-incidence outcome $\widehat{Y}_{ijt}^{\text{cf}}$. The implied counterfactual number of reviews generated by the focal-week order cohort is then

$$\widehat{N}_{j\tau(t)}^{\text{cf}} := \sum_{(i,j,s)\in\mathcal{O}_{j\tau(t)}} \widehat{Y}_{ijs}^{\text{cf}}.$$

Define the counterfactual change in review generation as

$$\Delta\widehat{N}_{j\tau(t)} := \widehat{N}_{j\tau(t)}^{\text{cf}} - \widehat{N}_{j\tau(t)}^{\text{base}}.$$

E.2 Step 2: Counterfactual Buyer Outcomes

We next feed the counterfactual shift in review generation into the estimated demand model to map the resulting change in review supply into subsequent orders and return rates.

For each product j and focal day t, define the counterfactual review input

$$N^{\mathrm{cf}}_{j\tau(t)} := N_{j\tau(t)} + \Delta\widehat{N}_{j\tau(t)},$$

where $N_{j\tau(t)}$ is the observed number of reviews generated by the focal-week order cohort. All other regressors in the demand model are held at their observed values. In particular, we leave the control-function residual $\widehat{v}_{j\tau(t)}$ unchanged, so that the counterfactual variation operates only through the exogenous, solicitation-driven change in review generation.

Using $\boldsymbol{\Theta}_{\mathrm{Demand}}$ together with the counterfactual review input $N^{\mathrm{cf}}_{j\tau(t)}$, we then obtain the counterfactual predicted demand-model outcomes

$$\left\{\widehat{Y}^{(O),\mathrm{cf}}_{j,\tau(t)+1}, \widehat{Y}^{(R),\mathrm{cf}}_{j,\tau(t)+1}\right\}.$$

The corresponding baseline outcomes $\widehat{Y}^{(O),\mathrm{base}}_{j,\tau(t)+1}$ and $\widehat{Y}^{(R),\mathrm{base}}_{j,\tau(t)+1}$ are computed similarly using $\boldsymbol{\Theta}_{\mathrm{Demand}}$ and the observed data.

E.3 Step 3: Net-Revenue Gain per Solicitation

We then convert the predicted changes in orders and return rates into net revenue and express the result on a per-solicitation basis.

For each product j and focal day t, define baseline and counterfactual net revenue over the outcome week $\tau(t) + 1$ as

$$\widehat{\pi}^{\mathrm{base}}_{j,\tau(t)+1} := p_j \cdot \widehat{Y}^{(O),\mathrm{base}}_{j,\tau(t)+1} \cdot \left(1 - \frac{\widehat{Y}^{(R),\mathrm{base}}_{j,\tau(t)+1}}{100}\right),$$
$$\widehat{\pi}^{\mathrm{cf}}_{j,\tau(t)+1} := p_j \cdot \widehat{Y}^{(O),\mathrm{cf}}_{j,\tau(t)+1} \cdot \left(1 - \frac{\widehat{Y}^{(R),\mathrm{cf}}_{j,\tau(t)+1}}{100}\right).$$

Here, p_j denotes the unit price of product j.

The implied firm-level change in net revenue is then

$$\Delta\Pi := \frac{1}{7}\sum_{j,t} I\left\{S^{(0)}_{j,t-1} = 1\right\}\left(\widehat{\pi}^{\mathrm{cf}}_{j,\tau(t)+1} - \widehat{\pi}^{\mathrm{base}}_{j,\tau(t)+1}\right).$$

The notation $I\{S^{(0)}_{j,t-1} = 1\}$ denotes the zero-review state indicator and restricts the aggregation to product–day observations in the zero-review state, which is the margin affected by the counterfactual. The factor 1/7 adjusts for the rolling-window construction of the demand model: because $\tau(t) + 1$ is a seven-day outcome window and the focal day t advances one day at a time, each calendar day of revenue is counted in seven adjacent outcome-week observations.

To express the gain on a per-solicitation basis, we divide $\Delta\Pi$ by the total number of solicitation messages sent to zero-review-state focal-week cohorts:

$$\Delta\Pi_{\text{per-solicitation}} := \frac{\Delta\Pi}{\frac{1}{7}\sum_{j,t} I\left\{S^{(0)}_{j,t-1} = 1\right\} Z_{j,\tau(t)}}. \tag{6}$$

Here, $Z_{j,\tau(t)}$ denotes the number of solicitation messages sent to focal-week order cohorts. The same 1/7 adjustment appears in the denominator for the same reason: each focal-week solicitation cohort is counted in seven overlapping rolling windows.