The Effect of Reviewer Heterogeneity Between One-Time and Multi-Review Reviewers, Price Sentiment, and Pre-Purchase Information on Star Ratings for Amazon Subscription Boxes

Jikhan Jeong[1]

Preliminary draft. Please do not circulate.

Version: March 4, 2026

Abstract

This study examines the presence, prevalence, and negative polarity of one-time reviewers compared to those who have left multiple reviews across all categories, based on an analysis of 571,544,746 Amazon product reviews. Since sellers and consumers may be interested in specific products relevant to their business or purchasing interests within a category rather than across all categories, this research focuses on the Subscription Boxes category for econometric analysis. This category exhibits an unusual rating distribution, likely presents fewer incentives for promotional reviews, and faces fewer constraints. However, price data is unavailable in this category. Therefore, this study develops price sentiment variables from review text data using generative LLMs and manual annotation. Specifically, the fine-tuned GPT-4.1 model outperforms other GPT models. Moreover, pre-purchase information from past reviews is extracted through data mining. Because star ratings are ordinal responses that indicate the strength of consumer preference, this study employs an ordered probit model and marginal effects analysis to interpret the results. The findings reveal variability in the marginal effects of reviewer type, purchase verification, and price sentiment on the likelihood of giving 1- or 5-star ratings. Additionally, consumers consider pre-purchase information from previous reviews when assigning star ratings. Thus, this approach can help market players make data-driven decisions when relevant consumer and business data are limited, and it can help bridge the information gap with Amazon, which operates both as a marketplace and as a seller on its own platform.

Keywords

Online Product Reviews, Abnormal Reviewers, Reviewer Heterogeneity, Price Sentiment,

Pre-Purchase Information, Consumer Preferences

[1] I am an assistant professor of Economics at Missouri University of Science and Technology (MST), Rolla, MO 65409, USA E-Mail: jikhan.jeong@mst.edu

1. Introduction

Rapid advances in big data and artificial intelligence (AI) within digital economies have considerably increased the significance and influence of digital platforms on consumers, businesses, and economic systems during and after the pandemic. The volume, speed, and variety of user-generated content (UGC) on these platforms have increased substantially. Consumers produce UGC—such as online product reviews—on digital retail platforms to share real product experiences, and this content remains publicly accessible at no cost. Pre-purchase information in online reviews can improve consumer welfare (Reimers and Waldfogel, 2021). Additionally, Remiers et al. (2025) showed that more pre-purchase information can increase consumer surplus (CS) by analyzing post-purchase video game usage data (e.g., playtime). Therefore, extracting valuable insights from large, rapidly growing, and diverse UGC is vital for understanding consumer behavior, firm operations, digital platforms, and the broader digital economy. For example, Timoshenko and Hauser (2019) identified consumer needs from product reviews, while Liu et al. (2019) examined how review content affects product sales. Numerous studies have explored how UGC influences sales on digital platforms (Anderson and Magruder, 2012; Chen, 2018; Chevalier and Mayzlin, 2006). However, some digital retail platforms with one-sided review systems (e.g., Amazon.com) allow consumers to post reviews freely while preventing sellers from publicly reviewing consumers in return (Cui et al., 2012; Tadelis, 2016). Moreover, reviews on such platforms typically lack demographic information about reviewers. As a result, researchers often analyze UGC without demographic data unless consumers voluntarily provide it or the platform shares it. Since many online consumers depend on pre-purchase information in UGC, biases (e.g., self-selection biases) and misinformation (e.g., fake reviews) can reduce consumer surplus, undermine product credibility, erode trust in platforms, and distort fair market competition (Askalidis et al., 2017; De Langhe et al., 2016; He et al., 2022; Hu et al., 2009; Hu et al., 2017; Li and Hitt, 2008; Luca

and Zervas, 2016; Mayzlin et al., 2014). Tackling these issues is essential for improving digital shopping experiences, optimizing business operations, and maintaining platform integrity.

Although there has been extensive research on fake reviews and reviewers, little attention has been given to classifying one-time reviewers, analyzing their rating patterns, or comparing their rating behavior with that of multi-review reviewers. For example, Mayzlin et al. (2014) suggested that one-time reviewers on TripAdvisor might be suspicious. Similarly, Luca and Zervas (2016) found that Yelp users with fewer reviews are more likely to have their reviews flagged as fake by Yelp's filtering system. Reviews from first-time reviewers generally have the highest filtering rate. Fang et al. (2024) demonstrated that the percentage of one-time reviewers on Yelp.com is linked to more extreme ratings, highlighting differences in how reviewers rate.

However, no research has yet defined reviewer segmentation, distinguished between one-time and multi-review reviewers, examined the prevalence of one-time reviewers across all product categories on Amazon.com, analyzed the skewness of their rating distributions, or estimated the impact of reviewer heterogeneity on star ratings, especially using large-scale, real-world datasets with sufficient temporal depth.

To address this gap, this study aims to: (1) develop a framework for segmenting consumers into subgroups to compare one-time reviewers (within a category or across all categories) with multi-review reviewers; (2) identify one-time reviewers and analyze their review counts and rating patterns in each category over multiple years in comparison to multi-review reviewers; and (3) estimate the impact of reviewer heterogeneity, price sentiment, and pre-purchase information from previous reviews on star ratings. Specifically, this study employs a descriptive, associational empirical approach—involving big data analysis, AI-generated variables, and an ordered probit model—rather than a causal inference framework, due to limited access to internal consumer data.

The present study first defines consumer segmentation into subgroups (Figure 1). Amazon consumers (AC) are divided into two groups: non-reviewers (NR), who do not write any reviews, and reviewers (R), who write one or more reviews. Furthermore, reviewers (R) are split into two categories: non-one-time reviewers within a category (NOT), who write more than one review within a specific category, and one-time reviewers within a product category (OT), who submit their first and only review for a particular product category. In detail, OT is divided into two subgroups: one-time reviewers across all categories (OTA) and one-time reviewers within a category but not across all categories (OTNOTA). This study identifies and compares the number of reviews and rating distributions for NOT, OTA, and OTNOTA across categories over time to ensure a fair comparison.

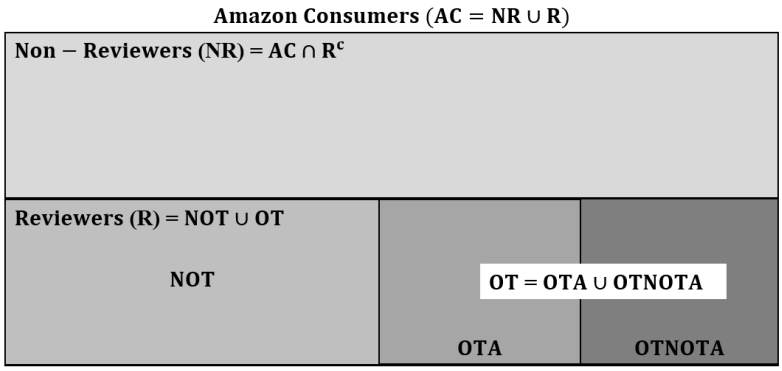


Figure 1. Framework for consumer segmentation into subgroups

As shown in Table 1, the analysis identifies 142,513,957 reviews (24.93% of all reviews) written by one-time reviewers within a product category (OT). Among these, 10,309,129 reviews (1.8% of all reviews) are written by one-time reviewers across all categories (OTA), and 132,204,828 reviews (23.13% of all reviews) are written by one-time reviewers within a category but not across categories (OTNOTA). Additionally, the number of reviews written by non-one-time reviewers within a category (NOT), who write more than one review, is 429,030,789 (75.07% of total reviews). In fact, every product category includes reviews from both OTA and OTNOTA.

Previous studies have shown skewed rating distributions with extreme positive and negative ratings on digital platforms (Brandes et al., 2022; Hu et al., 2006; Schoenmueller et al., 2020). Luca and Zervas (2016) found that filtered reviews, which are more suspicious than others on Yelp, tend to receive extreme ratings. Notably, Mayzlin et al. (2014) observed that one-time reviewers on TripAdvisor give more polarized ratings than the overall reviewing population. Additionally, Fang et al. (2024) reported that one-time reviewers are more likely to give 1- and 5-star ratings, especially more 1-star ratings, compared to reviewers who leave multiple reviews on Yelp.com. These prior studies suggest that the rating distributions of one-time reviewers on TripAdvisor and Yelp are more skewed toward positive and negative ratings.

In this study, the review rating distribution for OT (i.e., the union of OTA and OTNOTA) shows a stronger skew toward 1-star ratings compared to reviews from multi-review reviewers (i.e., NOT) on Amazon.com. However, the rating distribution of OT reviews shows a less pronounced skew toward 5-star ratings than NOT reviews. Overall, the higher proportion of 1-star ratings among one-time reviewers on Amazon.com in this study resembles the patterns seen on Yelp.com. Conversely, the lower share of 5-star ratings among one-time reviewers on Amazon.com in this study contrasts with the higher share on Yelp.com.

Table 1. Rating distribution of all reviews across all product categories (Appendix)

Reviewer Types	Reviews (% in all)	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 1&5
All (=1. OT+ 2. NOT)	571,544,746 (100%)	58,930,529 (10.31%)	28,106,622 (4.92%)	40,329,433 (7.06%)	72,502,401 (12.69%)	371,675,761 (65.03%)	430,606,290 (75.34%)
1. **OT** (=OTA+OTNOTA)	142,513,957 (24.93%)	21,242,136 (14.91%)	7,811,981 (5.48%)	9,052,369 (6.35%)	15,278,127 (10.72%)	89,129,344 (62.54%)	110,371,480 (77.45%)
1.1. OTA	10,309,129 (1.80%)	2,097,655 (20.35%)	572,779 (5.56%)	569,293 (5.52%)	1,079,061 (10.47%)	5,990,341 (58.11%)	8,087,996 (78.45%)
1.2. OTNOTA	132,204,828 (23.13%)	19,144,481 (14.48%)	7,239,202 (5.48%)	8,483,076 (6.42%)	14,199,066 (10.74%)	83,139,003 (62.89%)	102,283,484 (77.37%)
2. **NOT**	429,030,789 (75.07%)	37,688,393 (8.78%)	20,294,641 (4.73%)	31,277,064 (7.29%)	57,224,274 (13.34%)	282,546,417 (65.86%)	320,234,810 (74.64%)

He et al. (2022) argued that sellers have fewer incentives to purchase negative promotional reviews because of higher costs and lower expected returns. Therefore, OT may be less likely to write negative promotional reviews of competitors' products than to write

positive reviews of their own products. Additionally, Fang et al. (2024) showed that the rise in 1-star ratings on Yelp.com is driven by actual consumer dissatisfaction. Consequently, one plausible explanation for the negative rating behavior of one-time reviewers is a negative self-selection effect, in which consumers with negative experiences that exceed their expectations or tolerance levels leave their first—and only—review with a 1-star rating within a category or across all categories.

To understand how reviewer heterogeneity between one-time and multi-review reviewers affects their star ratings, this study first chooses specific product groups for analysis. Sellers and consumers may be more interested in business- or purchase-related products within a particular category than in products across all categories. Additionally, analyzing review data across all categories demands more computational resources, higher budgets, and more time. Therefore, this study focuses on the Subscription Boxes category, which shows the greatest difference in extreme rating shares between OTA and multi-review reviewers. The analysis uses only the first review from each reviewer to ensure a fair comparison between one-time and multi-review reviewers.

Consumers subscribe to a curated collection of products each subscription cycle with autopay. Consumer uncertainty about the quality of products in a subscription box is relatively high compared to other categories. Therefore, the value of pre-purchase information in previous reviews written by other reviewers for the subscription box is significant. Importantly, promotional reviewers may have less incentive to write positive reviews of their own products and negative reviews of competitors' products, because the subscription box changes the product selection each cycle and requires a subscription rather than a one-time payment. Overall, selecting the Subscription Boxes category as the target mitigates concerns about promotional reviews in the sample data.

In Economics, price is a crucial variable for demand estimation. For example, Chevalier and Goolsbee (2003) evaluated the impact of price on demand for books on digital platforms. However, numerical price data for products in the Subscription Boxes category is not available. To address this issue, this study creates discrete price-sentiment variables from review texts using generative large language models (LLMs) in combination with manual annotation. Specifically, it uses OpenAI's Generative Pre-trained Transformer (GPT) models. The GPT-4.1 model, fine-tuned on 1,000 training samples, achieves the best predictive performance for price sentiment classification on a manually labeled test set, outperforming the GPT-4o and GPT-5 models with and without fine-tuning. Therefore, the fine-tuned GPT-4.1 is employed to generate the price sentiment variables, which serve as independent variables in the econometric analysis.

Star ratings indicate how strongly consumers prefer the products being reviewed. Therefore, this study uses a random utility model for ordinal responses, specifically ordered probit models, to analyze the impact of reviewer differences, price sentiment, and pre-purchase information on star ratings. Marginal effect analysis is also performed for clearer interpretation.

The empirical results show heterogeneity in the marginal effects of reviewer types, purchase verification, and price sentiments. Additionally, consumers consider pre-purchase information when rating a product. Furthermore, unsatisfied reviewers tend to write longer review titles and bodies, especially longer titles. As a result, sellers, consumers, researchers, and regulators should account for reviewer heterogeneity and its varied effects on consumer preferences.

To the author's knowledge, this is the first comprehensive big data analysis comparing individual one-time reviewers with multi-review reviewers across all categories on Amazon.com, using cross-categorical analysis over multiple years. The findings enhance understanding of abnormal reviewer behavior and its impact on rating distributions.

A further contribution of this paper is the development of a consumer segmentation framework that differentiates between one-time and multi-review reviewers. Using this framework, heterogeneity among one-time reviewers within and across categories, and among multi-review reviewers, is identified through big data and empirical analyses. Specifically, the rating polarity of one-time reviewers is compared with that of multi-review reviewers to ensure a fair comparison. Additionally, the empirical results show heterogeneity in the marginal effects of reviewer types, purchase verification, and price sentiments on star ratings. These findings and the approach can be used to segment consumers and create personalized business strategies for diverse consumer groups.

This paper also advances the literature on pre-purchase information by showing how to extract hidden pre-purchase data from past reviews using data mining techniques and how to turn this data into variables. Additionally, empirical results show that consumers react to previous reviewers' (i.e., the crowd's) average star ratings, as well as to the proportions of 5-star and 2-star ratings and to the positive imbalance (i.e., the share of 5- and 4-star ratings) when assigning their own star ratings. These methods will be useful when consumers are uncertain about product quality and face relatively high search costs.

Another contribution of this study is the use of price sentiment analysis with generative LLMs and manual annotation. This paper details a manual annotation approach to building a validation dataset for evaluating generative LLMs. Additionally, it demonstrates the predictive performance of generative LLMs both with and without fine-tuning. The paper also shows how to combine manually annotated price sentiments with LLM-predicted price sentiments to create variables for econometric analysis. The empirical results offer a solid economic interpretation of these price sentiment variables, with their signs and magnitudes remaining consistent across different models. Hence, this study introduces a method that employs generative LLMs along with manual annotation to classify sentiment toward product prices in user-generated content

(UGC), rather than overall sentiment. The approach presented here will be useful when key numerical information, such as income and price, is missing, but unstructured data about it is available.

This study also helps reduce the information gap between platforms and other market participants and may address Amazon's self-preference. Amazon serves as a marketplace for both sellers and consumers, as well as a seller of its own products. Therefore, self-preference raises concerns for other market players, including regulators. Waldfogel (2024) studied the impact of the European Union's Digital Market Act (DMA) on Amazon's self-preferencing by analyzing the search rankings of Amazon's own products on its retail platform. Specifically, user information, such as consumer diversity, and data from other sellers are available to Amazon to enhance its business decisions, while this information remains inaccessible to other players. The approach outlined in this paper can help narrow the information gap between Amazon and other market participants, including consumers, sellers, and regulators, by making online product reviews publicly accessible and reflecting consumer opinions.

Overall, this paper's findings and proposed approach offer a novel way to understand reviewer heterogeneity and the underlying preferences that drive it. By leveraging latent information from historical reviews and unstructured data, the methodology presented is not only directly applicable but also readily extendable and scalable to other domains and categories.

The remainder of this paper is structured as follows. Section 2 outlines the research process and data methodology. Section 3 presents the results of the big-data analysis, focusing on identifying one-time and multi-review reviewers. In Section 4, we detail the empirical analysis and estimate the influence of reviewer heterogeneity, price sentiment, and pre-purchase information on star ratings. Finally, Section 5 concludes the paper by discussing the findings and their implications.

2. Research Process

This section provides a detailed overview of the research process. Figure 2 describes the big-data analysis process used to identify reviewer heterogeneity within each category across the entire sample period. Then, the empirical analysis for the target category, Subscription Boxes, which estimates the effects of reviewer heterogeneity, price sentiment, and pre-purchase information on star ratings, is presented.

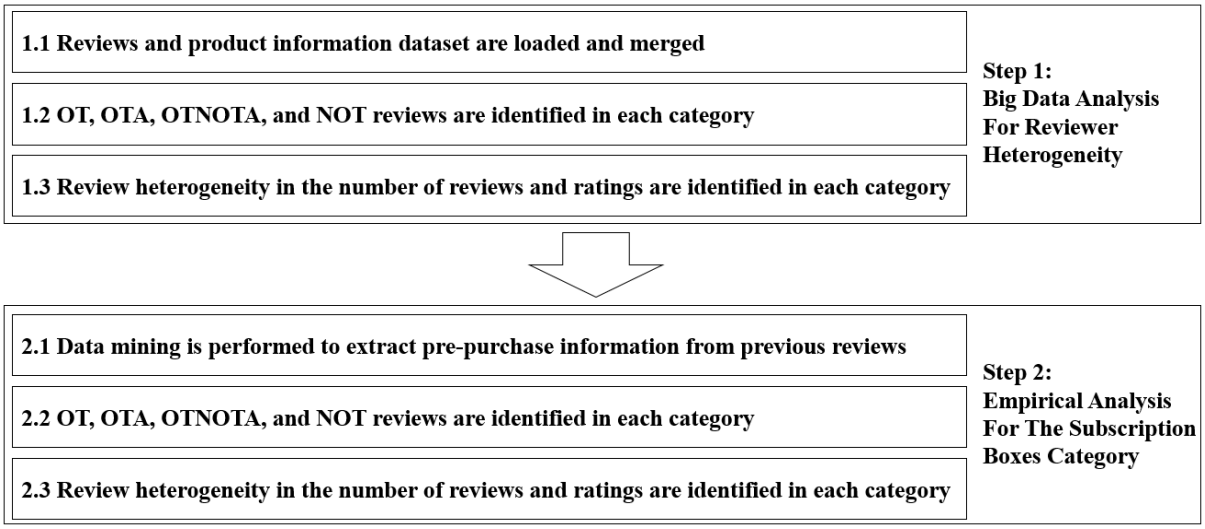


Figure 2. Pipeline diagram of the research process

2.1 Processing for identifying reviewer heterogeneity using big data analytics

To accurately classify individual reviewers, it is important to analyze each reviewer's complete review history, both within and across product categories over multiple years. This study examines the Amazon review dataset as secondary data (Hou et al., 2024), which includes reviewer and product identifiers, along with timestamps, enabling tracking of a consumer's digital footprint over several years. However, Amazon does not provide demographic information about reviewers, and the secondary data used in this study employs an encrypted reviewer ID; therefore, no personally identifiable information is included. This study does not explicitly mention any specific reviewer ID, product ID, product name, or store ID in the manuscript.

Specifically, the Google Cloud Platform (GCP) is primarily used to analyze consumer behavior and reviewer patterns in UGC across diverse product categories. It is configured with 64 cores, 128 virtual CPUs (vCPUs), and 864 GB of memory. Most of the code and its execution occur in the JupyterLab interface within the GCP Vertex AI workbench environment, with or without background execution. After the $5,000 in GCP credits are exhausted, high-performance computing (HPC) is used for further analysis, including the application of LLMs via the OpenAI API, with a maximum memory of 512 GB. Custom Python functions are developed and used for each pre-processing step.

This study uses specific data preprocessing steps to identify one-time reviewers within a product category (OT), who submit their first and only review in that category, and non-one-time reviewers within a category (NOT) through cross-categorical digital footprint analysis. Specifically, OT is divided into one-time reviewers across all categories (OTA) and one-time reviewers within a single category but not across all categories (OTNOTA).

First, product review and product information datasets are loaded and stored in GCP. Product reviews are then merged with these datasets using Amazon Standard Identification Numbers (ASINs). ASIN is a product ID for the reviewed item in this study. Due to the dataset size, each category-level dataset for reviews and product information is uploaded in the background separately. After uploading, each dataset is filtered to keep only key columns, preventing out-of-memory (OOM) issues. Then, the reviews and product information datasets are merged to match each review with its corresponding product details. Review timestamps are converted from Unix time to a human-readable format. Reviews with 0-star ratings are removed from the raw secondary data, as valid star ratings range from 1 to 5. Overall, a dataset containing 571,544,746 reviews for 48,185,048 unique products from 54,514,263 unique reviewers, spanning 17 June 1996 to 14 September 2023, is prepared.

Second, OT, OTA, OTNOTA, and NOT reviews are identified by counting each

reviewer's total reviews over the entire sample period. First, OT reviews are determined by counting the number of reviews per reviewer in each category from 1996 to 2023. If a reviewer has only one review in a category during the sample period, they are classified as OT. The total number of OT reviews and the number of reviewers are the same across categories. Second, an OTA is identified if a reviewer writes only one review throughout all categories during the entire period, and the number of OTAs matches the number of reviews they write within and across categories. Third, OTNOTA reviews are identified by subtracting OTA reviews from OT reviews. Fourth, NOT reviews are identified by subtracting OT reviews from all reviews. Some reviews are categorized as 'Unknown' because they lack category information; however, these reviews do include reviewer ID, product ID, and review timestamp. Therefore, reviews in the 'Unknown' category are included in the big data analysis.

Third, the number of reviews and rating distributions for each reviewer type in each category, as well as across all categories, are analyzed to identify the presence, prevalence, rating polarity, and temporal trends of one-time reviewers compared to multi-review reviewers.

During each pre-processing step for identifying reviewer heterogeneity, ratings and counts of reviews, reviewers, and products in each category for each year are carefully double-checked. The updated dataset from JupyterLab in GCP was transferred to a Google Cloud Storage (GCS) bucket and then downloaded to a local computer after undergoing big-data analysis, as large files cannot be downloaded directly from JupyterLab in GCP. These files are uploaded to HPC.

2.2 Processing for empirical analysis for the target product category

The Subscription Boxes category is selected as a target because one-time reviewers show abnormal rating patterns compared to multi-review reviewers. Also, there may be fewer incentives for promotional consumers to write promotional reviews, so concerns about promotional reviews in this category might be lower than in others. Additionally, the usefulness

of pre-purchase information from previous reviews may be higher, given the high consumer uncertainty about product quality and the changing nature of products from one subscription period to the next. Most importantly, this study's econometric analysis focuses on the Subscription Boxes category and involves three main steps. First, digital footprint mining is conducted to extract pre-purchase information for each reviewed product from previous other reviewers within the Subscription Boxes category. Dai et al. (2018) observed that existing reviews are likely to influence subsequent reviewers. This study, therefore, assumes that a consumer may be influenced by earlier reviews from previous reviewers when submitting their first review of a product. If a reviewer '*i* ' posts a review for the product '*j*' at time '*t*' in the category, all previous other reviewers' reviews by '*t-1*' since 1996 are collected and used to construct the variable for pre-purchase information. The timestamp for each review at '*t*' includes year, month, day, hour, minute, and second. Therefore, '*t-1*' refers to the most recent prior timestamp. As can be seen in Figure 3, all initial reviews for the product '*j*' between the time of the first review,$t^{a}_{-i,j}$, for the product '*j*' written by another reviewer (i.e., '-*i*') and the time of the most recent review,$t^{b}_{-i,j}$, for the product '*j*' written by another reviewer (i.e., '-*i*') are considered.

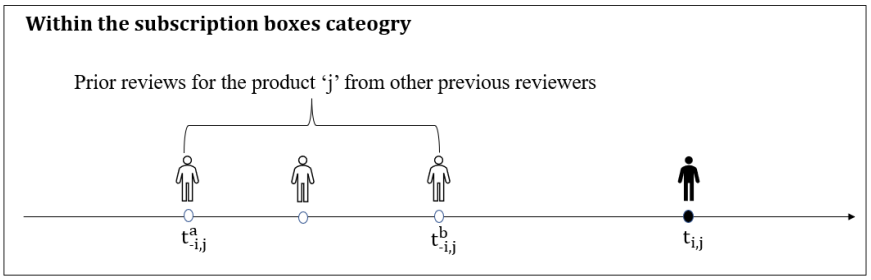


Figure 3. Digital Footprint mining for the product '*j*' in the Subscription Boxes category

Second, sentiment analysis of the price of a reviewed product is performed using review text data, because numerical price information is unavailable for products in this category. The review text captures actual consumer experiences and opinions. LLMs are well-suited for sentiment analysis of user-generated textual data. They have been widely used to

extract semantic information from unstructured data such as text, images, and videos. In particular, generative LLMs like OpenAI's GPT have demonstrated strong performance across a range of tasks and domains. Therefore, this study uses GPT models to extract price sentiment from review text.

To evaluate the predictive performance of GPT models, the author manually annotates the price sentiment for each review in the validation dataset (about 10% of the data). Based on this evaluation, the fine-tuned GPT-4.1, trained on around 1,000 manually labeled samples, is selected for price sentiment analysis. Fine-tuning involves training a pre-trained model with a small, task-specific dataset. Finally, the price sentiment variable is created by combining approximately 90% of the price sentiment produced by GPT-4.1 and about 10% of human-labeled price sentiment.

Third, a random-utility model for ordinal responses is used to estimate the effects of reviewer heterogeneity, price sentiment, and pre-purchase information on consumer preferences, such as star ratings. Since Amazon's star rating is an ordinal response, the ordered probit model is applied for empirical analysis. Then, marginal effect analysis is conducted for interpretation, as coefficients for independent variables in nonlinear models, such as ordered probit, cannot be interpreted directly.

The following subsection describes the results of big-data analysis of reviewer heterogeneity between one-time and multi-review reviewers across categories over the entire sample period from 1996 to 2023.

3. Big Data Analysis Results

3.1 Reviewer heterogeneity over time

This study examines one-time reviews over 28 years. As shown in Table 2, although there are yearly fluctuations, the overall trend in the share of OTA and OTNOTA reviews

among all reviews has increased. Compared to their share in 2019 (before the Covid-19 pandemic), the share of OTA and OTNOTA in 2020 (during the pandemic) rose by 0.59% and 3.4%, respectively. As the number of one-time reviews has grown, their impact on consumers, sellers, recommendation systems, and Amazon.com may have expanded as well.

Table 2. The number of NOT, OTA, and OTNOTA reviews in all categories over time

Year	All	NOT		OT			
				OTA		OTNOTA	
		# of reviews	Share in All	# of reviews	Share in All	# of reviews	Share in All
1996	26	24	92.31%	0	0.00%	2	7.69%
1997	5,616	4,323	76.98%	897	15.97%	396	7.05%
1998	35,274	26,716	75.74%	4,873	13.81%	3,685	10.45%
1999	92,764	69,803	75.25%	11,772	12.69%	11,189	12.06%
2000	323,514	267,109	82.56%	25,023	7.73%	31,382	9.70%
2001	333,820	278,943	83.56%	20,439	6.12%	34,438	10.32%
2002	357,870	298,465	83.40%	20,480	5.72%	38,925	10.88%
2003	385,518	319,637	82.91%	21,871	5.67%	44,010	11.42%
2004	472,360	392,368	83.07%	23,148	4.90%	56,844	12.03%
2005	662,651	540,763	81.61%	30,122	4.55%	91,766	13.85%
2006	740,442	591,630	79.90%	31,939	4.31%	116,873	15.78%
2007	1,164,701	901,148	77.37%	36,314	3.12%	227,239	19.51%
2008	1,340,481	1,034,781	77.19%	37,798	2.82%	267,902	19.99%
2009	1,713,696	1,307,519	76.30%	47,888	2.79%	358,289	20.91%
2010	2,418,972	1,850,512	76.50%	58,327	2.41%	510,133	21.09%
2011	3,962,343	3,031,365	76.50%	84,243	2.13%	846,735	21.37%
2012	7,617,919	5,933,526	77.89%	141,933	1.86%	1,542,460	20.25%
2013	19,443,201	15,403,511	79.22%	288,817	1.49%	3,750,873	19.29%
2014	30,969,378	24,745,079	79.90%	465,663	1.50%	5,758,636	18.59%
2015	43,271,614	34,329,445	79.33%	629,778	1.46%	8,312,391	19.21%
2016	50,743,595	39,928,762	78.69%	706,216	1.39%	10,108,617	19.92%
2017	49,206,890	38,196,178	77.62%	682,690	1.39%	10,328,022	20.99%
2018	51,863,757	40,050,879	77.22%	666,664	1.29%	11,146,214	21.49%
2019	65,732,087	49,265,567	74.95%	977,322	1.49%	15,489,198	23.56%
2020	73,271,141	51,988,791	70.95%	1,526,626	2.08%	19,755,724	26.96%
2021	75,268,993	53,408,674	70.96%	1,646,047	2.19%	20,214,272	26.86%
2022	64,470,243	47,018,131	72.93%	1,405,504	2.18%	16,046,608	24.89%
2023	25,675,880	17,847,140	69.51%	716,735	2.79%	7,112,005	27.70%
Total	571,544,746	429,030,789	75.07%	10,309,129	1.80%	132,204,828	23.13%

* Note: Total 571,544,746 reviews from 54,514,263 unique reviewers for 48,185,048 unique products between 17 June 1996 and 14 September 2023.

As shown in Table 3, the distribution of ratings across all 571,544,746 reviews is heavily skewed toward 1-star and 5-star ratings. Specifically, the share of 1-star ratings in OTA reviews (20.35%) is 11.56 percentage points higher than in NOT reviews (8.78%). The share of 1-star ratings in OTNOTA reviews (14.48%) is 5.70% higher than in NOT reviews (8.78%). Overall, the highest proportion of 1-star ratings occurs in OTA, followed by OTNOTA, then NOT. Conversely, the share of 5-star ratings in OTA reviews (58.11%) is 7.75 percentage points lower than in NOT reviews (65.86%). The share of 5-star ratings in OTNOTA reviews (62.89%) is 2.97% lower than in NOT reviews (65.86%). The highest proportion of 5-star ratings is in NOT, followed by OTNOTA and OTA. The difference in the share of extreme ratings is greater between OTA and NOT than between OTNOTA and NOT.

The share of 1-star ratings has increased over time in NOT, OTA, and OTNOTA, though there are year-to-year fluctuations. In 2020, the share of 1-star ratings for NOT, OTA, and OTNOTA rose by 1.31%, 1.19%, and 1.6%, respectively, compared to 2019. Relative to NOT, OTA's share of 1-star ratings decreased less, while OTNOTA showed a pronounced increase between 2019 and 2020. By 2023, the shares of 1-star ratings are 10.31% for NOT, 29.20% for OTA, and 18.83% for OTNOTA.

The shares of 5-star ratings for NOT, OTA, and OTNOTA fluctuated before 2019. Since then, these shares have generally declined, although there are still annual increases and decreases. In 2020, compared to 2019, the share of 5-star ratings for NOT decreased by 2.04%, OTA by 1.03%, and OTNOTA by 2.2%. Currently, the share of 5-star ratings stands at 65.59% for NOT, 49.42% for OTA, and 59.08% for OTNOTA.

In summary, the share of 1-star ratings for OTA and OTNOTA has grown over time, especially for OTA since 2000 and for OTNOTA since 1999, compared to NOT. The share of 5-star ratings for NOT has also increased over time, compared to OTA since 2012 and OTNOTA since 2013.

Table 3. Rating distribution of NOT, OTA, and OTNOTA reviews over time (Appendix)

Year	All Reviews														
	NOT					OT									
						OTA					OTNOTA				
	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
1996	0.00%	0.00%	4.17%	16.67%	79.17%	NA	NA	NA	NA	NA	0.00%	0.00%	0.00%	50.00%	50.00%
1997	3.42%	5.23%	7.01%	21.81%	62.53%	2.45%	2.34%	3.34%	13.71%	78.15%	2.78%	5.05%	4.55%	15.66%	71.97%
1998	4.29%	4.59%	7.72%	20.99%	62.41%	3.22%	2.81%	3.80%	14.20%	75.97%	3.72%	2.61%	4.61%	16.42%	72.65%
1999	5.32%	5.10%	8.70%	19.84%	61.04%	4.65%	2.91%	4.43%	13.47%	74.54%	5.44%	3.15%	4.86%	15.83%	70.72%
2000	4.33%	4.95%	9.75%	23.94%	57.03%	5.59%	3.60%	5.54%	16.29%	68.97%	5.88%	4.30%	6.40%	18.32%	65.10%
2001	4.77%	5.28%	10.28%	24.24%	55.44%	6.61%	3.99%	5.83%	15.79%	67.79%	7.82%	5.30%	7.02%	17.69%	62.17%
2002	5.24%	5.53%	10.24%	23.66%	55.33%	7.38%	4.33%	5.52%	15.53%	67.24%	9.34%	5.91%	6.65%	16.53%	61.57%
2003	6.04%	5.76%	10.36%	23.06%	54.78%	9.20%	4.42%	5.81%	15.09%	65.47%	10.08%	6.03%	6.87%	16.59%	60.42%
2004	7.70%	6.01%	10.19%	22.09%	54.01%	11.13%	4.99%	6.28%	15.31%	62.29%	12.82%	6.21%	7.11%	15.70%	58.16%
2005	8.18%	6.14%	10.25%	21.57%	53.87%	12.33%	4.94%	6.37%	15.77%	60.60%	13.05%	6.27%	7.40%	16.80%	56.48%
2006	7.82%	5.98%	9.79%	21.49%	54.92%	9.60%	4.25%	5.39%	14.30%	66.47%	12.04%	6.07%	7.17%	17.00%	57.73%
2007	6.64%	5.38%	9.17%	21.30%	57.52%	7.99%	3.88%	5.77%	17.58%	64.78%	9.12%	5.01%	7.13%	18.57%	60.16%
2008	7.32%	5.70%	9.23%	21.22%	56.53%	9.83%	4.30%	5.43%	16.92%	63.51%	10.13%	5.34%	6.99%	18.62%	58.93%
2009	7.81%	5.72%	9.30%	20.66%	56.52%	10.72%	4.32%	5.79%	17.00%	62.16%	10.69%	5.48%	7.04%	17.97%	58.83%
2010	8.37%	5.89%	9.29%	20.33%	56.12%	12.83%	5.16%	6.56%	16.44%	59.01%	11.76%	5.77%	7.62%	17.52%	57.34%
2011	8.77%	5.86%	9.12%	19.62%	56.63%	14.57%	5.62%	6.63%	15.86%	57.32%	12.36%	6.02%	7.49%	16.86%	57.27%
2012	7.82%	5.21%	8.68%	19.16%	59.13%	13.37%	5.27%	6.60%	16.33%	58.43%	11.17%	5.51%	7.34%	16.74%	59.24%
2013	5.96%	4.66%	8.87%	19.46%	61.05%	12.45%	5.51%	6.90%	16.60%	58.54%	9.17%	5.24%	7.87%	17.31%	60.41%
2014	6.13%	4.44%	8.26%	17.70%	63.47%	12.33%	4.75%	6.53%	16.60%	59.79%	9.65%	5.00%	7.36%	15.99%	62.00%
2015	6.48%	4.36%	7.75%	16.19%	65.23%	13.32%	4.72%	5.91%	14.71%	61.33%	10.14%	4.91%	6.87%	14.48%	63.59%
2016	6.71%	4.37%	7.47%	15.54%	65.92%	14.26%	4.83%	5.55%	13.42%	61.96%	10.61%	4.94%	6.58%	13.49%	64.38%
2017	7.72%	4.65%	7.28%	14.26%	66.09%	17.19%	5.29%	5.36%	11.46%	60.70%	12.55%	5.40%	6.46%	11.87%	63.73%
2018	8.70%	4.73%	7.00%	12.92%	66.65%	20.79%	5.60%	5.08%	9.93%	58.59%	14.35%	5.47%	6.14%	10.46%	63.57%
2019	8.50%	4.12%	6.08%	11.12%	70.18%	18.64%	4.62%	4.56%	9.61%	62.57%	13.07%	4.56%	5.23%	9.55%	67.59%
2020	9.81%	4.66%	6.61%	10.78%	68.14%	19.83%	4.86%	4.76%	9.02%	61.54%	14.67%	5.07%	5.70%	9.17%	65.39%
2021	11.31%	5.20%	7.17%	10.49%	65.84%	23.75%	5.88%	5.33%	8.26%	56.79%	17.40%	5.94%	6.46%	8.55%	61.65%
2022	11.56%	5.36%	7.55%	11.15%	64.38%	29.30%	7.04%	6.08%	7.44%	50.15%	19.85%	6.64%	7.05%	8.26%	58.20%
2023	10.31%	4.95%	7.24%	11.90%	65.59%	29.20%	7.56%	6.33%	7.49%	49.42%	18.83%	6.59%	6.98%	8.53%	59.08%
Sum	**8.78%**	4.73%	7.29%	13.34%	**65.86%**	**20.35%**	5.56%	5.52%	10.47%	**58.11%**	**14.48%**	5.48%	6.42%	10.74%	**62.89%**

* Note: Total 571,544,746 reviews from 54,514,263 unique reviewers for 48,185,048 unique products between 17 June 1996 and 14 September 2023.

3.2 Reviewer heterogeneity across categories

As shown in Table 4, all 34 product categories include reviews from OT (i.e., OTA and OTNOTA). OT reviews make up 24.93% of the total review dataset. Specifically, 142,513,957 reviews are authored by 50,939,994 unique OTs for 20,208,041 unique products between 15 November 1996 and 14 September 2023. The rating distribution of OT reviews is similarly polarized, with 14.91% receiving 1-star and 62.54% receiving 5-star ratings.

All product categories also include reviews from OTA. Between 8 August 1997 and 12 September 2023, there were 10,309,129 reviews from 10,309,129 unique OTAs, identifying 1,908,223 unique products. OTA reviews make up 1.8% of the entire review dataset. The proportion of 1-star ratings for OTA exceeds that for NOT across all product categories. Conversely, the share of 5-star ratings for OTA is lower than that for NOT, except in five product categories. OTA reviews tend to be more negatively polarized than NOT reviews across all categories, while NOT reviews are generally more positive than OTA reviews, except in a few categories. Notably, the largest gap in the share of 1- and 5-star ratings between NOT and OTA is in the Subscription Boxes category, at 9.58%, with OTA ratings significantly more extreme than those for NOT.

OTNOTA appears in all categories. A total of 132,204,828 reviews from 40,630,865 unique OTNOTA are identified between 15 November 1996 and 14 September 2023, covering 20,000,660 unique products. The share of 1-star ratings for OTNOTA is consistently higher than NOT across all product categories. The share of 5-star ratings for OTNOTA is usually lower than for NOT, except in five categories. The largest gap in the share of 1- and 5-star ratings between NOT and OTNOTA is in the Books category at 8.19%, where OTNOTA has a higher share of extreme ratings than NOT. Overall, one-time reviews tend to show more negative polarity across all categories and less positive polarity in most cases compared to multi-review reviewers.

Table 4. Rating distribution of NOT, OTA, and OTNOTA reviews in each product category (Appendix)

	Product Category	All Reviews														
		NOT					OT									
							OTA					OTNOTA				
		Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
1	All Beauty	11.56%	5.82%	9.01%	13.46%	60.15%	21.79%	5.54%	5.14%	9.13%	58.40%	15.05%	6.21%	7.87%	10.91%	59.96%
2	Amazon Fashion	9.96%	6.75%	10.65%	15.61%	57.03%	20.56%	6.63%	7.22%	12.10%	53.49%	13.01%	7.02%	9.47%	13.43%	57.07%
3	Appliances	8.64%	3.18%	4.87%	10.49%	72.82%	24.87%	4.51%	3.94%	8.61%	58.07%	12.75%	3.97%	4.92%	9.51%	68.85%
4	Arts Crafts and Sewing	7.61%	4.10%	6.64%	10.70%	70.94%	22.04%	6.19%	6.14%	9.17%	56.45%	13.65%	5.65%	7.06%	10.19%	63.45%
5	Automotive	9.94%	3.97%	5.92%	11.15%	69.01%	24.86%	5.30%	5.04%	9.20%	55.60%	16.28%	5.21%	6.07%	10.08%	62.36%
6	Baby Products	7.63%	5.10%	7.78%	12.31%	67.17%	19.22%	5.68%	5.30%	8.70%	61.10%	11.62%	5.32%	6.51%	10.15%	66.40%
7	Beauty and Personal Care	10.21%	5.25%	7.59%	11.53%	65.42%	23.22%	5.29%	5.04%	8.72%	57.73%	15.30%	5.55%	6.27%	9.59%	63.29%
8	Books	3.86%	3.67%	7.41%	16.86%	68.19%	8.10%	3.85%	5.15%	13.21%	69.69%	6.75%	3.63%	5.21%	10.92%	73.49%
9	CDs and Vinyl	3.27%	2.88%	6.29%	15.05%	72.50%	8.68%	3.35%	4.29%	10.67%	73.01%	5.76%	2.98%	4.48%	9.26%	77.53%
10	Cell Phones and Accessories	11.83%	5.87%	7.78%	12.29%	62.23%	22.12%	5.89%	5.34%	9.92%	56.72%	15.98%	6.29%	6.87%	10.86%	60.00%
11	Clothing Shoes and Jewelry	7.87%	5.37%	8.54%	14.00%	64.23%	17.31%	6.34%	6.83%	11.65%	57.86%	12.69%	6.22%	7.64%	12.35%	61.09%
12	Digital Music	3.37%	2.14%	5.58%	14.00%	74.92%	13.80%	3.78%	5.73%	7.67%	69.02%	5.25%	2.55%	4.56%	9.33%	78.31%
13	Electronics	10.75%	5.00%	6.76%	13.14%	64.35%	22.07%	5.77%	5.33%	11.23%	55.60%	16.37%	5.61%	6.02%	11.10%	60.89%
14	Gift Cards	2.89%	0.50%	1.44%	4.06%	91.11%	27.57%	2.37%	3.24%	5.50%	61.32%	8.93%	1.40%	2.31%	4.45%	82.92%
15	Grocery and Gourmet Food	10.23%	4.65%	6.58%	9.54%	69.00%	32.41%	5.73%	4.88%	5.92%	51.05%	17.82%	5.27%	5.68%	7.25%	63.99%
16	Handmade Products	4.23%	2.34%	3.86%	6.82%	82.75%	11.97%	3.67%	3.55%	4.71%	76.11%	6.70%	3.28%	4.69%	6.84%	78.50%
17	Health and Household	9.06%	4.45%	6.49%	11.17%	68.83%	20.09%	4.47%	4.25%	9.22%	61.97%	14.38%	5.07%	5.76%	9.77%	65.01%
18	Health and Personal Care	9.78%	5.15%	8.12%	14.71%	62.25%	23.35%	5.20%	4.73%	9.09%	57.63%	14.47%	5.95%	7.44%	11.17%	60.97%
19	Home and Kitchen	9.32%	4.84%	6.90%	11.40%	67.53%	23.78%	6.55%	5.77%	9.06%	54.84%	16.11%	6.07%	6.53%	10.22%	61.06%
20	Industrial and Scientific	8.15%	3.93%	6.04%	11.68%	70.20%	26.68%	5.90%	5.29%	8.68%	53.46%	13.89%	5.08%	6.17%	10.04%	64.82%
21	Kindle Store	2.59%	3.12%	7.74%	21.66%	64.88%	4.76%	6.08%	6.13%	17.96%	68.00%	5.29%	3.65%	6.40%	16.08%	68.57%
22	Magazine Subscriptions	7.50%	4.43%	7.64%	13.85%	66.59%	30.60%	3.15%	7.46%	10.03%	45.83%	15.20%	5.91%	6.81%	10.67%	61.41%
23	Movies and TV	7.33%	4.59%	7.89%	16.03%	64.15%	13.28%	3.53%	4.49%	11.27%	67.44%	10.76%	3.86%	5.14%	11.51%	68.74%
24	Musical Instruments	6.53%	3.97%	6.74%	14.09%	68.67%	18.28%	5.26%	5.46%	12.08%	58.91%	11.37%	4.84%	6.30%	12.20%	65.29%
25	Office Products	7.96%	4.10%	6.27%	11.46%	70.22%	23.15%	5.82%	5.26%	8.87%	56.89%	13.71%	5.25%	6.20%	10.18%	64.66%
26	Patio Lawn and Garden	11.60%	4.92%	6.67%	11.27%	65.54%	27.98%	6.46%	5.42%	9.05%	51.10%	17.75%	5.92%	6.42%	9.94%	59.97%
27	Pet Supplies	10.34%	5.50%	7.71%	11.19%	65.25%	23.50%	6.36%	5.36%	8.51%	56.27%	15.97%	6.15%	6.65%	9.55%	61.68%
28	Software	12.22%	4.85%	9.21%	18.23%	55.49%	15.25%	5.71%	6.72%	17.97%	56.36%	18.17%	5.11%	7.54%	16.22%	52.95%
29	Sports and Outdoors	7.50%	4.40%	6.88%	13.58%	67.63%	19.65%	7.13%	5.33%	10.65%	58.67%	12.53%	5.43%	6.60%	11.52%	63.92%
30	Subscription Boxes	13.81%	7.79%	10.6%	10.49%	57.31%	18.16%	3.70%	5.75%	6.44%	62.53%	16.07%	9.43%	10.99%	10.34%	53.16%
31	Tools and Home Improvement	9.45%	4.48%	6.40%	11.92%	67.75%	24.92%	6.03%	5.39%	9.14%	54.53%	15.73%	5.58%	6.22%	10.13%	62.35%
32	Toys and Games	7.66%	4.24%	6.78%	11.51%	69.81%	23.38%	6.27%	5.96%	9.15%	55.25%	14.53%	5.72%	6.85%	10.14%	62.75%
33	Unknown	11.19%	5.21%	7.66%	13.89%	62.05%	18.47%	4.91%	5.51%	11.76%	59.36%	17.24%	5.79%	6.65%	11.59%	58.73%
34	Video Games	9.54%	5.05%	7.92%	14.64%	62.85%	24.43%	6.18%	5.81%	10.76%	52.82%	16.35%	5.85%	6.67%	11.72%	59.41%
	Sum	8.78%	4.73%	7.29%	13.34%	65.86%	20.35%	5.56%	5.52%	10.47%	58.11%	14.48%	5.48%	6.42%	10.74%	62.89%

* Note: Total 571,544,746 reviews from 54,514,263 unique reviewers for 48,185,048 unique products between 17 June 1996, and 14 September 2023.

Long-term analysis of big data across all categories reveals that (1) one-time reviewers exist in every category, and the share of OTNOTA opinions surpasses that of OTA, (2) their review counts and overall review share have generally increased over time, (3) one-time reviewers tend to exhibit more negative polarity across all categories and show less positive polarity in most categories compared to multi-review reviewers. Overall, these descriptive findings suggest potential differences in reviewer ratings between one-time and multi-review reviewers. Furthermore, the Subscription Boxes category displays the largest gap in the share of extreme ratings between OTA and NOT. Based on this, the study initially selects the Subscription Boxes category as the target product group. It conducts empirical analysis to identify factors, including reviewer heterogeneity, that influence consumers' star ratings of a product.

4. Empirical Analysis

This section presents the findings of an empirical analysis estimating the effects of reviewer heterogeneity, price sentiment, and pre-purchase information on consumers' star ratings in the Subscription Boxes category. This category exhibits an abnormal rating distribution compared to other categories, may have fewer incentives for promotional reviews, and faces fewer constraints. Consumers and sellers might also want to explore a specific product category to find relevant products for their purchase or business. Furthermore, the empirical analysis framework for the Subscription Boxes category can be extended to other product categories.

The section first outlines the Subscription Boxes category for empirical analysis and examines reviewer heterogeneity within it. Next, it discusses price sentiment analysis of review text using LLMs, including manual annotation and the results. Third, it introduces variables, the model, and empirical findings. Specifically, an ordered probit model is estimated, and marginal effects are used for interpretation.

4.1 Target category selection and its reviewer heterogeneity

Amazon Subscription Boxes often offer themed collections of products curated by experts, such as beauty items, snacks, and educational toys, to subscribers. Consumers may feel more uncertain about the quality of these collections compared to regular products in other categories because the selection changes with each subscription cycle. Therefore, the value and influence of pre-purchase information, such as reviews from other consumers of Subscription Boxes, can be significant.

Previous studies (Luca & Zervas, 2016; Mayzlin et al., 2014) have noted that suspicious reviewers may post positive reviews for their own businesses and negative reviews for their competitors. He et al. (2022) reported that fake reviewers are more likely to submit positive reviews than negative ones for competitors because targeting competitors on Amazon.com is more costly. Therefore, Subscription Boxes may offer less incentive for promotional reviewers—who aim to boost their products with positive reviews or undermine competitors with negative ones—to write positive promotional reviews for their business-related products in the Subscription Boxes category compared to other categories. This is because the subscription box is not for a single product but for a collection of products that often change with each subscription cycle. Additionally, promotional reviewers might have less motivation to write negative reviews of competitors in the Subscription Boxes category, as the cost of posting a negative review after subscription may be higher than the cost for a one-time purchase of other products in different categories. The Subscription Boxes category exhibits the greatest difference between OTA and NOT reviews in the share of extreme ratings across all categories. As shown in Table 5, a total of 16,216 reviews from 15,237 unique reviewers for 641 unique products were identified between 24 January 2017 and 30 August 2023. The first review in the Subscription Boxes category appeared on 24 January 2017, during the entire sample period.

Table 5 shows that the gap in the share of extreme ratings between NOT (71.12%) and OTA reviews (80.69%) in this category reaches 9.57%, the largest difference observed across all categories. Specifically, the share of 1-star ratings in OTA reviews within the category (18.16%) is 4.35 percentage points higher than the share across all reviews in the same category (13.81%). Similarly, the share of 5-star ratings in OTA reviews within the category (62.53%) is 5.22 percentage points higher than the share in the same category (57.31%). The gap in the share of extreme ratings between NOT (71.12%) and OTNOTA reviews (69.23%) is 1.89 percentage points. The share of 1-star ratings in OTNOTA reviews within the category (16.07%) is 2.26 percentage points higher than the share across all reviews in the same category (13.81%). In contrast, the share of 5-star ratings in OTNOTA reviews within the category (53.16%) is 4.15 percentage points lower than the share across all reviews in the same category (57.31%).

Overall, one-time reviewers in the Subscription Boxes category display more negative polarity and less positive polarity than multi-review reviewers. Additionally, the gap in the share of extreme ratings between OTA and NOT is the largest across all product categories. Therefore, this study chooses the Subscription Boxes category as the target for empirical analysis, as it shows significant reviewer heterogeneity in extreme ratings.

Table 5. Rating distribution of reviews in the Subscription Boxes category

Subscription Boxes Category	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Count (% in All)	Share of 1- and 5- ratings	Unique reviewers	Unique items
All (1.OT + 2.NOT)	2,576 (15.89%)	1,491 (9.19%)	1,752 (10.80%)	1,663 (10.26%)	8,734 (53.86%)	16,216 (100%)	69.75%	15,237	641
1. OT (OTA+OTNOTA)	2,335 (16.14%)	1,355 (9.36%)	1,567 (10.83%)	1,480 (10.23%)	7,734 (53.44%)	14,471 (89.24%)	69.58%	14,471	602
1.1 OTA	79 (18.16%)	31 (7.13%)	25 (5.75%)	28 (6.44%)	272 (62.53%)	435 (2.68%)	80.69%	435	61
1.2 OTNOTA	2,256 (16.07%)	1,324 (9.43%)	1,542 (10.99%)	1,452 (10.34%)	7,462 (53.16%)	14,036 (86.56%)	69.23%	14,036	601
2. NOT	241 (13.81%)	136 (7.79%)	185 (10.60%)	183 (10.49%)	1,000 (57.31%)	1,745 (10.76%)	71.12%	766	340
2.1 NOT (first review)*	116 (15.14%)	52 (6.79%)	79 (10.31%)	75 (9.79%)	444 (57.96%)	766 (4.72%)	73.11%	766	251

* Note: The number of OTA, OTNOTA, NOT (first review) reviewers, and the number of reviews are the same because only the first review from each unique reviewer in each subgroup is considered for a fair comparison. NOT (first review) counts only the initial review from each unique NOT reviewer in the Subscription Boxes category.

In addition, analyzing the entire individual review dataset—using data mining to extract pre-purchase information from crowd reviews, sentiment analysis with generative LLMs, and labor-intensive manual annotation—demands significant computational resources, budgets, and time. Moreover, sellers and consumers may prefer to focus on relevant products within a specific category for their purchases or businesses, rather than considering all products across all categories. Given these reasons and constraints, this study focuses on the Subscription Boxes category for econometric analysis. Furthermore, OTAs are assumed to write their first and only review across all product categories during the entire sample period, while OTNOTA in the category are assumed to write their first and only review for a product within the category (but not the only review across all categories) during the sample period. To ensure a fair comparison between one-time reviewers (i.e., OTA and OTNOTA) and multi-review reviewers (i.e., NOT) in the Subscription Boxes category, this study limits the econometric sample to the first review submitted by each unique reviewer in the category. Consequently, the number of reviews matches the number of unique reviewers for all reviewer types. A total of 15,237 reviews from 766 unique reviewers for 251 unique products were identified between 24 January 2017 and 30 August 2023. However, all 16,216 reviews are used in the data mining process to extract pre-purchase information about product quality from previous reviewers' reviews. Since numerical price information for products in this category is unavailable, each consumer's sentiment toward the price of the reviewed product is classified as 'positive,' 'neutral,' or 'negative' using popular generative LLMs combined with human annotation through prompt engineering, with and without fine-tuning. The detailed process for this price sentiment analysis is described in the following subsection.

4.2 Price Sentiment Analysis Using LLMs combined with manual annotation

Deep learning has been widely used in economic research (Farrell et al., 2021; Ramsey and Bergtold, 2021), online product review analysis (Liu et al., 2019; Timoshenko and Hauser, 2019), and consumer behavior prediction (Safara, 2022). Additionally, employing pre-trained large language models (LLMs) and generative AI (GenAI), which are based on deep learning techniques, has opened new opportunities for analyzing unstructured data in economic studies (Bybee, 2023; Han, 2024; Korinek, 2023; Ludwig et al., 2025), measuring inflation through firms' earnings call transcripts (Chava et al., 2022), identifying consumer needs (Timoshenko et al., 2025), and exploring factors behind Yelp review polarization (Fang et al., 2024). Consequently, this research advances multimodal economic analysis by combining econometric techniques for structured data with generative models to analyze unstructured textual content in user-generated content (UGC).

This study uses widely adopted generative LLMs, specifically OpenAI's GPT models, which have demonstrated impressive performance across various tasks, both with and without fine-tuning on example outputs. Specifically, the study employs GPT models for price sentiment analysis via OpenAI's API. GPT is a pre-trained model based on the Transformer decoder architecture (Vaswani et al., 2017). The Transformer model forms the foundation of most modern LLMs in natural language processing (NLP), originally developed for machine translation, and features an encoder-decoder architecture. The decoder in the Transformer model generates output sequences for different tasks from numerical data produced by the encoder, using unstructured data such as text. Therefore, Transformer-based decoder models, such as GPT, are well-suited for text generation. Kheiri and Karimi (2024) demonstrated the outstanding predictive performance of GPT-3.5 Turbo, a predecessor of GPT-4, which operates only on text data, using prompt engineering for three-class sentiment classification on a Twitter dataset. Prompt engineering involves creating effective prompts for generative LLMs (Korinek,

2023); well-designed prompts can enhance LLM performance for specific tasks (Zhao et al., 2023). As a result, this study employs GPT-4o, GPT-4.1, and GPT-5 models for three-class price sentiment classification via prompt engineering, using Amazon product reviews in the Subscription Boxes category.

As shown in Table 6, the GPT-4o, GPT-4.1, and GPT-5 models are trained on datasets up to 1 October 2023, 1 June 2024, and 30 September 2024, respectively. Therefore, the knowledge cutoff date for all models is after the last date of the sample period in this study. The context window refers to the maximum number of tokens (i.e., the basic units of text) that GPT models can consider at once, including both input and output. Maximum output tokens indicate the highest number of tokens that GPT models can generate in their responses. Temperature measures the randomness of output from GPT models, where 0 means deterministic output, and 1 indicates standard randomness, which is the default setting for GPT-5. In this study, the temperature for GPT-4o and GPT-4.1 is set to 0 to ensure deterministic, reproducible outputs. Fine-tuning is an additional training process applied to pre-trained large language models using a smaller, task- or domain-specific dataset. GPT-4o and GPT-4.1 support fine-tuning, but GPT-5 does not. Fine-tuning requires a dataset tailored to a specific task; therefore, the study uses a manually labeled dataset for price sentiment classification during the fine-tuning of GPT-4o and GPT-4.1.

Table 6. Model comparison between GPT-4o, GPT-4.1, and GPT-5

Parameters	GPT-4o	GPT-4.1	GPT-5
API Model ID	gpt-4o-2024-08-06	gpt-4.1-2025-04-14	gpt-5
Knowledge cutoff	Oct 1, 2023	Jun 1, 2024	Sep 30, 2024
Context window	128,000	1,047,576	400,000
Max output tokens	16,384	32,768	128,000
Temperature flexibility	Yes	Yes	No (Fixed as 1)
Fine-tunning capability	Yes	Yes	No

* Source: OpenAI. Compare models. (modified).

The overall sentiment analysis process in this study follows that of Liu et al. (2019), who initially used a manually labeled, balanced product-review dataset to validate the

predictive performance of deep-learning-based sentiment classification models. Subsequently, Liu et al. (2019) used a trained deep learning model that achieved 93.3% accuracy on a manually labeled test dataset to classify the binary sentiment (i.e., positive vs. negative) across the entire product review dataset. Therefore, this study adopts the aforementioned process for three-class price sentiment analysis rather than binary sentiment classification, and uses GPT models rather than convolutional neural networks (CNNs; Kim, 2014).

This study begins by randomly selecting 1,500 reviews (about 10% of all 15,237 reviews) as a validation dataset to evaluate the predictive performance of GPT models with and without fine-tuning. Before analysis, the titles and bodies of all 15,237 reviews are combined and converted to lowercase. Second, the author then manually annotates 1,500 reviews twice, providing reasons for each review's positive or negative sentiment classification to ensure consistency and minimize human error over 17 days, from 6 October to 22 October 2025. The trigger phrase in each review is manually classified as either positive or negative. In contrast, all neutral reviews are labeled with the reason "there is no information about price in the review" (Table 7). The author has prior experience annotating sentiment in Amazon product reviews. During the manual annotation process to build a dataset for training GPT models on price sentiment analysis and to evaluate their performance, OpenAI's ChatGPT was also used to translate seven Spanish reviews into English. None of these Spanish reviews contained any direct mention of price-related sentiment; therefore, their price sentiments were all classified as neutral. GPT models are multilingual; consequently, they can classify the sentiment of prices in Spanish reviews using English prompts without modification.

Table 7. Manual annotation examples for three-class price sentiment classification

Reason for the manual price sentiment class	Manual price sentiment class
the phrase "well worth the money" indicates a positive sentiment towards the price.	positive
there is no information about price in the review	neutral
the phrase "expensive for quality i was disappointed" indicates a negative sentiment towards the price.	negative

The annotated validation dataset is split into training (1,000 reviews) and test (500 reviews) datasets. As shown in Table 8, the dataset is imbalanced, with most price sentiments neutral (73.87%), followed by negative (16.87%) and positive (9.27%).

Table 8. Sentiment distribution in training, test, and combined training and test datasets

	Training (A)	Test (B)	Total (A+B)
Positive sentiment	92 (9.2%)	47 (9.4%)	139 (9.27%)
Neutral sentiment	737 (73.7%)	371 (74.2%)	1,108 (73.87%)
Negative sentiment	171 (17.1%)	82 (16.4%)	253 (16.87%)
Total	1,000	500	1,500

During the validation phase, the study evaluates the overall predictive performance of three-class sentiment classification across GPT models on the test set, both with and without fine-tuning. This design permits a direct comparison between baseline prompting strategies and model-specific parameter updates.

In this context, the direct use of a pre-trained model with prompt engineering—where a small number of example inputs and outputs are provided within the prompt—is defined as the "few-shot learning" approach (Brown et al., 2020). By contrast, the “zero-shot learning” approach supplies no labeled examples in the prompt. All GPT models are evaluated under zero-shot conditions to classify price sentiment in test-set reviews without fine-tuning. These results are then compared with those obtained from fine-tuned versions of the same models, enabling a structured assessment of how additional supervised training may influence classification accuracy.

Fine-tuning a pre-trained model with labeled data is called supervised fine-tuning in generative LLMs. Using this approach, GPT-4o and GPT-4.1 are fine-tuned on manually labeled training datasets containing 50 or 1,000 samples to classify the sentiment of prices in a 500-review test set. GPT-5 is excluded from fine-tuning because it does not support this process. The prompt used for the fine-tuned model is identical to the one used in the zero-shot approach,

but is applied after fine-tuning. The best price sentiment classification model among GPT models is chosen solely based on its performance on the test set.

Prompts in GPT models consist of two parts: the system prompt and the user prompt. The system prompt in GPT models defines guidelines for how GPT should behave, establishing its AI persona and/or personality. The system prompt for price sentiment remains consistent across all GPT models (see Table 9). In this study, the system first instructs GPT to classify the price sentiment in each Amazon review as positive, neutral, or negative. Next, it directs GPT to categorize reviews without price information or those it cannot classify as neutral. Then, GPT is asked to generate a one-sentence explanation for its sentiment analysis of the price in each review. Afterward, the system prompts GPT to separate the classified price sentiment from the explanation with a comma and to provide all responses in lowercase. Finally, three examples illustrating the desired output formats with different sentiment classes are included.

Table 9. System prompt for price sentiment analysis

You will be provided with Amazon online product reviews. Your task is to classify their sentiment toward price as positive, neutral, or negative. When you cannot classify the sentiment toward price, or there is no information about price in the review, please classify it as neutral. Additionally, please provide the reason for your sentiment analysis of the price in the review in one sentence. Separate the sentiment toward price and reason with a comma. All responses must be in lowercase. The examples of responses are: negative, there is no information about the price in the review. positive, the phrase "worth the money" indicates a positive sentiment toward the price. negative, the phrase "not worth the money" indicates a negative sentiment toward the price.

The user prompt in GPT models refers to the input provided by a user, such as a specific question, request, or task. In this study, each review acts as a user prompt. Custom Python functions utilize the OpenAI API to classify the price sentiment of each review (i.e., user prompt) using a consistent system prompt applied across all GPT models, regardless of fine-tuning. Another Python function splits the GPT model response into two columns: one for price sentiment and the other for the reason, separated by a comma. This function also filters responses to assign a neutral price sentiment whenever the GPT model's reply does not match any of the three categories ('positive', 'neutral', and 'negative'). Instead, it returns other outputs, such as 'none'.

Compared to binary classification of overall sentiment in reviews with a balanced dataset, where sentiment classes are equally distributed, three-class sentiment classification for price in reviews with an imbalanced dataset in this study presents a more challenging NLP task. In particular, the F1-score is a widely used metric for evaluating the predictive performance of classification models on imbalanced datasets. In contrast, accuracy is more commonly used for classification models with balanced datasets. Riyanto et al. (2023) stated that the F1-score is the most appropriate evaluation metric for classification tasks with an imbalanced dataset among the confusion-matrix-based metrics, including accuracy, precision, and recall. Accuracy measures the share of correctly classified samples across all classes relative to the total sample size. Precision is the share of correctly classified samples for a specific class out of all samples that are classified as that class. Recall is the proportion of correctly classified samples in a specific class out of the total sample size for that class. The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a model's accuracy, as follows:

$$\text{F1-score} = \frac{2}{\frac{1}{\text{precision}}+\frac{1}{\text{recall}}} = \frac{2(\text{precision}\cdot\text{recall})}{\text{precision}+\text{recall}}$$

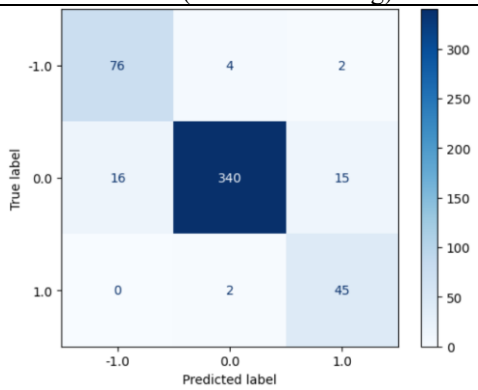
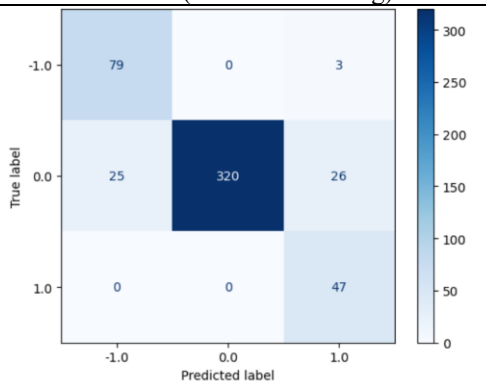
The macro-average scores of precision, recall, and F1-score are used to evaluate the predictive performance of GPT models for three-class price sentiment analysis in this study, as

they equally consider all classes in multi-class classification, regardless of class imbalance. Overall, the macro F1 score is the main metric for assessing GPT models' predictive performance in this study. As shown in Table 10, fine-tuning enhances the predictive performance of GPT-4o and GPT-4.1 across all evaluation metrics except macro recall, even when fine-tuning is done with only 50 samples in the training set. The top model for price sentiment analysis on the test set is GPT-4.1 fine-tuned with 1,000 training samples, achieving a macro F1-score of 93%, an accuracy of 96%, and a macro precision of 94%. The confusion matrices for each GPT model are also shown in Table 11, where the diagonal elements represent the number of correctly classified samples for each class.

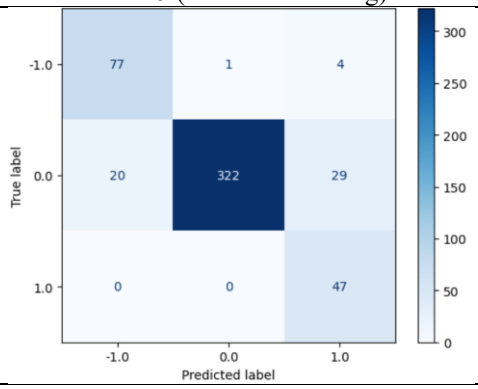
Table 10. Prediction results for price sentiment classification in the test set

Instruction strategy	Model	Accuracy	Macro precision	Macro recall	**Macro F1**
Zero-shot learning	GPT-4o	0.92	0.84	0.93	0.88
	GPT-4.1	0.89	0.79	0.94	0.85
	GPT-5	0.89	0.79	0.94	0.84
Fine-tuning with 50 samples in the training set	GPT-4o	0.95	0.90	0.93	0.91
	GPT-4.1	0.95	0.89	0.95	0.92
Fine-tuning with 1,000 samples in the training set	GPT-4o	**0.96**	0.93	0.89	0.91
	GPT-4.1	**0.96**	0.94	0.93	**0.93**

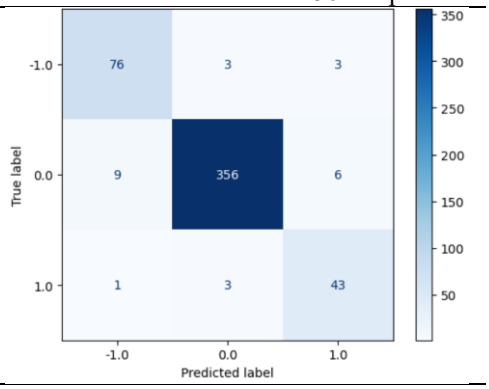
Table 11. Confusion matrix for price sentiment classification in the test set

GPT-4o (zero-shot learning)	GPT-4.1 (zero-shot learning)
	

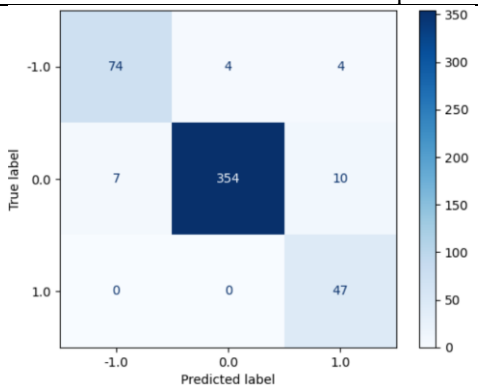
GPT-5 (zero-shot learning)



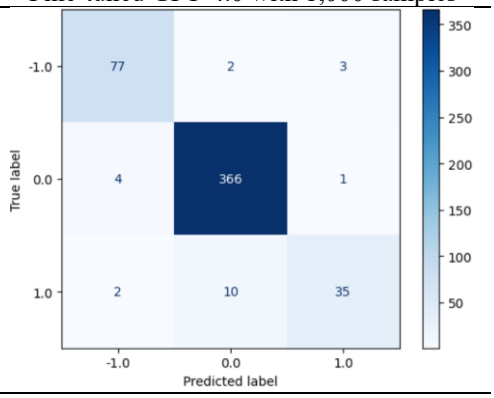
Fine-tuned GPT-4o with 50 samples



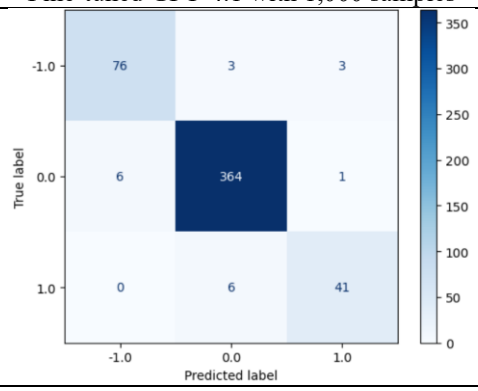
Fine-tuned GPT-4.1 with 50 samples



Fine-tuned GPT-4.0 with 1,000 samples



Fine-tuned GPT-4.1 with 1,000 samples



The price sentiment of 13,737 reviews, excluding 1,500 manually labeled reviews, is classified using the GPT-4.1 model fine-tuned with 1,000 manually labeled training samples. Ultimately, the dataset includes 1,500 manually labeled reviews and 13,737 GPT-labeled reviews. Specifically, the distribution of price sentiment classes is imbalanced across all label groups, with a pronounced skew toward neutral sentiment (Table 12). The combined price sentiment variables from both manual and GPT labels are used as independent variables in the econometric analysis in this study.

Table 12. Price sentiment distribution across manual, AI, and combined label sets

	Manual label (A)	GPT-4.1 label (B)	Total (A+B)
Positive sentiment	139 (9.27%)	1,426 (10.38%)	1,565(10.27 %)
Neutral sentiment	1,108 (73.87%)	10,071 (73.31%)	11,179(73.37%)
Negative sentiment	253 (16.87%)	2,240 (16.31%)	2,493(16.36%)
Total	1,500	13,737	15,237

4.3 Ordered Probit Model Analysis and Results

This paper uses a random utility model (RUM) for ordinal responses because consumers' star ratings for reviewed products on Amazon.com are ordinal in nature. The ordered probit model is a broadly used random utility model for sample data with an ordinal dependent variable (Chen and Tsai, 2012; Hausman et al., 1992; Greene, 2012; Kawakatsu and Largey, 2009; McCullagh, 1980; Train, 2009; Winship and Mare, 1984; Pudney and Shields, 2000). Hence, this study also uses an ordinal probit model to estimate the effect of reviewer heterogeneity between one-time and multi-review reviewers, price sentiment, and pre-purchase information on consumers' star ratings for Amazon Subscription Boxes.

The notation of the ordered probit model in this study follows Greene (2012). Specifically, the model in this study assumes that the observed ordinal star rating R_{ijt} In an Amazon product review for a subscription box, product 'j' is rated by a reviewer 'i' at time 't', ranging from one to five stars. The ordered probit model also assumes the presence of an unobserved continuous dependent variable. U_{ijt}^{*} on the real line, which represents the reviewer's unobservable true strength of preference (i.e., latent utility) toward the reviewed subscription box 'j' at the time 't'. Additionally, the latent utility, U_{ijt}^{*} is assumed as a linear function of observable independent variables (i.e., x_{ijt}) and its independent and identically distributed (i.i.d.) error terms (i.e., ε_{ijt}) follows a standard normal distribution as follows:

$$U_{ijt}^{*} = x_{ijt}\,\beta + \varepsilon_{ijt},$$

where β denotes unknown coefficients for the independent variables.

Importantly, the ordinal probit model assumes that the observable star rating. R_{ijt} In actual Amazon product reviews, it represents a continuous range of true latent utility. U_{ijt}^{*} over the real line as follows:

$$R_{ijt} = 1, \text{ if } \quad -\infty < U_{ijt}^{*} \leq t_1$$

$$R_{ijt} = 2, \text{ if } \quad t_1 < U_{ijt}^{*} \leq t_2$$

$$R_{ijt} = 3, \text{ if } \quad t_2 < U_{ijt}^{*} \leq t_3$$

$$R_{ijt} = 4, \text{ if } \quad t_3 < U_{ijt}^{*} \leq t_4$$

$$R_{ijt} = 5, \text{ if } \quad t_4 < U_{ijt}^{*} \leq \infty$$

where t_1, t_2, t_3, and t_4 represent latent thresholds required to map a range of the latent utility into observable five-star ratings, and these thresholds follow a strict ascending order:

$$t_1 < t_2 < t_3 < t_4 < t_5$$

Additionally, maximum likelihood estimation (MLE) is used to estimate unknown parameters, including both the coefficients of observable independent variables and the latent thresholds.

Table 13 shows the variables used in the ordered probit model. The dependent variable is a star rating in an Amazon product review. The independent variables fall into two categories: consumer-specific variables and pre-purchase information about the reviewed products from previous reviewers (i.e., crowd). The first category includes consumer-specific attributes at and after the time of review. These include dummy variables for reviewer type to capture reviewer diversity, the length of the review title and body, and the reviewer's price sentiment expressed in both the title and body. Reviewer type variables include binary dummy variables indicating purchase unverified NOT, purchase verified NOT, purchase unverified NOT, purchase verified OTA, and purchase unverified OTA, with the base group being purchase verified multi-review reviewers (i.e., verified NOT). Generally, this study assumes that purchase-verified reviews are less likely to be written by suspicious reviewers, as purchasing the product to write a

promotional review may increase the cost. Additionally, these variables aim to measure the effect of reviewer heterogeneity on each star rating, especially for extreme ratings (i.e., 1- and 5-star ratings). The lengths of the review title and body serve as indicators of reviewers' willingness to express and share their opinions about the reviewed products with others, since longer titles and/or bodies typically require more time and effort than shorter ones.

Price sentiment variables measure reviewers' opinions on whether the product's price is worth it. They are split into two binary variables indicating positive and negative sentiments. The neutral sentiment serves as the baseline for each price sentiment variable. Economically, a positive price sentiment is likely to increase the likelihood of a 5-star rating, while a negative price sentiment is likely to increase the likelihood of a 1-star rating.

The other independent variables representing pre-purchase information about the quality of the reviewed products, based on previous reviews, are constructed using digital footprints derived from all 16,216 reviews, not only the first reviews by unique reviewers in the Subscription Boxes category. This study assumes that a reviewer may be influenced by the previous review's title, body, average rating, and the share of each rating when assigning a star rating to a product at '*t*'. Therefore, this study includes the average length of previous other reviews' title for the revied product by '*t-1*', the average length of previous other reviews' body text for the reviewed product by '*t-1*', the average rating of previous reviews for the reviewed product by '*t-1*', the share of each rating from previous reviews for the reviewed products by '*t-1*', excluding the share of 3-star ratings as it is a minority class. Additionally, the measure of positive imbalance (i.e., the share of 5- and 4-star ratings) in previous reviews of the reviewed product by 't-1' is considered, following Schoenmueller et al. (2020). If a reviewer in the Subscription Boxes category responds to previous reviewers' reviews, the marginal effects of these pre-purchase information variables will be statistically significant.

Furthermore, store ID dummy variables are included to capture the time-invariant

characteristics of Subscription Boxes sold by each store. Dummy variables for year, month, and day of the week are also incorporated. Specifically, the base year for the year dummy variables is 2019, which is one year before the COVID-19 pandemic. If the pandemic influenced consumers' star rating behaviors on Amazon.com, the marginal effect of the 2020 dummy variable will be statistically significant. Additionally, Bayerl et al. (2025) demonstrate that submitting ratings on weekends negatively impacts star ratings across platforms. Therefore, the day-of-week variables are also considered.

Table 13. Variable description (*N*=15,237)

Variable	Description
Star rating	Star rating for the reviewed product at t (dependent variable)
Unverified NOT (n=53; 0.35%)	Whether the reviewer is an unverified NOT or verified NOT (n=713; 4.68%)
Verified OTNOTA (n=12,343; 81.01%)	Whether the reviewer is a verified OT or verified NOT
Unverified OTNOTA (n=1,693; 11.11%)	Whether the reviewer is an unverified NOT or a verified NOT
Verified OTA (n=304; 2%)	Whether the reviewer is a verified OTA or verified NOT
Unverified OTA (n=131; 0.86%)	Whether the reviewer is an unverified OTA or a verified NOT
Length of review title	Length of review title for the product by i
Length of review body	Length of review body for the product by i
Crowd's average length of review title	Crowd's average length of review title for the product by t-1
Crowd's average length of review body	Crowd's average length of review for the product by t-1
Negative price sentiment	Whether negative or neutral price sentiment toward the product by i at t
Positive price sentiment	Whether positive or neutral price sentiment toward the product by i at t
Crowd's average rating	Crowd's average rating for the product by t-1
Crowd's share of 5-star ratings	Crowd's share of 5 ratings on the product by t-1
Crowd's share of 4-star ratings	Crowd's share of 4 ratings on the product by t-1
Crowd's share of 2-star ratings	Crowd's share of 2 ratings on the product by t-1
Crowd's share of 1-star ratings	Crowd's share of 1 rating on the product by t-1
Crowd's positive imbalance	Crowd's share of 5-and 4-star ratings on the product by t-1
Store id	Multiple dummy variables for the effect of the product's store
Year	Multiple dummy variables for year effect compared to 2019 (base)
Month	Multiple dummy variables for month effect compared to January (base)
Day of week	Multiple dummy variables for day of week compared to Monday (base)

* Note: 'i' indicates a reviewer or their first review in the Subscription Boxes category; 't' indicates review time, including year, month, day, hour, minute, and second. The crowd indicates the previous reviewers.

Unlike linear regression models, the coefficients of independent variables in ordered probit models cannot be directly interpreted as marginal effects because these models are nonlinear. Therefore, this study conducts marginal effect analysis to facilitate interpretation.

Before analyzing the marginal effect results, the model specification, model-fit, and the signs and magnitudes of the coefficients for independent variables across models are described to identify the best model based on model-fit measures and the robustness of the results. As shown in Table 14, all models include dummy variables for store, year, month, and day of the week. Except for Model 1, the other models incorporate positive and negative price sentiment variables. Specifically, Model 2 adds positive and negative sentiments to Model 1. Models 3 through 8 each add one crowd rating variable, such as average star rating, the share of star ratings, or positive imbalance, to Model 2. Model-fit is assessed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), where lower scores on both measures indicate better model-fit.

Model 2 demonstrates a better model-fit than Model 1, with AIC and BIC scores that are 2,913.2 and 2,898.1 points lower, respectively. Models 3, 4, 6, and 8 show improved model-fit compared to Model 2. Notably, Model 8 has the best model-fit among all models, with the lowest AIC and BIC scores; therefore, the interpretation of marginal effects will be based on Model 8 in the following subsection.

The sign, magnitude, and statistical significance of the independent variables in Model 1, a baseline model without sentiment and crowd variables, are the same as or similar to those in other models, except for the unverified OTA variable, which is insignificant in Model 1 but statistically significant in the other models. The robustness of results across different specifications suggests that potential omitted-variable bias or measurement error in LLM-generated price sentiment variables is likely small or negligible when interpreting the overall trends of the marginal effects in the following subsection.

Table 14. Results from ordered probit model (Appendix)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Unverified NOT reviewers	-0.4818*** (0.1760)	-0.5249*** (0.1785)	-0.5174*** (0.1787)	-0.5203*** (0.1786)	-0.5235*** (0.1785)	-0.5271*** (0.1786)	-0.5213*** (0.1786)	-0.5176*** (0.1788)
Verified OT reviewers	-0.1599*** (0.0487)	-0.1621*** (0.0495)	-0.1585*** (0.0495)	-0.1588*** (0.0495)	-0.1622*** (0.0495)	-0.1636*** (0.0495)	-0.1604*** (0.0495)	-0.1620*** (0.0496)
Unverified OT reviewers	-0.4650*** (0.0569)	-0.4515*** (0.0580)	-0.4452*** (0.0581)	-0.4473*** (0.0581)	-0.4512*** (0.0580)	-0.4539*** (0.0580)	-0.4481*** (0.0581)	-0.4493*** (0.0582)
Verified OTA reviewers	-0.1885** (0.0846)	-0.2013** (0.0867)	-0.1970** (0.0867)	-0.1981** (0.0867)	-0.2012** (0.0867)	-0.2027** (0.0867)	-0.1991** (0.0867)	-0.2006** (0.0868)
Unverified OTA reviewers	-0.1706 (0.1244)	-0.2270* (0.1290)	-0.2297* (0.1292)	-0.2338* (0.1292)	-0.2251* (0.1291)	-0.2299* (0.1291)	-0.2263* (0.1291)	-0.2315* (0.1292)
Length of review title	-0.0046*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)	-0.0045*** (0.0006)
Length of review body	-0.0003*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)	-0.0002*** (0.00003)
Crowd's average length of review title	-0.0044 (0.0028)	-0.0009 (0.0029)	0.0002 (0.0029)	0.0001 (0.0029)	-0.0009 (0.0029)	0.0001 (0.0029)	-0.0007 (0.0029)	0.0002 (0.0030)
Crowd's average length of review body	-0.0002 (0.0001)	-0.0002 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0002 (0.0001)	-0.0001 (0.0001)	-0.0002 (0.0001)	-0.0001 (0.0001)
2017	-0.2285*** (0.0650	-0.1628** (0.0661)	-0.1326** (0.0668)	-0.1414** (0.0665)	-0.1606** (0.0662)	-0.1396** (0.0665)	-0.1530** (0.0665)	-0.1251* (0.0669)
2018	-0.2807*** (0.0391)	-0.2317*** (0.0400)	-0.2128*** (0.0404)	-0.2176*** (0.0403)	-0.2301*** (0.0401)	-0.2238*** (0.0401)	-0.2250*** (0.0403)	-0.2087*** (0.0406)
2020	-0.0942*** (0.0323)	-0.0534 (0.0332)	-0.0479 (0.0332)	-0.0478 (0.0332)	-0.0534 (0.0332)	-0.0554* (0.0332)	-0.0512 (0.0332)	-0.0473 (0.0332)
2021	-0.2253*** (0.0357)	-0.1647*** (0.0367)	-0.1519*** (0.0369)	-0.1525*** (0.0369)	-0.1651*** (0.0367)	-0.1647*** (0.0367)	-0.1599*** (0.0369)	-0.1493*** (0.0369)
2022	-0.2370*** (0.0421)	-0.1943*** (0.0431)	-0.1812*** (0.0433)	-0.1814*** (0.0434)	-0.1949*** (0.0432)	-0.1939*** (0.0431)	-0.1895*** (0.0433)	-0.1725*** (0.0434)
2023	-0.1570** (0.0729)	-0.1484** (0.0745)	-0.1450* (0.0746)	-0.1399* (0.0746)	-0.1528** (0.0748)	-0.1594** (0.0746)	-0.1455* (0.0746)	-0.1443* (0.0747)
Negative price sentiment		-1.2807*** (0.0276)	-1.2788*** (0.0276)	-1.2784*** (0.0276)	-1.2810*** (0.0276)	-1.2796*** (0.0276)	-1.2802*** (0.0276)	-1.2789*** (0.0277)
Positive price sentiment		0.7302*** (0.0393)	0.7316*** (0.0393)	0.7318*** (0.0393)	0.7299*** (0.0393)	0.7285*** (0.0393)	0.7310*** (0.0393)	0.7307*** (0.0393)
Crowd's average rating			0.0836*** (0.0255)					
Crowd's share of 5-star ratings				0.2627*** (0.0831)				
Crowd's share of 4-star ratings					0.1104 (0.1545)			
Crowd's share of 2-star ratings						-0.4763*** (0.1398)		
Crowd's share of 1-star ratings							-0.1524 (0.1133)	
Crowd's positive imbalance								0.3144*** (0.0834)
Days of the week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loglikelihood	-18,387.0	-16,928.3	-16,923.0	-16,923.3	-16,928.1	-16,922.5	-16,927.4	-16,874.5
AIC	37,331.9	34,418.7	34,409.9	34,410.7	34,420.2	34,409.1	34,418.9	34,309.0
BIC	39,448.5	36,550.4	36,549.3	36,550.0	36,559.5	36,548.4	36,558.2	36,432.5
N	14,565	14,565	14,565	14,565	14,565	14,565	14,565	14,526

* Note: *p < .10; **p < .05; **p < .01. Sample period: 24 January 2017–30 August 2023. Days of the week dummies, month dummies, and store dummies are excluded for space. Perfectly predicted samples are dropped.

4.4 Marginal Effect Analysis

Table 15 displays the marginal effect results for the statistically significant and/or primary independent variables. The focus of the interpretation is specifically on how these variables affect the likelihood of 1-star and 5-star ratings.

The interpretation of the marginal effects for the review heterogeneity and price sentiment variables is based on Model 8, which shows the best model-fit among all models. Unverified multi-review reviewers (i.e., unverified NOT) are about 10.05% more likely to give a 1-star rating and 16.95% less likely to give a 5-star rating than verified multi-review reviewers (i.e., verified NOT), who serve as the baseline group. The only difference between these two groups is whether the reviewers are purchase verified. Therefore, these marginal effects suggest that purchase verification among multi-review reviewers reduces the likelihood of a 1-star rating and increases the likelihood of a 5-star rating. Notably, for multi-review reviewers, the negative marginal effect of purchase unverification on a 5-star rating exceeds the positive marginal effect on a 1-star rating.

Verified and unverified OTNOTA are 2.7% and 8.5% more likely to give a 1-star rating and 5.2% and 14.7% less likely to give a 5-star rating, respectively, compared to verified multi-review reviewers. In general, OTNOTA shows a higher probability of giving a 1-star rating and a lower probability of giving a 5-star rating than verified multi-review reviewers. The magnitudes of both positive and negative marginal effects for unverified OTNOTA on the likelihood of giving a 1-star or 5-star rating are larger than those for verified OTNOTA. Verified and unverified OTAs are 3.4% and 3.9% more likely to receive a 1-star rating and 6.5% and 7.5% less likely to receive a 5-star rating, respectively, than verified multi-review reviewers (i.e., verified NOT). OTA reviewers are more likely to give a 1-star rating and less likely to give a 5-star rating than verified multi-review reviewers. The gap in the probability of giving a 5-star rating is larger between OTA and verified NOT than the gap in the likelihood

of giving a 1-star rating. The impact of purchase verification on the difference in the probabilities of giving 1-star or 5-star ratings between OTA and verified NOT is smaller than between OTNOTA and verified NOT.

Increasing the length of the review title and body raises the likelihood of a 1-star rating by 0.082% and 0.004%, respectively. It lowers the chances of giving a 5-star rating by 0.147% and 0.007%, respectively. Longer titles and bodies make it more likely to assign a 1-star rating and less likely to give a 5-star rating. The impact of an extra character in the title is greater than that in the body. The effect of an additional character is stronger for a 5-star rating than for a 1-star rating. This indicates that dissatisfied consumers tend to write longer titles and bodies, especially titles, compared to satisfied consumers.

Negative price sentiment increases the likelihood of a 1-star rating by 36.54% and decreases the likelihood of a 5-star rating by 41.74% compared to neutral price sentiment (the baseline). In contrast, positive price sentiment reduces the probability of a 1-star review by 7.55% and raises the chance of a 5-star review by 23.33%. Both negative and positive price sentiment have a greater impact on 5-star ratings than on 1-star ratings. Moreover, the effect of negative price sentiment is larger than that of positive price sentiment in both 1-star and 5-star ratings. Accordingly, sellers should focus more on negative price sentiment than on positive. These findings are reasonable because consumers are more likely to give a 1-star or 5-star rating when they are unhappy or happy with the product price. Therefore, potential measurement errors in price sentiment variables created by combining generative LLMs with manual annotation may be small or negligible.

Table 15. Marginal effect of reviewer heterogeneity, length of review, and price sentiment (Appendix)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Unverified NOT 1	0.112617**	0.102432**	0.100869**	0.101541**	0.102097**	0.102754**	0.101700**	0.100520**
	(0.048671)	(0.041193)	(0.041104)	(0.041159)	(0.041151)	(0.041198)	(0.041164)	(0.040998)
2	0.033364***	0.032979***	0.032510***	0.032694***	0.032893***	0.033113***	0.032753***	0.032595***
	(0.010627)	(0.010498)	(0.010530)	(0.010515)	(0.010504)	(0.010502)	(0.010514)	(0.010579)
3	0.022723***	0.025948***	0.025602***	0.025715***	0.025898***	0.026060***	0.025784***	0.025737***
	(0.005493)	(0.006540)	(0.006595)	(0.006562)	(0.006556)	(0.006540)	(0.006567)	(0.006654)
4	0.007708***	0.010636***	0.010515***	0.010523***	0.010640***	0.010709***	0.010577***	0.010676***
	(0.001958)	(0.001845)	(0.001840)	(0.001838)	(0.001846)	(0.001850)	(0.001843)	(0.001861)
5	-0.176411***	-0.171995***	-0.169496***	-0.170473***	-0.171529***	-0.172637***	-0.170815***	-0.169528***
	(0.063790)	(0.058442)	(0.058510)	(0.058475)	(0.058445)	(0.058449)	(0.058489)	(0.058573)
Verified OTNOTA 1	0.032105***	0.027145***	0.026577***	0.026619***	0.027163***	0.027365***	0.026886***	0.027072***
	(0.009068)	(0.007769)	(0.007791)	(0.007788)	(0.007769)	(0.007757)	(0.007782)	(0.007771)
2	0.011790***	0.010364***	0.010134***	0.010155***	0.010369***	0.010458***	0.010257***	0.010366***
	(0.003618)	(0.003144)	(0.003144)	(0.003145)	(0.003143)	(0.003143)	(0.003145)	(0.003153)
3	0.009402***	0.009392***	0.009177***	0.009196***	0.009396***	0.009480***	0.009293***	0.009393***
	(0.003054)	(0.002978)	(0.002974)	(0.002975)	(0.002978)	(0.002980)	(0.002977)	(0.002987)
4	0.004762***	0.005311***	0.005182***	0.005191***	0.005315***	0.005371***	0.005251***	0.005322***
	(0.001725)	(0.001835)	(0.001827)	(0.001827)	(0.001835)	(0.001840)	(0.001832)	(0.001843)
5	-0.058060***	-0.052211***	-0.051070***	-0.051160***	-0.052244***	-0.052674***	-0.051686***	-0.052152***
	(0.017422)	(0.015690)	(0.015701)	(0.015700)	(0.015689)	(0.015682)	(0.015700)	(0.015717)
Unverified OTNOTA 1	0.107895***	0.085562***	0.084335***	0.084759***	0.085495***	0.085928***	0.084897***	0.084909***
	(0.012379)	(0.010445)	(0.010454)	(0.010455)	(0.010444)	(0.010437)	(0.010458)	(0.010443)
2	0.032364***	0.028621***	0.028218***	0.028349***	0.028600***	0.028766***	0.028403***	0.028524***
	(0.004103)	(0.003717)	(0.003719)	(0.003719)	(0.003717)	(0.003718)	(0.003720)	(0.003731)
3	0.022255***	0.023264***	0.022944***	0.023032***	0.023250***	0.023385***	0.023096***	0.023209***
	(0.003180)	(0.003230)	(0.003229)	(0.003228)	(0.003230)	(0.003231)	(0.003232)	(0.003242)
4	0.007787***	0.010378***	0.010238***	0.010256***	0.010380***	0.010452***	0.010308***	0.010403***
	(0.001724)	(0.001846)	(0.001839)	(0.001839)	(0.001847)	(0.001851)	(0.001843)	(0.001855)
5	-0.170301***	-0.147824***	-0.145734***	-0.146396***	-0.147725***	-0.148532***	-0.146705***	-0.147044***
	(0.020457)	(0.018644)	(0.018661)	(0.018655)	(0.018643)	(0.018637)	(0.018667)	(0.018683)
Verified OTA 1	0.038387**	0.034303**	0.033593**	0.033776**	0.034290**	0.034497**	0.033942**	0.034109**
	(0.017902)	(0.015320)	(0.015318)	(0.015321)	(0.015320)	(0.015306)	(0.015322)	(0.015293)
2	0.013856**	0.012885**	0.012606**	0.012677**	0.012879**	0.012972**	0.012742**	0.012853**
	(0.006164)	(0.005552)	(0.005551)	(0.005552)	(0.005551)	(0.005551)	(0.005552)	(0.005562)
3	0.010917**	0.011524**	0.011269**	0.011329**	0.011519**	0.011605**	0.011396**	0.011497**
	(0.004724)	(0.004836)	(0.004836)	(0.004835)	(0.004836)	(0.004837)	(0.004838)	(0.004847)
4	0.005395**	0.006351**	0.006204**	0.006232**	0.006351**	0.006409**	0.006277**	0.006351**
	(0.002253)	(0.002566)	(0.002564)	(0.002561)	(0.002567)	(0.002571)	(0.002566)	(0.002578)
5	-0.068555**	-0.065063**	-0.063672**	-0.064014**	-0.065039**	-0.065482**	-0.064358**	-0.064810**
	(0.030859)	(0.028150)	(0.028150)	(0.028148)	(0.028150)	(0.028139)	(0.028157)	(0.028157)
Unverified OTA 1	0.034441	0.039114	0.039727*	0.040496*	0.038767	0.039598*	0.039035	0.039905*
	(0.026695)	(0.023879)	(0.024007)	(0.024073)	(0.023851)	(0.023893)	(0.023905)	(0.023938)
2	0.012567	0.014532*	0.014702*	0.014969*	0.014413*	0.014718*	0.014486*	0.014840*
	(0.009053)	(0.008268)	(0.008269)	(0.008270)	(0.008269)	(0.008269)	(0.008268)	(0.008284)
3	0.009977	0.012882*	0.012995*	0.013213*	0.012785*	0.013045*	0.012834*	0.013134*
	(0.006823)	(0.006912)	(0.006885)	(0.006867)	(0.006922)	(0.006907)	(0.006909)	(0.006905)
4	0.005008	0.006975**	0.006993**	0.007088**	0.006934**	0.007070**	0.006938**	0.007101**
	(0.003095)	(0.003338)	(0.003294)	(0.003267)	(0.003352)	(0.003334)	(0.003331)	(0.003321)
5	-0.061991	-0.073503*	-0.074417*	-0.075767*	-0.072899*	-0.074431*	-0.073292*	-0.074981*
	(0.045540)	(0.042261)	(0.042315)	(0.042331)	(0.042260)	(0.042263)	(0.042278)	(0.042306)
Length of review title 1	0.001016***	0.000821***	0.000818***	0.000819***	0.000820***	0.000812***	0.000821***	0.000819***
	(0.000131)	(0.000111)	(0.000111)	(0.000110)	(0.000111)	(0.000111)	(0.000111)	(0.000110)
2	0.000325***	0.000285***	0.000284***	0.000285***	0.000285***	0.000283***	0.000285***	0.000286***
	(0.000042)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)
3	0.000236***	0.000242***	0.000241***	0.000241***	0.000241***	0.000239***	0.000242***	0.000242***
	(0.000031)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)
4	0.000097***	0.000118***	0.000118***	0.000118***	0.000118***	0.000117***	0.000118***	0.000119***
	(0.000013)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000017)
5	-0.001673***	-0.001466***	-0.001461***	-0.001463***	-0.001465***	-0.001451***	-0.001466***	-0.001465***
	(0.000215)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)
Length of review body 1	0.000069***	0.000040***	0.000040***	0.000040***	0.000040***	0.000039***	0.000040***	0.000040***
	(0.000007)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)
2	0.000022***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***
	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)
3	0.000016***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***
	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)
4	0.000007***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***
	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)
5	-0.000113***	-0.000071***	-0.000071***	-0.000071***	-0.000071***	-0.000070***	-0.000071***	-0.000071***
	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)
Negative price sentiment 1		0.366525***	0.365663***	0.365563***	0.366611***	0.365891***	0.366342***	0.365410***
		(0.009341)	(0.009340)	(0.009341)	(0.009341)	(0.009339)	(0.009341)	(0.009354)
2		0.067186***	0.067200***	0.067204***	0.067185***	0.067242***	0.067179***	0.067386***
		(0.002266)	(0.002265)	(0.002265)	(0.002266)	(0.002266)	(0.002266)	(0.002272)
3		0.015111***	0.015233***	0.015211***	0.015112***	0.015218***	0.015144***	0.015323***
		(0.001872)	(0.001868)	(0.001868)	(0.001872)	(0.001869)	(0.001871)	(0.001871)
4		-0.030873***	-0.030734***	-0.030742***	-0.030873***	-0.030745***	-0.030843***	-0.030709***
		(0.002086)	(0.002084)	(0.002083)	(0.002086)	(0.002084)	(0.002086)	(0.002089)
5		-0.417949***	-0.417363***	-0.417235***	-0.418036***	-0.417605***	-0.417822***	-0.417410***
		(0.007119)	(0.007128)	(0.007131)	(0.007119)	(0.007125)	(0.007121)	(0.007139)
Positive price sentiment 1		-0.075611***	-0.075717***	-0.075738***	-0.075587***	-0.075505***	-0.075667***	-0.075527***
		(0.003017)	(0.003017)	(0.003017)	(0.003018)	(0.003021)	(0.003017)	(0.003015)
2		-0.053100***	-0.053151***	-0.053194***	-0.053073***	-0.052997***	-0.053128***	-0.053150***
		(0.002606)	(0.002605)	(0.002606)	(0.002606)	(0.002608)	(0.002605)	(0.002609)
3		-0.059780***	-0.059846***	-0.059884***	-0.059747***	-0.059607***	-0.059832***	-0.059799***
		(0.003298)	(0.003297)	(0.003298)	(0.003298)	(0.003298)	(0.003298)	(0.003299)
4		-0.044778***	-0.044841***	-0.044843***	-0.044760***	-0.044630***	-0.044833***	-0.044824***
		(0.002966)	(0.002965)	(0.002966)	(0.002966)	(0.002964)	(0.002967)	(0.002968)
5		0.233268***	0.233555***	0.233659***	0.233166***	0.232739***	0.233460***	0.233299***
		(0.010529)	(0.010522)	(0.010523)	(0.010530)	(0.010539)	(0.010525)	(0.010529)
N	14,565	14,565	14,565	14,565	14,565	14,565	14,565	14,526

* Note: marginal effects of other variables are not displayed due to the size of the table.

Tables 16 and 17 present the predicted probabilities of 1- and 5-star ratings for reviewer types with different price sentiments, based on the marginal effect analysis in Model 8. The likelihood of a 1-star rating is highest for negative price sentiment, followed by neutral and positive sentiment, respectively. These results make sense economically because dissatisfied reviewers are more likely to give a 1-star rating than indifferent or satisfied reviewers.

When price sentiment is negative, the predicted probability of a 1-star rating is highest for unverified NOT, followed by unverified OTNOTA, unverified OTA, verified OTA, verified OTNOTA, and verified NOT, respectively. Additionally, unverified reviewers with negative price sentiment are more likely to give a 1-star rating than unverified reviewers overall, regardless of reviewer type.

The predicted probability of a 5-star rating is highest with positive price sentiment, followed by neutral and negative sentiment, across all reviewer types. Economically, this makes sense because satisfied reviewers are more likely to give a 5-star rating than indifferent or unsatisfied reviewers. Specifically, when price sentiment is positive, the predicted probability of a 5-star rating is highest for verified NOT, followed by OTNOTA and OTA. Verified reviewers with positive price sentiment are more likely to give a 5-star rating than unverified reviewers, regardless of reviewer type.

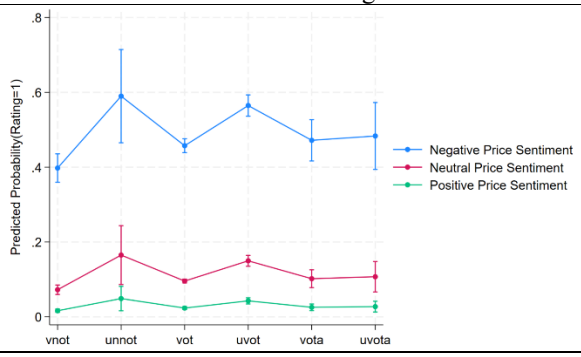
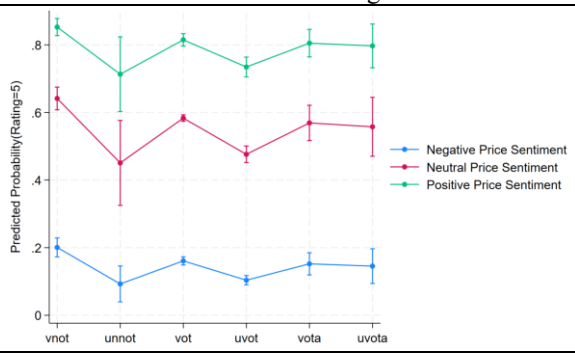
Overall, the variations in the predicted probabilities of 1- and 5-star ratings across reviewer types, price sentiments, and purchase verification suggest that consumer preferences should be carefully analyzed by reviewer segmentation and price sentiment, as each reviewer group may have different rating tendencies. Additionally, when price information is missing in online product reviews, consumer review text data can serve as an alternative source for gaining insights into latent consumer and product information.

Table 16. Predicted probabilities of star ratings by reviewer types with different price sentiment

Variables	The predicted probability of a 1-star rating	The predicted probability of a 5-star rating
Verified NOT with negative price sentiment	0.397666***	0.200746***
	(0.019346)	(0.014224)
Verified NOT with neutral price sentiment	0.072459***	0.641470***
	(0.006440)	(0.017147)
Verified NOT with positive price sentiment	0.016356***	0.852640***
	(0.002315)	(0.012923)
Unverified NOT with negative price sentiment	0.589735***	0.092675***
	(0.063552)	(0.027127)
Unverified NOT with neutral price sentiment	0.165075***	0.450725***
	(0.040128)	(0.064135)
Unverified NOT with positive price sentiment	0.048829***	0.713128***
	(0.016622)	(0.056344)
Verified OTNOTA with negative price sentiment	0.457303***	0.160849***
	(0.009447)	(0.006162)
Verified OTNOTA with neutral price sentiment	0.095690***	0.583265***
	(0.002705)	(0.004936)
Verified OTNOTA with positive price sentiment	0.023509***	0.814896***
	(0.002057)	(0.009395)
Unverified OTNOTA with negative price sentiment	0.564648***	0.103744***
	(0.014508)	(0.006918)
Unverified OTNOTA with neutral price sentiment	0.149703***	0.476186***
	(0.007411)	(0.012477)
Unverified OTNOTA with positive price sentiment	0.042739***	0.734531***
	(0.004199)	(0.014909)
Verified OTA with negative price sentiment	0.471730***	0.152155***
	(0.028158)	(0.016755)
Verified OTA with neutral price sentiment	0.101972***	0.569047***
	(0.012124)	(0.026627)
Verified OTA with positive price sentiment	0.025564***	0.805067***
	(0.004499)	(0.020789)
Unverified OTA with negative price sentiment	0.483303***	0.145427***
	(0.045626)	(0.026150)
Unverified OTA with neutral price sentiment	0.107213***	0.557603***
	(0.020766)	(0.044451)
Unverified OTA with positive price sentiment	0.027317***	0.796976***
	(0.007311)	(0.033025)

* Note: * p<.1; ** p<.05; *** p<.01.

Table 17. Marginal Plot for 1-star and 5-star ratings

The predicted probability of a 1-star rating	The predicted probability of a 5-star rating
	

* Note: vnot (verified NOT), unnot (unverified NOT), vot (verified OTNOTA), uvot (unverified OTNOTA), vota (verified OTA), and uvota (unverified OTA).

Table 18 presents the marginal effect of pre-purchase information from the previous reviews on the probability of giving a star rating. Specifically, when a reviewer '*i*' gives a rating for the subscription box '*j*' at time '*t*', the reviewer is influenced by previous reviews written by other reviewers (i.e., crowd). In detail, a one-star increase in the average star rating from previous crowd reviews for the product decreases the probability of giving a 1-star rating by 1.53%. It increases the probability of giving a 5-star rating by 2.73%. Importantly, the crowd's average star rating for the product may reflect the crowd's aggregate evaluation of its quality. The marginal effect of the crowd's average star rating on an individual star rating can be further decomposed into the effects of the shares of different ratings in the previous crowd reviews. Specifically, a 1% increase in the crowd's share of 5-star ratings in previous reviews decreases the probability of giving a 1-star rating by 0.0479% and increases the probability of giving a 5-star rating by 0.0856%, respectively. In addition, a 1% increase in the share of 2-star ratings in the previous crowd reviews increases the probability of giving a 1-star rating by 0.087%. It decreases the probability of giving a 5-star rating by 0.155%. Moreover, a 1% increase in the share of 5- and 4-star ratings (i.e., positive imbalance) in previous crowd reviews decreases the probability of giving a 1-star rating by 0.057% and increases the probability of giving a 5-star rating by 0.102%, respectively. The magnitude of the marginal effect of the share of 2-star ratings in previous reviews on the probability of giving extreme ratings is larger than that for both the share of 5-star ratings and the positive imbalance in previous reviews. Therefore, reviewers are more sensitive to the crowd's share of 2-star ratings than the crowd's positive imbalance.

These results indicate that consumers respond to prior crowd ratings when assigning star ratings; therefore, improving the quality of pre-purchase information in reviews may enhance consumer purchase decisions while reducing their uncertainty about product quality.

Table 18. Marginal effect of previous other reviewer's rating distributions (Appendix)

Variable	Model 3	Model 4	Model 6	Model 8
Crowd's average rating 1	-0.015256***			
	(0.004649)			
2	-0.005307***			
	(0.001621)			
3	-0.004494***			
	(0.001373)			
4	-0.002199***			
	(0.000676)			
5	0.027255***			
	(0.008298)			
Crowd's share of 5-star ratings 1		-0.047927***		
		(0.015166)		
2		-0.016676***		
		(0.005289)		
3		-0.014117***		
		(0.004478)		
4		-0.006902***		
		(0.002200)		
5		0.085623***		
		(0.027073)		
Crowd's share of 2-star ratings 1			0.086900***	
			(0.025511)	
2			0.030237***	
			(0.008901)	
3			0.025588***	
			(0.007532)	
4			0.012522***	
			(0.003712)	
5			-0.155246***	
			(0.045539)	
Crowd's positive imbalance 1				-0.057286***
				(0.015200)
2				-0.019984***
				(0.005317)
3				-0.016916***
				(0.004503)
4				-0.008291***
				(0.002226)
5				0.102477***
				(0.027161)
N	14,565	14,565	14,565	14,526

* Note: marginal effects of other variables are not displayed due to the size of the table.

5. Conclusion

This study investigates the existence, prevalence, and negative rating tendencies of one-time reviewers compared to multi-review reviewers across all product categories over several decades, analyzing 571,544,746 individual reviews from Amazon.com. One possible explanation for the negative ratings from one-time reviewers is negative self-selection, where dissatisfied consumers give their first and only review a 1-star rating in a specific category or across all categories.

Sellers and consumers may focus on specific products related to their business or

purchase, rather than all products across every category. Additionally, performing empirical analysis across all reviews and categories demands significant computational resources, budgets, and time. Therefore, the study focuses on the Subscription Boxes category for further econometric analysis. Subscribers regularly receive a curated collection of products that is updated each subscription period. Consumers‘ uncertainty about the quality of the subscription box is high; thus, the value of pre-purchase information from previous reviews can be significant. Moreover, promotional reviewers might have less incentive to write promotional reviews in this category because the cost of posting a promotional review after subscribing may be higher than for a one-time purchase, and the collection is often updated with new products. Additionally, this category shows the largest gap in the share of extreme ratings between OTA and multi-review reviewers. Overall, the first review from each unique reviewer in the Subscription Boxes category is used for further econometric analysis to ensure a fair comparison between one-time and multi-review reviewers.

Amazon’s star rating is an ordinal response that reflects the strength of consumer preference for a reviewed product; therefore, an ordered probit model is used to estimate the effects of reviewer heterogeneity, price sentiment, and pre-purchase information on star ratings. Since numerical price data are unavailable for the Subscription Boxes category, this study develops a discrete price sentiment variable from review text using generative LLMs and manual annotation. Additionally, pre-purchase information is extracted from previous reviews through data mining.

Marginal effect analysis is used for interpretation because the ordered probit model is nonlinear. The analysis shows heterogeneity in the marginal effects of reviewer type on the likelihood of giving 1- or 5-star ratings. Specifically, when price sentiment is negative, unverified NOT has the highest predicted probability of a 1-star rating, followed by unverified OTNOTA and OTA. Conversely, when price sentiment is positive, verified NOT (the baseline

reviewers) has the highest predicted probability of a 5-star rating, followed by verified OTNOTA and OTA. Additionally, purchase-verified reviewers are more likely to give a 5-star rating than unverified reviewers. In contrast, purchase-unverified reviewers are more likely to give a 1-star rating, regardless of reviewer type. Furthermore, positive price sentiment increases the probability of a 5-star rating, while negative price sentiment increases the likelihood of a 1-star rating.

Further marginal-effect analysis results indicate that consumers may respond to pre-purchase information in prior reviews. Specifically, an increase in the average star rating, the share of 5-star ratings, and the positive imbalance in previous reviews raises the likelihood of giving a 5-star rating. Conversely, a higher share of 2-star ratings in prior reviews increases the chances of giving a 1-star rating.

When a consumer writes a longer review title and body, the likelihood of giving a 1-star rating increases, while the chance of giving a 5-star rating decreases. The effect is stronger for the title than for the body. These results suggest that unhappy consumers are more likely to write longer review titles and bodies, especially the title.

These findings indicate that sellers, consumers, and regulators should differentiate between one-time reviewers and multi-review reviewers based on their diversity. The consumer segmentation framework introduced in this study can help categorize one-time reviewers into distinct subgroups. Furthermore, the existence, frequency, and behaviors of one-time reviewers should be compared with those of multi-review reviewers to ensure a fair assessment. Although all groups show overall increases in the number and negative polarity of one-time reviewers, the extent of the patterns varies across groups. Consequently, one-time reviewers should be analyzed both within and across categories to uncover diverse review behaviors.

The findings in this paper also show that purchase-verified reviewers are more likely to give a 5-star rating. In contrast, unverified reviewers tend to give a 1-star rating, regardless of reviewer type. Even within the same one-time reviewer group, rating patterns differ by verification status, and the magnitude of these differences varies by reviewer type. This indicates that reviewer heterogeneity should be carefully examined, given verification status, as rating behavior can differ by verification status. Overall, when market participants make decisions about atypical reviewers, such as one-time reviewers, these choices should be adapted to different consumer segments and reviewer differences.

The study also shows how to extract pre-purchase information from previous reviews using data mining and how to identify missing business details, such as price, from unstructured user-generated content with generative LLMs combined with manual annotation. Additionally, this paper demonstrates how to enhance the performance of generative LLMs for specific tasks through fine-tuning with manual annotation. These findings and methods can help reduce asymmetric information in Amazon's recommendation system by uncovering hidden insights from user-generated content. Importantly, these computational techniques can also benefit sellers, buyers, and regulators by narrowing the information gap between Amazon and other market players. Amazon acts not only as a marketplace for sellers but also as a seller itself on its retail platform. Amazon has access to consumers‘ and sellers‘ data, while other market players cannot directly access this internal information for their business decisions. Therefore, other market players can close this information gap by leveraging publicly available user-generated content. Specifically, they can extract valuable business insights from structured and unstructured content generated by users, sellers, and the platform on Amazon.com to make data-driven decisions, using data mining, generative LLMs, manual annotation, and econometric analysis as outlined in this study. Additionally, regulators can apply these multimodal methods to reduce potential biases and harmful content in user-generated material.

Overall, the hybrid multimodal approach outlined in this paper is applicable, adaptable, and scalable across various product categories on different digital retail platforms.

However, price sentiment variables generated by LLMs may be subject to measurement error, hallucinations, and biases. To address these issues, several methods discussed in this study are useful, such as prompt engineering, fine-tuning, manual annotation for specific tasks, and custom Python functions for filtering out abnormal outputs from generative LLMs. Ultimately, the predictive performance of the fine-tuned GPT models with manual annotation is relatively high. The empirical results are consistent across models, and the interpretation of marginal effects is also economically feasible. Therefore, the approach in this study can help mitigate concerns related to LLMs.

This study relies on a descriptive, associational empirical approach rather than a causal inference framework, as it lacks access to individual demographics and relevant business information for each reviewer. Such access would require researchers to obtain internal data directly from the platform or to collaborate with the platform itself. Typically, platforms do not share internal data such as quantity information (Waldfogel, 2024). For example, assigning randomized treatments to individual consumers—who may be one-time or multi-review reviewers—requires extensive internal data to track large numbers of individual reviewers across all categories over time. Since the share of one-time reviewers is relatively small compared to multi-review reviewers, and because one-time reviewers may later become multi-review reviewers by posting additional reviews within or across different categories, this is problematic. Design-based quasi-experiments are also limited because researchers cannot access individuals' demographic data to construct personal counterfactual outcomes. Furthermore, internal instrument approaches, such as dynamic panel models with lagged dependent variables, are limited because this study uses only the first review of each reviewer. External instrument variable methods for linear models and control function approaches for

nonlinear models face similar constraints, as both require individual demographic information to identify individual-level instrument variables.

Notably, the findings and approaches in this study can be valuable for future research on one-time reviewers using causal inference, as they demonstrate the existence, prevalence, abnormal rating behaviors, and heterogeneity of one-time reviewers across all product categories over time, compared to multi-review reviewers. Therefore, future studies can use these findings and approaches to formulate hypotheses, treatment variables, instrument variables, and design large-scale field experiments or design-based quasi-experiments, while reducing computational resources, budgets, and time constraints. For example, before conducting field experiments to understand why some one-time reviewers remain as single-time reviewers after their first and only review, rather than writing additional reviews, researchers can analyze user-generated content to uncover insights for causal inference from observational data.

Acknowledgments

The paper will be presented at the Business Implications of Generative AI@MIT on April 2 and 3, 2026. English editing of this article was completed using editorworld.com. Additionally, OpenAI's ChatGPT, Microsoft's Copilot, and Grammarly are used to polish and correct grammar in the manually written draft to improve the quality of the manuscript.

Declarations

Author Contributions

The author declares sole responsibility for the study concept, design, material preparation, data collection, analysis, writing, and manuscript preparation. The author read and approved the final manuscript.

Funding

Partial funding for the OpenAI API is supported by the Missouri S&T OVCRI Seed Funding for Arts, Humanities, Social, and Behavioral Science Research. In this paper, cloud computing is supported by the Google Cloud Research Credits program with the award GCP19980904.

Data Availability

Data is available from https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023; see Sect. 2.

Code Availability

Manual code is written in Python and Stata, see Sections 2, 3, and 4.

Conflict of interest

The author declares no conflict of interest.

Reference

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

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

Bayerl, A., Schoenmueller, V., Goldenberg, J., & Stahl, F. EXPRESS: The Weekend Effect in Online Reviews. Journal of Marketing Research, 00222437251391808.

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

Bybee, J. L. (2023). The ghost in the machine: Generating beliefs with large language models. *arXiv preprint arXiv:2305.02823*, 349-389.

Chava, S., Du, W., Shah, A., & Zeng, L. (2022). *Measuring firm-level inflation exposure: A deep learning approach*. SSRN.

Chen, J. H., and Tsai, W. C. (2012). A comparison of international income inequality: An ordered probit model analysis. *Applied Economics*, *44*(13), 1701-1716.

Chen, Y. (2018). User-generated physician ratings—Evidence from Yelp. URL: https://www.softwareadvice.com/resources/how-patientsuse.

Chevalier, J., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and BarnesandNoble. com. Quantitative marketing and Economics, 1, 203-222.

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

Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. International Journal of Electronic Commerce, 17(1), 39-58.

Dai, W., Jin, G., Lee, J., & Luca, M. (2018). Aggregation of consumer ratings: an application to yelp. com. Quantitative Marketing and Economics, 16, 289-339.

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. Journal of Consumer Research, 42(6), 817-833.

Fang, L., Wu, C., & Sun, B. (2024). Shifting Standards or Changing Preferences? Unraveling Review Polarization via LLMs. *Unraveling Review Polarization via LLMs (October 29, 2024)*.

Farrell, M. H., Liang, T., & Misra, S. (2021). Deep neural networks for estimation and inference. Econometrica, 89(1), 181-213.

Greene, William H. (2012), *Econometric Analysis*, 7th ed, Harlow: Pearson Education.

Han, S. (2024). Mining causality: AI-assisted search for instrumental variables. arXiv preprint arXiv:2409.14202.

Hausman, J. A., Lo, A. W., and MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, *31*(3), 319-379.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. Marketing Science, 41(5), 896-921.

Hou, Y., Li, J., He, Z., Yan, A., Chen, X., & McAuley, J. (2024). Bridging language and items for retrieval and recommendation. arXiv preprint arXiv:2403.03952.

Hu, N., Pavlou, P. A., & Zhang, J. (2006, June). Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In Proceedings of the 7th ACM conference on Electronic commerce (pp. 324-330).

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On self-selection biases in online product reviews. MIS Quarterly, 41(2), 449-475.

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

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

Kawakatsu, H., and Largey, A. G. (2009). EM algorithms for ordered probit models with endogenous regressors. *The Econometrics Journal*, *12*(1), 164-186.

Kheiri, K., & Karimi, H. (2024, December). SentimentGPT: Leveraging GPT for Advancing Sentiment Analysis. In 2024 IEEE International Conference on Big Data (BigData) (pp. 7051-7060). IEEE.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. Journal of Economic Literature, 61(4), 1281-1317.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. Information Systems Research, 19(4), 456-474.

Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. Journal of Marketing Research, 56(6), 918-943.

Ludwig, J., Mullainathan, S., & Rambachan, A. (2025). Large language models: An applied econometric framework (No. w33344). National Bureau of Economic Research.

Luca, M., & 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., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. American Economic Review, 104(8), 2421-2455.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*(2), 109-127.

OpenAI. (2025) Compare models. Retrieved November 29, 2025. From https://platform.openai.com/docs/models/compare

Ramsey, S. M., & Bergtold, J. S. (2021). Examining inferences from neural network estimators of binary choice processes: Marginal effects, and willingness-to-pay. Computational Economics, 58(4), 1137-1165.

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

Reimers, I., Riedl, C., and & Waldfogel, J. (2025). Information and the Welfare Benefits from Differentiated Products (No.) (No. w33401). National Bureau of Economic Research.

Riyanto, S., Imas, S. S., Djatna, T., & Atikah, T. D. (2023). Comparative analysis using various performance metrics in imbalanced data for multi-class text classification. *International Journal of Advanced Computer Science and Applications*, *14*(6).

Safara, F. (2022). A computational model to predict consumer behaviour during COVID-19 pandemic. Computational Economics, 59(4), 1525-1538.

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. Journal of Marketing Research, 57(5), 853-877.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. Annual Review of Economics, 8(1), 321-340.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. Marketing Science, 38(1), 1-20.

Timoshenko, A., Mao, C., & Hauser, J. R. (2025). Can Large Language Models Extract Customer Needs as well as Professional Analysts?. arXiv preprint arXiv:2503.01870.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Pudney, S., and Shields, M. (2000). Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics*, *15*(4), 367-399.

Waldfogel, J. (2024). Amazon self-preferencing in the shadow of the digital markets act (No. w32299). National Bureau of Economic Research.

Winship, C., and Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review*, 512-525.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223, 1(2).

Appendix

Table A. 1 Rating distribution of NOT reviews over years

Year	NOT										
	Rating 1		Rating 2		Rating 3		Rating 4		Rating 5		Total
	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	
1996	0	0.00%	0	0.00%	1	4.17%	4	16.67%	19	79.17%	24
1997	148	3.42%	226	5.23%	303	7.01%	943	21.81%	2,703	62.53%	4,323
1998	1,147	4.29%	1,226	4.59%	2,063	7.72%	5,607	20.99%	16,673	62.41%	26,716
1999	3,713	5.32%	3,563	5.10%	6,073	8.70%	13,848	19.84%	42,606	61.04%	69,803
2000	11,562	4.33%	13,230	4.95%	26,037	9.75%	63,948	23.94%	152,332	57.03%	267,109
2001	13,292	4.77%	14,721	5.28%	28,672	10.28%	67,612	24.24%	154,646	55.44%	278,943
2002	15,636	5.24%	16,508	5.53%	30,559	10.24%	70,610	23.66%	165,152	55.33%	298,465
2003	19,309	6.04%	18,417	5.76%	33,100	10.36%	73,709	23.06%	175,102	54.78%	319,637
2004	30,205	7.70%	23,562	6.01%	39,998	10.19%	86,667	22.09%	211,936	54.01%	392,368
2005	44,209	8.18%	33,197	6.14%	55,435	10.25%	116,617	21.57%	291,305	53.87%	540,763
2006	46,253	7.82%	35,391	5.98%	57,918	9.79%	127,124	21.49%	324,944	54.92%	591,630
2007	59,819	6.64%	48,443	5.38%	82,602	9.17%	191,928	21.30%	518,356	57.52%	901,148
2008	75,761	7.32%	58,972	5.70%	95,499	9.23%	219,565	21.22%	584,984	56.53%	1,034,781
2009	102,083	7.81%	74,769	5.72%	121,571	9.30%	270,126	20.66%	738,970	56.52%	1,307,519
2010	154,958	8.37%	108,935	5.89%	171,903	9.29%	376,222	20.33%	1,038,494	56.12%	1,850,512
2011	265,769	8.77%	177,547	5.86%	276,571	9.12%	594,705	19.62%	1,716,773	56.63%	3,031,365
2012	463,823	7.82%	309,270	5.21%	514,890	8.68%	1,136,901	19.16%	3,508,642	59.13%	5,933,526
2013	918,337	5.96%	718,168	4.66%	1,365,977	8.87%	2,996,978	19.46%	9,404,051	61.05%	15,403,511
2014	1,517,499	6.13%	1,099,734	4.44%	2,043,115	8.26%	4,378,955	17.70%	15,705,776	63.47%	24,745,079
2015	2,224,225	6.48%	1,496,591	4.36%	2,659,440	7.75%	5,557,053	16.19%	22,392,136	65.23%	34,329,445
2016	2,679,146	6.71%	1,742,960	4.37%	2,981,833	7.47%	6,203,673	15.54%	26,321,150	65.92%	39,928,762
2017	2,947,609	7.72%	1,774,290	4.65%	2,782,045	7.28%	5,447,170	14.26%	25,245,064	66.09%	38,196,178
2018	3,484,909	8.70%	1,893,131	4.73%	2,802,789	7.00%	5,176,181	12.92%	26,693,869	66.65%	40,050,879
2019	4,189,385	8.50%	2,029,282	4.12%	2,993,648	6.08%	5,477,010	11.12%	34,576,242	70.18%	49,265,567
2020	5,101,472	9.81%	2,424,757	4.66%	3,434,667	6.61%	5,602,512	10.78%	35,425,383	68.14%	51,988,791
2021	6,040,510	11.31%	2,776,504	5.20%	3,827,206	7.17%	5,601,231	10.49%	35,163,223	65.84%	53,408,674
2022	5,436,723	11.56%	2,517,920	5.36%	3,551,059	7.55%	5,242,981	11.15%	30,269,448	64.38%	47,018,131
2023	1,840,891	10.31%	883,327	4.95%	1,292,090	7.24%	2,124,394	11.90%	11,706,438	65.59%	17,847,140
Total	37,688,393	8.78%	20,294,641	4.73%	31,277,064	7.29%	57,224,274	13.34%	282,546,417	65.86%	429,030,789

* Note: Total reviews 429,030,789 from 3182,7161 unique reviewers for 42,472,318 unique products between 17 June 1996 and 14 September 2023.

Table A. 2 Rating distribution of OTA reviews over years

Year	OTA										
	Rating 1		Rating 2		Rating 3		Rating 4		Rating 5		
	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	# of Reviews	% in the year	Total
1996	0	NA	0	NA	0	NA	0	NA	0	NA	0
1997	22	2.45%	21	2.34%	30	3.34%	123	13.71%	701	78.15%	897
1998	157	3.22%	137	2.81%	185	3.80%	692	14.20%	3,702	75.97%	4,873
1999	547	4.65%	342	2.91%	522	4.43%	1,586	13.47%	8,775	74.54%	11,772
2000	1,400	5.59%	902	3.60%	1,387	5.54%	4,075	16.29%	17,259	68.97%	25,023
2001	1,350	6.61%	815	3.99%	1,191	5.83%	3,227	15.79%	13,856	67.79%	20,439
2002	1,512	7.38%	887	4.33%	1,130	5.52%	3,181	15.53%	13,770	67.24%	20,480
2003	2,012	9.20%	967	4.42%	1,271	5.81%	3,301	15.09%	14,320	65.47%	21,871
2004	2,576	11.13%	1,155	4.99%	1,453	6.28%	3,545	15.31%	14,419	62.29%	23,148
2005	3,713	12.33%	1,487	4.94%	1,918	6.37%	4,751	15.77%	18,253	60.60%	30,122
2006	3,065	9.60%	1,356	4.25%	1,720	5.39%	4,567	14.30%	21,231	66.47%	31,939
2007	2,901	7.99%	1,409	3.88%	2,095	5.77%	6,385	17.58%	23,524	64.78%	36,314
2008	3,716	9.83%	1,627	4.30%	2,053	5.43%	6,396	16.92%	24,006	63.51%	37,798
2009	5,133	10.72%	2,070	4.32%	2,775	5.79%	8,142	17.00%	29,768	62.16%	47,888
2010	7,483	12.83%	3,011	5.16%	3,828	6.56%	9,588	16.44%	34,417	59.01%	58,327
2011	12,271	14.57%	4,734	5.62%	5,587	6.63%	13,359	15.86%	48,292	57.32%	84,243
2012	18,975	13.37%	7,475	5.27%	9,366	6.60%	23,183	16.33%	82,934	58.43%	141,933
2013	35,964	12.45%	15,907	5.51%	19,928	6.90%	47,948	16.60%	169,070	58.54%	288,817
2014	57,426	12.33%	22,126	4.75%	30,389	6.53%	77,286	16.60%	278,436	59.79%	465,663
2015	83,884	13.32%	29,757	4.72%	37,230	5.91%	92,670	14.71%	386,237	61.33%	629,778
2016	100,674	14.26%	34,085	4.83%	39,168	5.55%	94,747	13.42%	437,542	61.96%	706,216
2017	117,382	17.19%	36,099	5.29%	36,583	5.36%	78,233	11.46%	414,393	60.70%	682,690
2018	138,631	20.79%	37,332	5.60%	33,899	5.08%	66,210	9.93%	390,592	58.59%	666,664
2019	182,124	18.64%	45,164	4.62%	44,561	4.56%	93,960	9.61%	611,513	62.57%	977,322
2020	302,770	19.83%	74,124	4.86%	72,606	4.76%	137,666	9.02%	939,460	61.54%	1,526,626
2021	390,890	23.75%	96,710	5.88%	87,661	5.33%	135,984	8.26%	934,802	56.79%	1,646,047
2022	411,761	29.30%	98,892	7.04%	85,416	6.08%	104,605	7.44%	704,830	50.15%	1,405,504
2023	209,316	29.20%	54,188	7.56%	45,341	6.33%	53,651	7.49%	354,239	49.42%	716,735
Total	2,097,655	20.35%	572,779	5.56%	569,293	5.52%	1,079,061	10.47%	5,990,341	58.11%	10,309,129

* Note: Total 10,309,129 reviews from 10,309,129 unique reviewers for 1,908,223 unique products between 8 August 1997 and 12 September 2023.

Table A. 3 Rating distribution of OTNOTA reviews over years

Year	OTNOTA										
	Rating 1		Rating 2		Rating 3		Rating 4		Rating 5		Total
	# of	% in	# of	% in	# of	% in	# of	% in	# of	% in	
1996	0	0.00%	0	0.00%	0	0.00%	1	50.00%	1	50.00%	2
1997	11	2.78%	20	5.05%	18	4.55%	62	15.66%	285	71.97%	396
1998	137	3.72%	96	2.61%	170	4.61%	605	16.42%	2,677	72.65%	3,685
1999	609	5.44%	352	3.15%	544	4.86%	1,771	15.83%	7,913	70.72%	11,189
2000	1,846	5.88%	1,349	4.30%	2,008	6.40%	5,750	18.32%	20,429	65.10%	31,382
2001	2,692	7.82%	1,824	5.30%	2,418	7.02%	6,093	17.69%	21,411	62.17%	34,438
2002	3,634	9.34%	2,301	5.91%	2,589	6.65%	6,433	16.53%	23,968	61.57%	38,925
2003	4,438	10.08%	2,654	6.03%	3,025	6.87%	7,301	16.59%	26,592	60.42%	44,010
2004	7,287	12.82%	3,529	6.21%	4,040	7.11%	8,925	15.70%	33,063	58.16%	56,844
2005	11,972	13.05%	5,758	6.27%	6,788	7.40%	15,414	16.80%	51,834	56.48%	91,766
2006	14,071	12.04%	7,091	6.07%	8,379	7.17%	19,867	17.00%	67,465	57.73%	116,873
2007	20,714	9.12%	11,395	5.01%	16,210	7.13%	42,208	18.57%	136,712	60.16%	227,239
2008	27,126	10.13%	14,303	5.34%	18,719	6.99%	49,887	18.62%	157,867	58.93%	267,902
2009	38,288	10.69%	19,617	5.48%	25,214	7.04%	64,375	17.97%	210,795	58.83%	358,289
2010	59,980	11.76%	29,443	5.77%	38,853	7.62%	89,353	17.52%	292,504	57.34%	510,133
2011	104,671	12.36%	50,987	6.02%	63,382	7.49%	142,801	16.86%	484,894	57.27%	846,735
2012	172,366	11.17%	84,932	5.51%	113,149	7.34%	258,276	16.74%	913,737	59.24%	1,542,460
2013	343,784	9.17%	196,434	5.24%	295,305	7.87%	649,276	17.31%	2,266,074	60.41%	3,750,873
2014	555,809	9.65%	287,959	5.00%	423,700	7.36%	921,045	15.99%	3,570,123	62.00%	5,758,636
2015	842,685	10.14%	408,543	4.91%	571,447	6.87%	1,203,840	14.48%	5,285,876	63.59%	8,312,391
2016	1,072,595	10.61%	499,503	4.94%	665,582	6.58%	1,363,161	13.49%	6,507,776	64.38%	10,108,617
2017	1,296,495	12.55%	557,339	5.40%	666,687	6.46%	1,225,872	11.87%	6,581,629	63.73%	10,328,022
2018	1,599,341	14.35%	610,125	5.47%	684,674	6.14%	1,166,036	10.46%	7,086,038	63.57%	11,146,214
2019	2,023,783	13.07%	706,765	4.56%	810,524	5.23%	1,479,293	9.55%	10,468,833	67.59%	15,489,198
2020	2,898,721	14.67%	1,001,768	5.07%	1,125,340	5.70%	1,810,821	9.17%	12,919,074	65.39%	19,755,724
2021	3,517,664	17.40%	1,200,371	5.94%	1,306,627	6.46%	1,728,039	8.55%	12,461,571	61.65%	20,214,272
2022	3,184,708	19.85%	1,066,270	6.64%	1,131,616	7.05%	1,325,667	8.26%	9,338,347	58.20%	16,046,608
2023	1,339,054	18.83%	468,474	6.59%	496,068	6.98%	606,894	8.53%	4,201,515	59.08%	7,112,005
Total	19,144,481	14.48%	7,239,202	5.48%	8,483,076	6.42%	14,199,066	10.74%	83,139,003	62.89%	132,204,828

* Note: Total reviews 132,204,828 from 40,630,865 unique reviewers for 20,000,660 unique products between 15 November 1996 and 14 September 2023.

Table A. 4 Rating distribution of all reviews in each product category

Product Category	1-star-rating	2-star-rating	3-star-rating	4-star-rating	5-star-rating	Count (% in total)	Share of 1- and 5-ratings
All Beauty	102,080 (14.55%)	43,034 (6.13%)	56,307 (8.03%)	79,381 (11.32%)	420,726 (59.97%)	701,528 (0.12%)	74.52%
Amazon Fashion	303,411 (12.13%)	173,462 (6.94%)	245,471 (9.82%)	352,327 (14.09%)	1,426,268 (57.03%)	2,500,939 (0.44%)	69.16%
Appliances	250,453 (11.77%)	79,664 (3.74%)	104,047 (4.89%)	208,216 (9.78%)	1,486,225 (69.82%)	2,128,605 (0.37%)	81.59%
Arts Crafts and Sewing	883,616 (9.85%)	417,130 (4.65%)	607,210 (6.77%)	942,344 (10.51%)	6,116,457 (68.21%)	8,966,757 (1.57%)	78.07%
Automotive	2,302,427 (11.54%)	851,553 (4.27%)	1,185,429 (5.94%)	2,172,554 (10.89%)	13,443,487 (67.37%)	19,955,450 (3.49%)	78.91%
Baby Products	562,663 (9.33%)	313,177 (5.19%)	437,934 (7.26%)	689,930 (11.44%)	4,025,179 (66.76%)	6,028,883 (1.05%)	76.10%
Beauty and Personal Care	2,840,878 (11.88%)	1,275,031 (5.33%)	1,715,821 (7.18%)	2,616,444 (10.94%)	15,463,216 (64.67%)	23,911,390 (4.18%)	76.55%
Books	1,316,085 (4.47%)	1,080,897 (3.67%)	2,054,057 (6.97%)	4,632,932 (15.72%)	20,391,478 (69.18%)	29,475,449 (5.16%)	73.65%
CDs and Vinyl	187,727 (3.89%)	140,557 (2.91%)	283,691 (5.88%)	664,514 (13.77%)	3,550,784 (73.56%)	4,827,273 (0.84%)	77.45%
Cell Phones and Accessories	2,802,212 (13.46%)	1,251,658 (6.01%)	1,543,387 (7.42%)	2,446,330 (11.75%)	12,769,358 (61.35%)	20,812,945 (3.64%)	74.82%
Clothing Shoes and Jewelry	5,778,281 (8.75%)	3,638,852 (5.51%)	5,529,471 (8.37%)	9,054,462 (13.71%)	42,032,280 (63.65%)	66,033,346(11.55%)	72.40%
Digital Music	6,136 (4.70%)	3,159 (2.42%)	6,392 (4.90%)	14,129 (10.83%)	100,618 (77.14%)	130,434 (0.02%)	81.85%
Electronics	5,358,800 (12.21%)	2,257,604 (5.14%)	2,883,065 (6.57%)	5,557,725 (12.66%)	27,829,748 (63.41%)	43,886,942 (7.68%)	75.62%
Gift Cards	12,326 (8.09%)	1,873 (1.23%)	3,271 (2.15%)	6,692 (4.39%)	128,248 (84.15%)	152,410 (0.03%)	92.23%
Grocery and Gourmet Food	1,855,049 (12.96%)	695,428 (4.86%)	898,317 (6.27%)	1,256,795 (8.78%)	9,612,930 (67.14%)	14,318,519 (2.51%)	80.09%
Handmade Products	41,592 (6.26%)	20,531 (3.09%)	29,944 (4.51%)	45,247 (6.81%)	526,848 (79.33%)	664,162 (0.12%)	85.59%
Health and Household	2,772,511 (10.82%)	1,186,802 (4.63%)	1,597,024 (6.23%)	2,750,389 (10.73%)	17,324,619 (67.59%)	25,631,345 (4.48%)	78.41%
Health and Personal Care	69,564 (14.08%)	28,895 (5.85%)	36,949 (7.48%)	57,000 (11.54%)	301,713 (61.06%)	494,121 (0.09%)	75.14%
Home and Kitchen	7,160,456 (10.62%)	3,410,803 (5.06%)	4,600,479 (6.82%)	7,536,382 (11.18%)	44,701,824 (66.31%)	67,409,944 (11.79%)	76.94%
Industrial and Scientific	584,133 (11.27%)	234,763 (4.53%)	315,792 (6.09%)	560,932 (10.82%)	3,487,385 (67.29%)	5,183,005 (0.91%)	78.56%
Kindle Store	738,729 (2.89%)	812,301 (3.18%)	1,940,802 (7.59%)	5,386,277 (21.06%)	16,699,504 (65.29%)	25,577,613 (4.48%)	68.18%
Magazine Subscriptions	9,685 (13.55%)	3,953 (5.53%)	5,033 (7.04%)	8,206 (11.48%)	44,620 (62.41%)	71,497 (0.01%)	75.95%
Movies and TV	1,410,250 (8.14%)	766,735 (4.42%)	1,260,217 (7.27%)	2,604,827 (15.03%)	11,286,285 (65.13%)	17,328,314 (3.03%)	73.27%
Musical Instruments	264,352 (8.76%)	131,441 (4.36%)	197,131 (6.53%)	400,387 (13.27%)	2,024,128 (67.08%)	3,017,439 (0.53%)	75.84%
Office Products	1,353,519 (10.54%)	589,120 (4.59%)	799,194 (6.22%)	1,400,675 (10.90%)	8,703,203 (67.75%)	12,845,711 (2.25%)	78.29%
Patio Lawn and Garden	2,282,783 (13.84%)	868,187 (5.26%)	1,083,552 (6.57%)	1,782,189 (10.81%)	10,473,336 (63.51%)	16,490,047 (2.89%)	77.36%
Pet Supplies	2,034,198 (12.09%)	957,197 (5.69%)	1,242,552 (7.38%)	1,801,732 (10.71%)	10,792,183 (64.13%)	16,827,862 (2.94%)	76.22%
Software	695,854 (14.26%)	239,253 (4.90%)	419,356 (8.59%)	857,082 (17.56%)	2,668,636 (54.68%)	4,880,181 (0.85%)	68.94%
Sports and Outdoors	1,836,990 (9.37%)	933,101 (4.76%)	1,324,911 (6.76%)	2,518,170 (12.85%)	12,981,998 (66.25%)	19,595,170 (3.43%)	75.63%
Subscription Boxes	2,576 (15.89%)	1,491 (9.19%)	1,752 (10.80%)	1,663 (10.26%)	8,734 (53.86%)	16,216 (0.003%)	69.75%
Tools and Home Improvement	3,057,747 (11.33%)	1,291,499 (4.79%)	1,709,616 (6.34%)	3,078,376 (11.41%)	17,845,01 (66.14%)	26,982,256 (4.72%)	77.47%
Toys and Games	1,631,131 (10.03%)	767,908 (4.72%)	1,103,648 (6.79%)	1,797,010 (11.05%)	10,960,709 (67.41%)	16,260,406 (2.84%)	77.44%
Unknown	7,832,796 (12.27%)	3,389,685 (5.31%)	4,767,525 (7.47%)	8,603,831 (13.48%)	39,220,135 (61.46%)	63,813,972 (11.17%)	73.73%
Video Games	589,519 (12.75%)	249,878 (5.40%)	340,086 (7.35%%)	617,251 (13.35%)	2,827,881 (61.15%)	4,624,615 (0.81%)	73.90%
Sum	58,930,529 (10.31%)	28,106,622 (4.92%)	40,329,433 (7.06%)	72,502,401 (12.69%)	371,675,761 (65.03%)	571,544,746 (100%)	75.34%

* Note: Total 571,544,746 reviews from 54,514,263 unique reviewers for 48,185,048 unique products between 17 June 1996 and 14 September 2023.

Table A. 5 Rating distribution of OT (i.e., the union of OTA and OTNOTA) reviews in each product category

Product Category	1-star-rating	2-star-rating	3-star-rating	4-star-rating	5-star-rating	Count (% in total OT)	Share of 1- and 5-ratings
All Beauty	88,445 (15.16%)	36,168 (6.20%)	45,674 (7.83%)	63,504 (10.88%)	349,762 (59.94%)	583,553 (0.41%)	75.09%
Amazon Fashion	227,319 (13.09%)	121,864 (7.02%)	164,081 (9.45%)	233,004 (13.42%)	990,438 (57.03%)	1,736,706 (1.22%)	70.12%
Appliances	195,823 (13.09%)	59,590 (3.98%)	73,236 (4.89%)	141,926 (9.48%)	1,025,936 (68.56%)	1,496,511 (1.05%)	81.64%
Arts Crafts and Sewing	439,988 (14.02%)	177,912 (5.67%)	220,189 (7.02%)	318,530 (10.15%)	1,981,689 (63.15%)	3,138,308 (2.20%)	77.17%
Automotive	784,735 (16.72%)	244,887 (5.22%)	282,131 (6.01%)	470,809 (10.03%)	2,909,762 (62.01%)	4,692,324 (3.29%)	78.73%
Baby Products	281,301 (12.01%)	125,053 (5.34%)	151,107 (6.45%)	236,075 (10.08%)	1,549,405 (66.13%)	2,342,941 (1.64%)	78.14%
Beauty and Personal Care	1,112,118 (15.92%)	386,625 (5.53%)	431,126 (6.17%)	665,171 (9.52%)	4,392,030 (62.86%)	6,987,070 (4.90%)	78.78%
Books	404,654 (6.87%)	215,194 (3.65%)	306,403 (5.20%)	655,282 (11.13%)	4,307,858 (73.15%)	5,889,391 (4.13%)	80.02%
CDs and Vinyl	65,420 (5.99%)	32,857 (3.01%)	48,758 (4.46%)	102,355 (9.37%)	843,165 (77.17%)	1,092,555 (0.77%)	83.16%
Cell Phones and Accessories	1,220,803 (16.40%)	466,463 (6.27%)	503,617 (6.76%)	803,908 (10.80%)	4,450,658 (59.78%)	7,445,449 (5.22%)	76.17%
Clothing Shoes and Jewelry	1,439,535 (13.24%)	677,818 (6.24%)	820,213 (7.55%)	1,333,765 (12.27%)	6,598,108 (60.70%)	10,869,439 (7.63%)	73.95%
Digital Music	4,708 (5.34%)	2,254 (2.56%)	4,030 (4.58%)	8,202 (9.31%)	68,890 (78.21%)	88,084 (0.06%)	83.55%
Electronics	1,750,289 (16.96%)	580,360 (5.62%)	614,116 (5.95%)	1,147,012 (11.12%)	6,226,940 (60.35%)	10,318,717 (7.24%)	77.31%
Gift Cards	11,386 (9.50%)	1,710 (1.43%)	2,802 (2.34%)	5,370 (4.48%)	98,592 (82.26%)	119,860 (0.08%)	91.76%
Grocery and Gourmet Food	862,087 (18.68%%)	244,645 (5.30%)	259,878 (5.63%)	330,937 (7.17%)	2,918,514 (63.23%)	4,616,061 (3.24%)	81.90%
Handmade Products	35,950 (6.78%)	17,407 (3.28%)	24,789 (4.67%)	36,135 (6.81%)	416,344 (78.46%)	530,625 (0.37%)	85.24%
Health and Household	1,162,341 (14.79%)	395,441 (5.03%)	444,390 (5.65%)	764,886 (9.73%)	5,092,294 (64.79%)	7,859,352 (5.51%)	79.58%
Health and Personal Care	64,147 (14.62%)	26,044 (5.94%)	32,451 (7.40%)	48,851 (11.14%)	267,217 (60.91%)	438,710 (0.31%)	75.53%
Home and Kitchen	1,942,445 (17.00%)	699,775 (6.12%)	736,204 (6.44%)	1,152,671 (10.09%)	6,894,294 (60.34%)	11,425,389 (8.02%)	77.34%
Industrial and Scientific	377,341 (14.26%)	135,138 (5.11%)	162,575 (6.14%)	264,597 (10.00%)	1,706,633 (64.49%)	2,646,284 (1.86%)	78.75%
Kindle Store	149,705 (5.23%)	102,697 (3.59%)	182,383 (6.37%)	466,603 (16.30%)	1,960,628 (68.51%)	2,862,016 (2.01%)	73.74%
Magazine Subscriptions	8,293 (15.67%)	3,130 (5.91%)	3,615 (6.83%)	5,635 (10.65%)	32,256 (60.94%)	52,929 (0.04%)	76.61%
Movies and TV	418,983 (11.01%)	145,725 (3.83%)	193,011 (5.07%)	437,202 (11.49%)	2,611,224 (68.61%)	3,806,145 (2.67%)	79.61%
Musical Instruments	152,431 (11.69%)	63,410 (4.86%)	81,715 (6.27%)	159,082 (12.20%)	847,608 (64.99%)	1,304,246 (0.92%)	76.68%
Office Products	755,577 (14.17%)	281,240 (5.28%)	328,307 (6.16%)	539,480 (10.12%)	3,426,497 (64.27%)	5,331,101 (3.74%)	78.45%
Patio Lawn and Garden	1,014,330 (18.26%)	330,512 (5.95%)	353,617 (6.37%)	549,537 (9.89%)	3,306,255 (59.53%)	5,554,251 (3.90%)	77.79%
Pet Supplies	785,176 (16.52%)	293,037 (6.17%)	311,297 (6.55%)	449,974 (9.47%)	2,912,078 (61.29%)	4,751,562 (3.33%)	77.81%
Software	312,586 (17.93%)	87,122 (5.00%)	130,360 (7.48%)	285,339 (16.37%)	928,045 (53.23%)	1,743,452 (1.22%)	71.16%
Sports and Outdoors	873,497 (12.93%)	367,669 (5.44%)	441,126 (6.53%)	774,859 (11.47%)	4,298,678 (63.63%)	6,755,829 (4.74%)	76.56%
Subscription Boxes	2,335 (16.14%)	1,355 (9.36%)	1,567 (10.83%)	1,480 (10.23%)	7,734 (53.44%)	14,471 (0.01%)	69.58%
Tools and Home Improvement	1,210,972 (16.26%)	417,171 (5.60%)	459,434 (6.17%)	750,193 (10.07%)	4,608,907 (61.89%)	7,446,677 (5.23%)	78.15%
Toys and Games	785,160 (15.05%)	300,082 (5.75%)	354,831 (6.80%)	526,012 (%)	3,250,466 (62.31%)	5,216,551 (3.66%)	77.36%
Unknown	1,963,463 (17.28%)	654,280 (5.76%)	751,344 (6.61%)	1,317,057 (%)	6,673,501 (58.75%)	11,359,645 (7.97%)	76.03%
Video Games	338,793 (16.96%)	117,346 (5.87%)	131,992 (6.61%)	232,684 (%)	1,176,938 (58.91%)	1,997,753 (1.40%)	75.87%
Sum	21,242,136 (14.91%)	7,811,981 (5.48%)	9,052,369 (6.35%)	15,278,127 (10.72%)	89,129,344 (62.54%)	142,513,957 (100%)	**77.45%**

* Note: Total 142,513,957 reviews from 50,939,994 unique reviewers for 20,208,041unique products between 15 November 1996 and 14 September 2023.

Table A. 6 Rating distribution of OTA reviews in each product category

Product Category	1-star-rating	2-star-rating	3-star-rating	4-star-rating	5-star-rating	Count (% in total OTA)	Share of 1- and 5-ratings
All Beauty	1,922 (21.79%)	489 (5.54%)	453 (5.14%)	805 (9.13%)	5,150 (58.40%)	8,819 (0.09%)	80.19%
Amazon Fashion	3,525 (20.56%)	1,137 (6.63%)	1,238 (7.22%)	2,075 (12.10%)	9,170 (53.49%)	17,145 (0.17%)	74.04%
Appliances	10,290 (24.87%)	1,865 (4.51%)	1,630 (3.94%)	3,564 (8.61%)	24,024 (58.07%)	41,373 (0.40%)	82.94%
Arts Crafts and Sewing	30,244 (22.04%)	8,490 (6.19%)	8,426 (6.14%)	12,586 (9.17%)	77,456 (56.45%)	137,202 (1.33%)	78.50%
Automotive	59,969 (24.86%)	12,783 (5.30%)	12,151 (5.04%)	22,187 (9.20%)	134,123 (55.60%)	241,213 (2.34%)	80.47%
Baby Products	23,019 (19.22%)	6,799 (5.68%)	6,347 (5.30%)	10,424 (8.70%)	73,181 (61.10%)	119,770 (1.16%)	80.32%
Beauty and Personal Care	125,705 (23.22%)	28,648 (5.29%)	27,254 (5.04%)	47,180 (8.72%)	312,495 (57.73%)	541,282 (5.25%)	80.96%
Books	43,532 (8.10%)	20,695 (3.85%)	27,638 (5.15%)	70,944 (13.21%)	374,312 (69.69%)	537,121 (5.21%)	77.79%
CDs and Vinyl	7,441 (8.68%)	2,873 (3.35%)	3,675 (4.29%)	9,148 (10.67%)	62,590 (73.01%)	85,727 (0.83%)	81.69%
Cell Phones and Accessories	111,297 (22.12%)	29,644 (5.89%)	26,880 (5.34%)	49,928 (9.92%)	285,380 (56.72%)	503,129 (4.88%)	78.84%
Clothing Shoes and Jewelry	224,498 (17.31%)	82,272 (6.34%)	88,625 (6.83%)	151,147 (11.65%)	750,318 (57.86%)	1,296,860 (12.58%)	75.17%
Digital Music	135 (13.80%)	37 (3.78%)	56 (5.73%)	75 (7.67%)	675 (69.02%)	978 (0.01%)	82.82%
Electronics	235,427 (22.07%)	61,596 (5.77%)	56,889 (5.33%)	119,803 (11.23%)	593,110 (55.60%)	1,066,82 (10.35%)	77.66%
Gift Cards	1,013 (27.57%)	87 (2.37%)	119 (3.24%)	202 (5.50%)	2,253 (61.32%)	3,674 (0.04%)	88.89%
Grocery and Gourmet Food	88,040 (32.41%)	15,551 (5.73%)	13,268 (4.88%)	16,089 (5.92%)	138,670 (51.05%)	271,618 (2.63%)	83.47%
Handmade Products	864 (11.97%)	265 (3.67%)	256 (3.55%)	340 (4.71%)	5,495 (76.11%)	7,220 (0.07%)	88.07%
Health and Household	112,829 (20.09%)	25,093 (4.47%)	23,883 (4.25%)	51,774 (9.22%)	348,061 (61.97%)	561,640 (5.45%)	82.06%
Health and Personal Care	1,718 (23.35%)	383 (5.20%)	348 (4.73%)	669 (9.09%)	4,241 (57.63%)	7,359 (0.07%)	80.98%
Home and Kitchen	314,325 (23.78%)	86,583 (6.55%)	76,210 (5.77%)	119,752 (9.06%)	724,903 (54.84%)	1,321,773 (12.82%)	78.62%
Industrial and Scientific	20,453 (26.68%)	4,521 (5.90%)	4,053 (5.29%)	6,655 (8.68%)	40,988 (53.46%)	76,670 (0.74%)	80.14%
Kindle Store	16,117 (4.76%)	10,673 (6.08%)	20,784 (6.13%)	60,843 (17.96%)	230,400 (68.00%)	338,817 (3.29%)	72.76%
Magazine Subscriptions	488 (30.60%)	97 (3.15%)	119 (7.46%)	160 (10.03%)	731 (45.83%)	1,595 (0.02%)	76.43%
Movies and TV	50,545 (13.28%)	13,442 (3.53%)	17,081 (4.49%)	42,894 (11.27%)	256,703 (67.44%)	380,665 (3.69%)	80.71%
Musical Instruments	11,104 (18.28%)	3,197 (5.26%)	3,319 (5.46%)	7,342 (12.08%)	35,792 (58.91%)	60,754 (0.59%)	77.19%
Office Products	60,685 (23.15%)	15,265 (5.82%)	13,790 (5.26%)	23,239 (8.87%)	149,111 (56.89%)	262,090 (2.54%)	80.05%
Patio Lawn and Garden	77,966 (27.98%)	17,997 (6.46%)	15,095 (5.42%)	25,231 (9.05%)	142,397 (51.10%)	278,686 (2.70%)	79.07%
Pet Supplies	81,549 (23.50%)	22,072 (6.36%)	18,589 (5.36%)	29,520 (8.51%)	195,226 (56.27%)	346,956 (3.37%)	79.77%
Software	21,846 (15.25%)	5,298 (5.71%)	9,629 (6.72%)	25,735 (17.97%)	80,720 (56.36%)	143,228 (1.39%)	71.61%
Sports and Outdoors	73,929 (19.65%)	21,485 (7.13%)	20,050 (5.33%)	40,076 (10.65%)	220,769 (58.67%)	376,309 (3.65%)	78.31%
Subscription Boxes	79 (18.16%)	31 (3.70%)	25 (5.75%)	28 (6.44%)	272 (62.53%)	435 (0.004%)	80.69%
Tools and Home Improvement	108,161 (24.92%)	26,163 (6.03%)	23,401 (5.39%)	39,657 (9.14%)	236,691 (54.53%)	434,073 (4.21%)	79.45%
Toys and Games	71,937 (23.38%)	19,290 (6.27%)	18,331 (5.96%)	28,148 (9.15%)	170,016 (55.25%)	307,722 (2.98%)	78.63%
Unknown	69,970 (18.47%)	18,594 (4.91%)	20,866 (5.51%)	44,535 (11.76%)	224,843 (59.36%)	378,808 (3.67%)	77.83%
Video Games	37,033 (24.43%)	9,364 (6.18%)	8,815 (5.81%)	16,306 (10.76%)	80,075 (52.82%)	151,593 (1.47%)	77.25%
Sum	2,097,655 (20.35%)	572,779 (5.56%)	569,293 (5.52%)	1,079,061 (10.47%)	5,990,341 (58.11%)	10,309,129 (100%)	**78.45%**

* Note: Total 10,309,129 reviews from 10,309,129 unique reviewers for 1,908,223 unique products between 8 August 1997 and 12 September 2023.

Table A. 7 Rating distribution of OTNOTA reviews in each product category

Product Category	1-star-rating	2-star-rating	3-star-rating	4-star-rating	5-star-rating	Count (% in total OTNOTA)	Share of 1- and 5- ratings
All Beauty	86,523 (15.05%)	35,679 (6.21%)	45,221 (7.87%)	62,699 (10.91%)	344,612 (59.96%)	574,734 (0.43%)	75.01%
Amazon Fashion	223,794 (13.01%)	120,727 (7.02%)	162,843 (9.47%)	230,929 (13.43%)	981,268 (57.07%)	1,719,561 (1.30%)	70.08%
Appliances	185,533 (12.75%)	57,725 (3.97%)	71,606 (4.92%)	138,362 (9.51%)	1,001,912 (68.85%)	1,455,138 (1.10%)	81.60%
Arts Crafts and Sewing	409,744 (13.65%)	169,422 (5.65%)	211,763 (7.06%)	305,944 (10.19%)	1,904,233 (63.45%)	3,001,106 (2.27%)	77.10%
Automotive	724,766 (16.28%)	232,104 (5.21%)	269,980 (6.07%)	448,622 (10.08%)	2,775,639 (62.36%)	4,451,111 (3.37%)	78.64%
Baby Products	258,282 (11.62%)	118,254 (5.32%)	144,760 (6.51%)	225,651 (10.15%)	1,476,224 (66.40%)	2,223,171 (1.68%)	78.02%
Beauty and Personal Care	986,413 (15.30%)	357,977 (5.55%)	403,872 (6.27%)	617,991 (9.59%)	4,079,535 (63.29%)	6,445,788 (4.88%)	78.59%
Books	361,122 (6.75%)	194,499 (3.63%)	278,765 (5.21%)	584,338 (10.92%)	3,933,546 (73.49%)	5,352,270 (4.05%)	80.24%
CDs and Vinyl	57,979 (5.76%)	29,984 (2.98%)	45,083 (4.48%)	93,207 (9.26%)	780,575 (77.53%)	1,006,828 (0.76%)	83.29%
Cell Phones and Accessories	1,109,506 (15.98%)	436,819 (6.29%)	476,737 (6.87%)	753,980 (10.86%)	4,165,278 (60.00%)	6,942,320 (5.25%)	75.98%
Clothing Shoes and Jewelry	1,215,037 (12.69%)	595,546 (6.22%)	731,588 (7.64%)	1,182,618 (12.35%)	5,847,790 (61.09%)	9,572,579 (7.24%)	73.78%
Digital Music	4,573 (5.25%)	2,217 (2.55%)	3,974 (4.56%)	8,127 (9.33%)	68,215 (78.31%)	87,106 (0.07%)	83.56%
Electronics	1,514,862 (16.37%)	518,764 (5.61%)	557,227 (6.02%)	1,027,209 (11.10%)	5,633,830 (60.89%)	9,251,892 (7.00%)	77.27%
Gift Cards	10,373 (8.93%)	1,623 (1.40%)	2,683 (2.31%)	5,168 (4.45%)	96,339 (82.92%)	116,186 (0.09%)	91.85%
Grocery and Gourmet Food	774,047 (17.82%)	229,094 (5.27%)	246,610 (5.68%)	314,848 (7.25%)	2,779,844 (63.99%)	4,344,443 (3.29%)	81.80%
Handmade Products	35,086 (6.70%)	17,142 (3.28%)	24,533 (4.69%)	35,795 (6.84%)	410,849 (78.50%)	523,405 (0.40%)	85.20%
Health and Household	1,049,512 (14.38%)	370,348 (5.07%)	420,507 (5.76%)	713,112 (9.77%)	4,744,233 (65.01%)	7,297,712 (5.52%)	79.39%
Health and Personal Care	62,429 (14.47%)	25,661 (5.95%)	32,103 (7.44%)	48,182 (11.17%)	262,976 (60.97%)	431,351 (0.33%)	75.44%
Home and Kitchen	1,628,120 (16.11%)	613,192 (6.07%)	659,994 (6.53%)	1,032,919 (10.22%)	6,169,391 (61.06%)	10,103,616 (7.64%)	77.18%
Industrial and Scientific	356,888 (13.89%)	130,617 (5.08%)	158,522 (6.17%)	257,942 (10.04%)	1,665,645 (64.82%)	2,569,614 (1.94%)	78.71%
Kindle Store	133,588 (5.29%)	92,024 (3.65%)	161,599 (6.40%)	405,760 (16.08%)	1,730,228 (68.57%)	2,523,199 (1.91%)	73.87%
Magazine Subscriptions	7,805 (15.20%)	3,033 (5.91%)	3,496 (6.81%)	5,475 (10.67%)	31,525 (61.41%)	51,334 (0.04%)	76.62%
Movies and TV	368,438 (10.76%)	132,283 (3.86%)	175,930 (5.14%)	394,308 (11.51%)	2,354,521 (68.74%)	3,425,480 (2.59%)	79.49%
Musical Instruments	141,327 (11.37%)	60,213 (4.84%)	78,396 (6.30%)	151,740 (12.20%)	811,816 (65.29%)	1,243,492 (0.94%)	76.65%
Office Products	694,892 (13.71%)	265,975 (5.25%)	314,517 (6.20%)	516,241 (10.18%)	3,277,386 (64.66%)	5,069,011 (3.83%)	78.36%
Patio Lawn and Garden	936,364 (17.75%)	312,515 (5.92%)	338,522 (6.42%)	524,306 (9.94%)	3,163,858 (59.97%)	5,275,565 (3.99%)	77.72%
Pet Supplies	703,627 (15.97%)	270,965 (6.15%)	292,708 (6.65%)	420,454 (9.55%)	2,716,852 (61.68%)	4,404,606 (3.33%)	77.66%
Software	290,740 (18.17%)	81,824 (5.11%)	120,731 (7.54%)	259,604 (16.22%)	847,325 (52.95%)	1,600,224 (1.21%)	71.12%
Sports and Outdoors	799,568 (12.53%)	346,184 (5.43%)	421,076 (6.60%)	734,783 (11.52%)	4,077,909 (63.92%)	6,379,520 (4.83%)	76.46%
Subscription Boxes	2,256 (16.07%)	1,324 (9.43%)	1,542 (10.99%)	1,452 (10.34%)	7,462 (53.16%)	14,036 (0.01%)	69.24%
Tools and Home Improvement	1,102,811 (15.73%)	391,008 (5.58%)	436,033 (6.22%)	710,536 (10.13%)	4,372,216 (62.35%)	7,012,604 (5.30%)	78.07%
Toys and Games	713,223 (14.53%)	280,792 (5.72%)	336,500 (6.85%)	497,864 (10.14%)	3,080,450 (62.75%)	4,908,829 (3.71%)	77.28%
Unknown	1,893,493 (17.24%)	635,686 (5.79%)	730,478 (6.65%)	1,272,522 (11.59%)	6,448,658 (58.73%)	10,980,837 (8.31%)	75.97%
Video Games	301,760 (16.35%)	107,982 (5.85%)	123,177 (6.67%)	216,378 (11.72%)	1,096,863 (59.41%)	1,846,160 (1.40%)	75.76%
Sum	19,144,481 (14.48%)	7,239,202 (5.48%)	8,483,076 (6.42%)	14,199,066 (10.74%)	83,139,003 (62.89%)	132,204,828 (100%)	77.37%

* Note: Total reviews 132,204,828 from 40,630,865 unique reviewers for 20,000,660 unique products between 15 November 1996 and 14 September 2023.

Table A. 8 Rating distribution of NOT reviews in each product category

Product Category	1-star-rating	2-star-rating	3-star-rating	4-star-rating	5-star-rating	Count (% in total OTA)	Share of 1- and 5-ratings
All Beauty	13,635 (11.56%)	6,866 (5.82%)	10,633 (9.01%)	15,877 (13.46%)	70,964 (60.15%)	117,975 (0.03%)	71.71%
Amazon Fashion	76,092 (9.96%)	51,598 (6.75%)	81,390 (10.65%)	119,323 (15.61%)	435,830 (57.03%)	764,233 (0.18%)	66.99%
Appliances	54,630 (8.64%)	20,074 (3.18%)	30,811 (4.87%)	66,290 (10.49%)	460,289 (72.82%)	632,094 (0.15%)	81.46%
Arts Crafts and Sewing	443,628 (7.61%)	239,218 (4.10%)	387,021 (6.64%)	623,814 (10.70%)	4,134,768 (70.94%)	5,828,449 (1.36%)	78.55%
Automotive	1,517,692 (9.94%)	606,666 (3.97%)	903,298 (5.92%)	1,701,745 (11.15%)	10,533,725 (69.01%)	15,263,126 (3.56%)	78.96%
Baby Products	281,362 (7.63%)	188,124 (5.10%)	286,827 (7.78%)	453,855 (12.31%)	2,475,774 (67.17%)	3,685,942 (0.86%)	74.80%
Beauty and Personal Care	1,728,760 (10.21%)	888,406 (5.25%)	1,284,695 (7.59%)	1,951,273 (11.53%)	11,071,186 (65.42%)	16,924,320 (3.94%)	75.63%
Books	911,431 (3.86%)	865,703 (3.67%)	1,747,654 (7.41%)	3,977,650 (16.86%)	16,083,620 (68.19%)	23,586,058 (5.50%)	72.06%
CDs and Vinyl	122,307 (3.27%)	107,700 (2.88%)	234,933 (6.29%)	562,159 (15.05%)	2,707,619 (72.50%)	3,734,718 (0.87%)	75.77%
Cell Phones and Accessories	1,581,409 (11.83%)	785,195 (5.87%)	1,039,770 (7.78%)	1,642,422 (12.29%)	8,318,700 (62.23%)	13,367,496 (3.12%)	74.06%
Clothing Shoes and Jewelry	4,338,746 (7.87%)	2,961,034 (5.37%)	4,709,258 (8.54%)	7,720,697 (14.00%)	35,434,172 (64.23%)	55,163,907 (12.86%)	72.10%
Digital Music	1,428 (3.37%)	905 (2.14%)	2,362 (5.58%)	5,927 (14.00%)	31,728 (74.92%)	42,350 (0.01%)	78.29%
Electronics	3,608,511 (10.75%)	1,677,244 (5.00%)	2,268,949 (6.76%)	4,410,713 (13.14%)	21,602,808 (64.35%)	33,568,225 (7.82%)	75.10%
Gift Cards	940 (2.89%)	163 (0.50%)	469 (1.44%)	1,322 (4.06%)	29,656 (91.11%)	32,550 (0.01%)	94.00%
Grocery and Gourmet Food	992,962 (10.23%)	450,783 (4.65%)	638,439 (6.58%)	925,858 (9.54%)	6,694,416 (69.00%)	9,702,458 (2.26%)	79.23%
Handmade Products	5,642 (4.23%)	3,124 (2.34%)	5,155 (3.86%)	9,112 (6.82%)	110,504 (82.75%)	133,537 (0.03%)	86.98%
Health and Household	1,610,170 (9.06%)	791,361 (4.45%)	1,152,634 (6.49%)	1,985,503 (11.17%)	12,232,325 (68.83%)	17,771,993 (4.14%)	77.89%
Health and Personal Care	5,417 (9.78%)	2,851 (5.15%)	4,498 (8.12%)	8,149 (14.71%)	34,496 (62.25%)	55,411 (0.01%)	72.03%
Home and Kitchen	5,218,011 (9.32%)	2,711,028 (4.84%)	3,864,275 (6.90%)	6,383,711 (11.40%)	37,807,530 (67.53%)	55,984,555 (13.05%)	76.85%
Industrial and Scientific	206,792 (8.15%)	99,625 (3.93%)	153,217 (6.04%)	296,335 (11.68%)	1,780,752 (70.20%)	2,536,721 (0.59%)	78.35%
Kindle Store	589,024 (2.59%)	709,604 (3.12%)	1,758,419 (7.74%)	4,919,674 (21.66%)	14,738,876 (64.88%)	22,715,597 (5.29%)	67.48%
Magazine Subscriptions	1,392 (7.50%)	823 (4.43%)	1,418 (7.64%)	2,571 (13.85%)	12,364 (66.59%)	18,568 (0.004%)	74.08%
Movies and TV	991,267 (7.33%)	621,010 (4.59%)	1,067,206 (7.89%)	2,167,625 (16.03%)	8,675,061 (64.15%)	13,522,169 (3.15%)	71.49%
Musical Instruments	111,921 (6.53%)	68,031 (3.97%)	115,416 (6.74%)	241,305 (14.09%)	1,176,520 (68.67%)	1,713,193 (0.40%)	75.21%
Office Products	597,942 (7.96%)	307,880 (4.10%)	470,887 (6.27%)	861,195 (11.46%)	5,276,706 (70.22%)	7,514,610 (1.75%)	78.18%
Patio Lawn and Garden	1,268,453 (11.60%)	537,675 (4.92%)	729,935 (6.67%)	1,232,652 (11.27%)	7,167,081 (65.54%)	10,935,796 (2.55%)	77.14%
Pet Supplies	1,249,022 (10.34%)	664,160 (5.50%)	931,255 (7.71%)	1,351,758 (11.19%)	7,880,105 (65.25%)	12,076,300 (2.81%)	75.60%
Software	383,268 (12.22%)	152,131 (4.85%)	288,996 (9.21%)	571,743 (18.23%)	1,740,591 (55.49%)	3,136,729 (0.73%)	67.71%
Sports and Outdoors	963,493 (7.50%)	565,432 (4.40%)	883,785 (6.88%)	1,743,311 (13.58%)	8,683,320 (67.63%)	12,839,341 (2.99%)	75.13%
Subscription Boxes	241 (13.81%)	136 (7.79%)	185 (10.6%)	183 (10.49%)	1,000 (57.31%)	1,745 (0.0004%)	71.12%
Tools and Home Improvement	1,846,775 (9.45%)	874,328 (4.48%)	1,250,182 (6.40%)	2,328,183 (11.92%)	13,236,111 (67.75%)	19,535,579 (4.55%)	77.21%
Toys and Games	845,971 (7.66%)	467,826 (4.24%)	748,817 (6.78%)	1,270,998 (11.51%)	7,710,243 (69.81%)	11,043,855 (2.57%)	77.47%
Unknown	5,869,333 (11.19%)	2,735,405 (5.21%)	4,016,181 (7.66%)	7,286,774 (13.89%)	32,546,634 (62.05%)	52,454,327 (12.23%)	73.24%
Video Games	250,726 (9.54%)	132,532 (5.05%)	208,094 (7.92%)	384,567 (14.64%)	1,650,943 (62.85%)	2,626,862 (0.61%)	72.39%
Sum	37,688,393 (8.78%)	20,294,641 (4.73%)	31,277,064 (7.29%)	57,224,274 (13.34%)	282,546,417 (65.86%)	429,030,789 (100%)	74.64%

* Note: Total reviews 429,030,789 from 3182,7161 unique reviewers for 42,472,318 unique products between 17 June 1996 and 14 September 2023.

Table A. 9 Results from ordered probit model

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Unverified NOT reviewers	-0.481798***	-0.524877***	-0.517357***	-0.520348***	-0.523474***	-0.527086***	-0.521254***	-0.517566***
	(0.176013)	(0.178471)	(0.178650)	(0.178600)	(0.178452)	(0.178561)	(0.178576)	(0.178767)
Verified OT reviewers	-0.159947***	-0.162079***	-0.158548***	-0.158834***	-0.162185***	-0.163613***	-0.160430***	-0.161977***
	(0.048681)	(0.049496)	(0.049511)	(0.049511)	(0.049495)	(0.049513)	(0.049510)	(0.049606)
Unverified OT reviewers	-0.464979***	-0.451522***	-0.445239***	-0.447258***	-0.451238***	-0.453902***	-0.448090***	-0.449343***
	(0.056860)	(0.058028)	(0.058067)	(0.058053)	(0.058028)	(0.058050)	(0.058084)	(0.058169)
Verified OTA reviewers	-0.188486**	-0.201297**	-0.197025**	-0.198077**	-0.201234**	-0.202716**	-0.199103**	-0.200622**
	(0.084612)	(0.086691)	(0.086711)	(0.086702)	(0.086694)	(0.086706)	(0.086709)	(0.086758)
Unverified OTA reviewers	-0.170647	-0.226956*	-0.229693*	-0.233804*	-0.225131*	-0.229928*	-0.226263*	-0.231549*
	(0.124396)	(0.129038)	(0.129178)	(0.129191)	(0.129059)	(0.129090)	(0.129081)	(0.129185)
Length of review title	-0.004578***	-0.004496***	-0.004482***	-0.004488***	-0.004493***	-0.004451***	-0.004496***	-0.004495***
	(0.000591)	(0.000604)	(0.000604)	(0.000604)	(0.000604)	(0.000604)	(0.000604)	(0.000604)
Length of review body	-0.000309***	-0.000218***	-0.000217***	-0.000217***	-0.000218***	-0.000215***	-0.000219***	-0.000217***
	(0.000031)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)
Crowd's average length of review title	-0.004419	-0.000862	0.000154	0.000148	-0.000876	0.000123	-0.000667	0.000240
	(0.002781)	(0.002883)	(0.002898)	(0.002903)	(0.002882)	(0.002898)	(0.002885)	(0.002979)
Crowd's average length of review body	-0.000169	-0.000155	-0.000120	-0.000106	-0.000158	-0.000114	-0.000156	-0.000111
	(0.000121)	(0.000126)	(0.000126)	(0.000127)	(0.000126)	(0.000126)	(0.000126)	(0.000129)
2017	-0.228484***	-0.162832**	-0.132635**	-0.141395**	-0.160633**	-0.139599**	-0.153032**	-0.125101*
	(0.064986)	(0.066124)	(0.066771)	(0.066488)	(0.066193)	(0.066489)	(0.066521)	(0.066931)
2018	-0.280669***	-0.231721***	-0.212761***	-0.217627***	-0.230130***	-0.223764***	-0.225001***	-0.208701***
	(0.039098)	(0.040002)	(0.040423)	(0.040264)	(0.040061)	(0.040075)	(0.040311)	(0.040555)
2020	-0.094155***	-0.053361	-0.047874	-0.047797	-0.053449	-0.055410*	-0.051208	-0.047282
	(0.032344)	(0.033184)	(0.033234)	(0.033239)	(0.033184)	(0.033193)	(0.033225)	(0.033246)
2021	-0.225294***	-0.164710***	-0.151921***	-0.152481***	-0.165080***	-0.164732***	-0.159893***	-0.149273***
	(0.035700)	(0.036682)	(0.036900)	(0.036897)	(0.036686)	(0.036688)	(0.036860)	(0.036861)
2022	-0.237030***	-0.194285***	-0.181244***	-0.181398***	-0.194894***	-0.193892***	-0.189514***	-0.172492***
	(0.042081)	(0.043147)	(0.043349)	(0.043356)	(0.043154)	(0.043149)	(0.043300)	(0.043379)
2023	-0.157015**	-0.148400**	-0.144989*	-0.139872*	-0.152849**	-0.159398**	-0.145455*	-0.144265*
	(0.072892)	(0.074512)	(0.074602)	(0.074617)	(0.074782)	(0.074626)	(0.074570)	(0.074727)
February	0.115309**	0.113141**	0.112437**	0.112573**	0.113037**	0.110756**	0.113293**	0.109574**
	(0.047392)	(0.048569)	(0.048567)	(0.048567)	(0.048571)	(0.048575)	(0.048568)	(0.048662)
March	0.033126	0.064916	0.066454	0.066598	0.064939	0.065285	0.065259	0.063142
	(0.049632)	(0.050901)	(0.050906)	(0.050906)	(0.050903)	(0.050910)	(0.050900)	(0.051007)
April	0.082552*	0.078673	0.079640	0.079135	0.078543	0.078523	0.079394	0.080276
	(0.048582)	(0.049659)	(0.049661)	(0.049662)	(0.049660)	(0.049668)	(0.049660)	(0.049747)
May	0.042768	0.081409*	0.080215*	0.080339*	0.081149*	0.079197	0.081483*	0.079052
	(0.047305)	(0.048471)	(0.048473)	(0.048472)	(0.048473)	(0.048480)	(0.048470)	(0.048536)
June	0.117342**	0.118030**	0.117828**	0.117630**	0.117708**	0.112999**	0.119113**	0.115260**
	(0.046752)	(0.047845)	(0.047847)	(0.047845)	(0.047849)	(0.047870)	(0.047853)	(0.047921)
July	0.069193	0.074208	0.069772	0.070800	0.073638	0.071023	0.073052	0.070628
	(0.045356)	(0.046459)	(0.046486)	(0.046477)	(0.046467)	(0.046479)	(0.046468)	(0.046579)
August	0.040745	0.047893	0.047370	0.046093	0.047727	0.044458	0.049121	0.050021
	(0.043944)	(0.044945)	(0.044953)	(0.044953)	(0.044947)	(0.044962)	(0.044957)	(0.045034)
September	0.081461*	0.108436**	0.103438**	0.103841**	0.107840**	0.101568**	0.108070**	0.103255**
	(0.047845)	(0.049057)	(0.049078)	(0.049074)	(0.049064)	(0.049095)	(0.049058)	(0.049150)
October	0.051211	0.115850**	0.111692**	0.112851**	0.114834**	0.109093**	0.115600**	0.107192**
	(0.046666)	(0.047856)	(0.047878)	(0.047870)	(0.047877)	(0.047901)	(0.047858)	(0.047962)
November	-0.048607	-0.008424	-0.013187	-0.011792	-0.008884	-0.013882	-0.009628	-0.011486
	(0.048097)	(0.049271)	(0.049295)	(0.049287)	(0.049276)	(0.049301)	(0.049278)	(0.049386)
December	-0.084121*	-0.054538	-0.056395	-0.055989	-0.055140	-0.059177	-0.054233	-0.055898
	(0.044841)	(0.045870)	(0.045873)	(0.045872)	(0.045879)	(0.045892)	(0.045869)	(0.045955)
Tuesday	-0.030051	-0.011969	-0.012399	-0.011703	-0.012492	-0.011823	-0.012174	-0.011597
	(0.036266)	(0.037192)	(0.037198)	(0.037194)	(0.037201)	(0.037196)	(0.037194)	(0.037236)
Wednesday	-0.003976	0.006532	0.004948	0.006292	0.006180	0.005691	0.005611	0.005864
	(0.036191)	(0.037038)	(0.037045)	(0.037041)	(0.037042)	(0.037044)	(0.037044)	(0.037090)
Thursday	0.012300	0.010575	0.011142	0.011961	0.010197	0.011949	0.010327	0.014307
	(0.036487)	(0.037365)	(0.037375)	(0.037374)	(0.037371)	(0.037376)	(0.037368)	(0.037427)
Friday	-0.003853	0.004666	0.003891	0.005173	0.003813	0.003905	0.004367	0.006476
	(0.037205)	(0.038114)	(0.038122)	(0.038122)	(0.038133)	(0.038123)	(0.038116)	(0.038161)
Saturday	0.028256	0.054921	0.055224	0.055568	0.054546	0.055201	0.055045	0.055804

	(0.037071)	(0.038054)	(0.038057)	(0.038054)	(0.038060)	(0.038056)	(0.038056)	(0.038083)
Sunday	0.007963	0.031880	0.032683	0.033247	0.031697	0.033214	0.031745	0.033231
	(0.038319)	(0.039267)	(0.039273)	(0.039272)	(0.039269)	(0.039273)	(0.039269)	(0.039325)
Negative Price Sentiment		-1.280668***	-1.278823***	-1.278442***	-1.280973***	-1.279618***	-1.280239***	-1.278949***
		(0.027615)	(0.027623)	(0.027626)	(0.027619)	(0.027620)	(0.027617)	(0.027666)
Positive Price Sentiment		0.730199***	0.731554***	0.731797***	0.729893***	0.728517***	0.730991***	0.730747***
		(0.039266)	(0.039279)	(0.039282)	(0.039267)	(0.039270)	(0.039273)	(0.039288)
Crowd's average rating			0.083620***					
			(0.025476)					
Crowd's share of 5-star ratings				0.262685***				
				(0.083113)				
Crowd's share of 4-star ratings					0.110368			
					(0.154465)			
Crowd's share of 2-star ratings						-0.476287***		
						(0.139801)		
Crowd's share of 1-star ratings							-0.152351	
							(0.113264)	
Crowd's positive imbalance								0.314440***
								(0.083417)
Threshold 1	-1.760533***	-1.894206***	-1.552516***	-1.749323***	-1.868377***	-2.018682***	-1.893339***	-1.651767***
	(0.316477)	(0.323219)	(0.339375)	(0.326002)	(0.325394)	(0.325239)	(0.323198)	(0.329878)
Threshold 2	-1.398986***	-1.477536***	-1.135638***	-1.332322***	-1.451735***	-1.601675***	-1.476704***	-1.234287***
	(0.316385)	(0.323103)	(0.339282)	(0.325901)	(0.325275)	(0.325114)	(0.323082)	(0.329775)
Threshold 3	-1.055594***	-1.069612***	-0.727506**	-0.924103***	-1.043859***	-1.193659***	-1.068742***	-0.825991**
	(0.316323)	(0.323017)	(0.339218)	(0.325828)	(0.325182)	(0.325026)	(0.322997)	(0.329699)
Threshold 4	-0.764126**	-0.726587**	-0.384279	-0.580930*	-0.700828**	-0.850514***	-0.725643**	-0.482480
	(0.316284)	(0.322962)	(0.339182)	(0.325780)	(0.325129)	(0.324969)	(0.322942)	(0.329657)
LL	-18,387	-16,928	-16,923	-16,923	-16,928	-16,923	-16,927	-16,875
AIC	37,332	34,419	34,410	34,411	34,420	34,409	34,419	34,309
BIC	39,449	36,550	36,549	36,550	36,560	36,548	36,558	36,432
N	14,565	14,565	14,565	14,565	14,565	14,565	14,565	14,526

* Note: *p < .10; **p < .05; **p < .01. Store dummies are excluded for space. Perfectly predicted samples are dropped.

Table A. 10 Marginal effect of ordered probit model

Marginal effect (variables)	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Unverified NOT – P(Y = 1)	0.112617**	0.102432**	0.100869**	0.101541**	0.102097**	0.102754**	0.101700**	0.100520**
	(0.048671)	(0.041193)	(0.041104)	(0.041159)	(0.041151)	(0.041198)	(0.041164)	(0.040998)
Unverified NOT– P(Y = 2)	0.033364***	0.032979***	0.032510***	0.032694***	0.032893***	0.033113***	0.032753***	0.032595***
	(0.010627)	(0.010498)	(0.010530)	(0.010515)	(0.010504)	(0.010502)	(0.010514)	(0.010579)
Unverified NOT– P(Y = 3)	0.022723***	0.025948***	0.025602***	0.025715***	0.025898***	0.026060***	0.025784***	0.025737***
	(0.005493)	(0.006540)	(0.006595)	(0.006562)	(0.006556)	(0.006540)	(0.006567)	(0.006654)
Unverified NOT– P(Y = 4)	0.007708***	0.010636***	0.010515***	0.010523***	0.010640***	0.010709***	0.010577***	0.010676***
	(0.001958)	(0.001845)	(0.001840)	(0.001838)	(0.001846)	(0.001850)	(0.001843)	(0.001861)
Unverified NOT– P(Y = 5)	-0.176411***	-0.171995***	-0.169496***	-0.170473***	-0.171529***	-0.172637***	-0.170815***	-0.169528***
	(0.063790)	(0.058442)	(0.058510)	(0.058475)	(0.058445)	(0.058449)	(0.058489)	(0.058573)
Verified OTNOTA– P(Y = 1)	0.032105***	0.027145***	0.026577***	0.026619***	0.027163***	0.027365***	0.026886***	0.027072***
	(0.009068)	(0.007769)	(0.007791)	(0.007788)	(0.007769)	(0.007757)	(0.007782)	(0.007771)
Verified OTNOTA– P(Y = 2)	0.011790***	0.010364***	0.010134***	0.010155***	0.010369***	0.010458***	0.010257***	0.010366***
	(0.003618)	(0.003144)	(0.003144)	(0.003145)	(0.003143)	(0.003143)	(0.003145)	(0.003153)
Verified OTNOTA– P(Y = 3)	0.009402***	0.009392***	0.009177***	0.009196***	0.009396***	0.009480***	0.009293***	0.009393***
	(0.003054)	(0.002978)	(0.002974)	(0.002975)	(0.002978)	(0.002980)	(0.002977)	(0.002987)
Verified OTNOTA– P(Y = 4)	0.004762***	0.005311***	0.005182***	0.005191***	0.005315***	0.005371***	0.005251***	0.005322***
	(0.001725)	(0.001835)	(0.001827)	(0.001827)	(0.001835)	(0.001840)	(0.001832)	(0.001843)
Verified OTNOTA– P(Y = 5)	-0.058060***	-0.052211***	-0.051070***	-0.051160***	-0.052244***	-0.052674***	-0.051686***	-0.052152***
	(0.017422)	(0.015690)	(0.015701)	(0.015700)	(0.015689)	(0.015682)	(0.015700)	(0.015717)
Unverified OTNOTA– P(Y = 1)	0.107895***	0.085562***	0.084335***	0.084759***	0.085495***	0.085928***	0.084897***	0.084909***
	(0.012379)	(0.010445)	(0.010454)	(0.010455)	(0.010444)	(0.010437)	(0.010458)	(0.010443)
Unverified OTNOTA– P(Y = 2)	0.032364***	0.028621***	0.028218***	0.028349***	0.028600***	0.028766***	0.028403***	0.028524***
	(0.004103)	(0.003717)	(0.003719)	(0.003719)	(0.003717)	(0.003718)	(0.003720)	(0.003731)
Unverified OTNOTA– P(Y = 3)	0.022255***	0.023264***	0.022944***	0.023032***	0.023250***	0.023385***	0.023096***	0.023209***
	(0.003180)	(0.003230)	(0.003229)	(0.003228)	(0.003230)	(0.003231)	(0.003232)	(0.003242)
Unverified OTNOTA– P(Y = 4)	0.007787***	0.010378***	0.010238***	0.010256***	0.010380***	0.010452***	0.010308***	0.010403***
	(0.001724)	(0.001846)	(0.001839)	(0.001839)	(0.001847)	(0.001851)	(0.001843)	(0.001855)
Unverified OTNOTA– P(Y = 5)	-0.170301***	-0.147824***	-0.145734***	-0.146396***	-0.147725***	-0.148532***	-0.146705***	-0.147044***
	(0.020457)	(0.018644)	(0.018661)	(0.018655)	(0.018643)	(0.018637)	(0.018667)	(0.018683)
Verified OTA– P(Y = 1)	0.038387**	0.034303**	0.033593**	0.033776**	0.034290**	0.034497**	0.033942**	0.034109**
	(0.017902)	(0.015320)	(0.015318)	(0.015321)	(0.015320)	(0.015306)	(0.015322)	(0.015293)
Verified OTA– P(Y = 2)	0.013856**	0.012885**	0.012606**	0.012677**	0.012879**	0.012972**	0.012742**	0.012853**
	(0.006164)	(0.005552)	(0.005551)	(0.005552)	(0.005551)	(0.005551)	(0.005552)	(0.005562)
Verified OTA– P(Y = 3)	0.010917**	0.011524**	0.011269**	0.011329**	0.011519**	0.011605**	0.011396**	0.011497**
	(0.004724)	(0.004836)	(0.004836)	(0.004835)	(0.004836)	(0.004837)	(0.004838)	(0.004847)
Verified OTA– P(Y = 4)	0.005395**	0.006351**	0.006204**	0.006232**	0.006351**	0.006409**	0.006277**	0.006351**
	(0.002253)	(0.002566)	(0.002564)	(0.002561)	(0.002567)	(0.002571)	(0.002566)	(0.002578)
Verified OTA– P(Y = 5)	-0.068555**	-0.065063**	-0.063672**	-0.064014**	-0.065039**	-0.065482**	-0.064358**	-0.064810**
	(0.030859)	(0.028150)	(0.028150)	(0.028148)	(0.028150)	(0.028139)	(0.028157)	(0.028157)
Unverified OTA– P(Y = 1)	0.034441	0.039114	0.039727*	0.040496*	0.038767	0.039598*	0.039035	0.039905*
	(0.026695)	(0.023879)	(0.024007)	(0.024073)	(0.023851)	(0.023893)	(0.023905)	(0.023938)
Unverified OTA– P(Y = 2)	0.012567	0.014532*	0.014702*	0.014969*	0.014413*	0.014718*	0.014486*	0.014840*
	(0.009053)	(0.008268)	(0.008269)	(0.008270)	(0.008269)	(0.008269)	(0.008268)	(0.008284)
Unverified OTA– P(Y = 3)	0.009977	0.012882*	0.012995*	0.013213*	0.012785*	0.013045*	0.012834*	0.013134*
	(0.006823)	(0.006912)	(0.006885)	(0.006867)	(0.006922)	(0.006907)	(0.006909)	(0.006905)
Unverified OTA– P(Y = 4)	0.005008	0.006975**	0.006993**	0.007088**	0.006934**	0.007070**	0.006938**	0.007101**
	(0.003095)	(0.003338)	(0.003294)	(0.003267)	(0.003352)	(0.003334)	(0.003331)	(0.003321)
Unverified OTA– P(Y = 5)	-0.061991	-0.073503*	-0.074417*	-0.075767*	-0.072899*	-0.074431*	-0.073292*	-0.074981*
	(0.045540)	(0.042261)	(0.042315)	(0.042331)	(0.042260)	(0.042263)	(0.042278)	(0.042306)
Length of review title– P(Y = 1)	0.001016***	0.000821***	0.000818***	0.000819***	0.000820***	0.000812***	0.000821***	0.000819***
	(0.000131)	(0.000111)	(0.000111)	(0.000110)	(0.000111)	(0.000111)	(0.000111)	(0.000110)
Length of review title– P(Y = 2)	0.000325***	0.000285***	0.000284***	0.000285***	0.000285***	0.000283***	0.000285***	0.000286***
	(0.000042)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)	(0.000039)
Length of review title– P(Y = 3)	0.000236***	0.000242***	0.000241***	0.000241***	0.000241***	0.000239***	0.000242***	0.000242***
	(0.000031)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)	(0.000033)
Length of review title– P(Y = 4)	0.000097***	0.000118***	0.000118***	0.000118***	0.000118***	0.000117***	0.000118***	0.000119***
	(0.000013)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000016)	(0.000017)
Length of review title– P(Y = 5)	-0.001673***	-0.001466***	-0.001461***	-0.001463***	-0.001465***	-0.001451***	-0.001466***	-0.001465***
	(0.000215)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)
Length of review body– P(Y = 1)	0.000069***	0.000040***	0.000040***	0.000040***	0.000040***	0.000039***	0.000040***	0.000040***
	(0.000007)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)	(0.000006)
Length of review body– P(Y = 2)	0.000022***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***	0.000014***
	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)
Length of review body– P(Y = 3)	0.000016***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***	0.000012***
	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)	(0.000002)
Length of review body– P(Y = 4)	0.000007***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***	0.000006***
	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)
Length of review body– P(Y = 5)	-0.000113***	-0.000071***	-0.000071***	-0.000071***	-0.000071***	-0.000070***	-0.000071***	-0.000071***
	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)	(0.000011)
Crowd's length of review title– P(Y = 1)	0.000981	0.000157	-0.000028	-0.000027	0.000160	-0.000022	0.000122	-0.000044
	(0.000617)	(0.000526)	(0.000529)	(0.000530)	(0.000526)	(0.000529)	(0.000527)	(0.000543)
Crowd's length of review title– P(Y = 2)	0.000313	0.000055	-0.000010	-0.000009	0.000056	-0.000008	0.000042	-0.000015
	(0.000197)	(0.000183)	(0.000184)	(0.000184)	(0.000183)	(0.000184)	(0.000183)	(0.000189)
Crowd's length of review title– P(Y = 3)	0.000228	0.000046	-0.000008	-0.000008	0.000047	-0.000007	0.000036	-0.000013
	(0.000143)	(0.000155)	(0.000156)	(0.000156)	(0.000155)	(0.000156)	(0.000155)	(0.000160)
Crowd's length of review title– P(Y = 4)	0.000093	0.000023	-0.000004	-0.000004	0.000023	-0.000003	0.000018	-0.000006
	(0.000059)	(0.000076)	(0.000076)	(0.000076)	(0.000076)	(0.000076)	(0.000076)	(0.000079)
Crowd's length of review title– P(Y = 5)	-0.001615	-0.000281	0.000050	0.000048	-0.000286	0.000040	-0.000218	0.000078
	(0.001016)	(0.000940)	(0.000945)	(0.000946)	(0.000940)	(0.000945)	(0.000941)	(0.000971)
Crowd's length of review body– P(Y = 1)	0.000037	0.000028	0.000022	0.000019	0.000029	0.000021	0.000028	0.000020
	(0.000027)	(0.000023)	(0.000023)	(0.000023)	(0.000023)	(0.000023)	(0.000023)	(0.000023)
Crowd's length of review body– P(Y = 2)	0.000012	0.000010	0.000008	0.000007	0.000010	0.000007	0.000010	0.000007
	(0.000009)	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)	(0.000008)
Crowd's length of review body– P(Y = 3)	0.000009	0.000008	0.000006	0.000006	0.000008	0.000006	0.000008	0.000006
	(0.000006)	(0.000007)	(0.000007)	(0.000007)	(0.000007)	(0.000007)	(0.000007)	(0.000007)
Crowd's length of review body– P(Y = 4)	0.000004	0.000004	0.000003	0.000003	0.000004	0.000003	0.000004	0.000003
	(0.000003)	(0.000003)	(0.000003)	(0.000003)	(0.000003)	(0.000003)	(0.000003)	(0.000003)
Crowd's length of review body– P(Y = 5)	-0.000062	-0.000051	-0.000039	-0.000035	-0.000052	-0.000037	-0.000051	-0.000036
	(0.000044)	(0.000041)	(0.000041)	(0.000041)	(0.000041)	(0.000041)	(0.000041)	(0.000042)
2017 – P(Y = 1)	0.048842***	0.028894**	0.023414*	0.025031**	0.028479**	0.024539**	0.027122**	0.022032*
	(0.015074)	(0.012356)	(0.012291)	(0.012307)	(0.012350)	(0.012222)	(0.012373)	(0.012263)
2017– P(Y = 2)	0.016400***	0.010372**	0.008452**	0.009010**	0.010232**	0.008898**	0.009749**	0.007983*
	(0.004515)	(0.004197)	(0.004245)	(0.004225)	(0.004202)	(0.004229)	(0.004224)	(0.004263)
2017– P(Y = 3)	0.012383***	0.009013***	0.007375**	0.007842**	0.008897**	0.007788**	0.008482**	0.006969*

	(0.003138)	(0.003486)	(0.003574)	(0.003539)	(0.003495)	(0.003564)	(0.003524)	(0.003598)
2017– P(Y = 4)	0.005584***	0.004665***	0.003850**	0.004071**	0.004614***	0.004097**	0.004400***	0.003649**
	(0.001191)	(0.001644)	(0.001733)	(0.001696)	(0.001654)	(0.001735)	(0.001677)	(0.001758)
2017– P(Y = 5)	-0.083210***	-0.052945**	-0.043091**	-0.045955**	-0.052222**	-0.045322**	-0.049752**	-0.040633*
	(0.023731)	(0.021626)	(0.021806)	(0.021725)	(0.021645)	(0.021709)	(0.021748)	(0.021849)
2018– P(Y = 1)	0.061490***	0.042325***	0.038845***	0.039779***	0.042008***	0.040754***	0.041100***	0.038070***
	(0.009039)	(0.007596)	(0.007640)	(0.007626)	(0.007602)	(0.007578)	(0.007641)	(0.007656)
2018– P(Y = 2)	0.019962***	0.014740***	0.013541***	0.013849***	0.014639***	0.014247***	0.014315***	0.013302***
	(0.002748)	(0.002551)	(0.002578)	(0.002567)	(0.002554)	(0.002558)	(0.002570)	(0.002590)
2018– P(Y = 3)	0.014701***	0.012483***	0.011470***	0.011717***	0.012404***	0.012087***	0.012124***	0.011259***
	(0.001957)	(0.002100)	(0.002128)	(0.002116)	(0.002104)	(0.002111)	(0.002118)	(0.002137)
2018– P(Y = 4)	0.006243***	0.006103***	0.005608***	0.005712***	0.006074***	0.005941***	0.005926***	0.005508***
	(0.000858)	(0.001006)	(0.001017)	(0.001010)	(0.001009)	(0.001015)	(0.001014)	(0.001020)
2018– P(Y = 5)	-0.102396***	-0.075651***	-0.069464***	-0.071056***	-0.075126***	-0.073029***	-0.073465***	-0.068139***
	(0.014219)	(0.013077)	(0.013218)	(0.013165)	(0.013096)	(0.013100)	(0.013179)	(0.013262)
2020– P(Y = 1)	0.018840***	0.009034	0.008149	0.008129	0.009049	0.009394*	0.008687	0.008053
	(0.006440)	(0.005604)	(0.005644)	(0.005639)	(0.005604)	(0.005612)	(0.005622)	(0.005649)
2020– P(Y = 2)	0.006863***	0.003392	0.003045	0.003040	0.003398	0.003525*	0.003256	0.003012
	(0.002367)	(0.002113)	(0.002117)	(0.002117)	(0.002113)	(0.002115)	(0.002116)	(0.002121)
2020– P(Y = 3)	0.005495***	0.003061	0.002736	0.002733	0.003065	0.003175*	0.002934	0.002701
	(0.001904)	(0.001910)	(0.001905)	(0.001906)	(0.001909)	(0.001908)	(0.001909)	(0.001905)
2020– P(Y = 4)	0.002812***	0.001713	0.001517	0.001517	0.001715	0.001772*	0.001636	0.001495
	(0.000989)	(0.001073)	(0.001061)	(0.001062)	(0.001073)	(0.001070)	(0.001070)	(0.001058)
2020– P(Y = 5)	-0.034011***	-0.017199	-0.015447	-0.015419	-0.017228	-0.017866*	-0.016513	-0.015261
	(0.011669)	(0.010691)	(0.010720)	(0.010719)	(0.010691)	(0.010698)	(0.010710)	(0.010727)
2021– P(Y = 1)	0.048087***	0.029251***	0.027039***	0.027121***	0.029322***	0.029268***	0.028421***	0.026560***
	(0.007646)	(0.006528)	(0.006577)	(0.006572)	(0.006530)	(0.006532)	(0.006563)	(0.006568)
2021– P(Y = 2)	0.016179***	0.010492***	0.009680***	0.009716***	0.010515***	0.010500***	0.010186***	0.009525***
	(0.002594)	(0.002359)	(0.002371)	(0.002371)	(0.002359)	(0.002360)	(0.002369)	(0.002372)
2021– P(Y = 3)	0.012234***	0.009111***	0.008389***	0.008423***	0.009129***	0.009107***	0.008840***	0.008244***
	(0.001970)	(0.002045)	(0.002053)	(0.002054)	(0.002045)	(0.002044)	(0.002054)	(0.002050)
2021– P(Y = 4)	0.005536***	0.004709***	0.004314***	0.004335***	0.004717***	0.004700***	0.004561***	0.004237***
	(0.000933)	(0.001069)	(0.001066)	(0.001067)	(0.001068)	(0.001066)	(0.001071)	(0.001064)
2021– P(Y = 5)	-0.082037***	-0.053562***	-0.049422***	-0.049595***	-0.053684***	-0.053575***	-0.052007***	-0.048565***
	(0.012971)	(0.011932)	(0.012008)	(0.012005)	(0.011933)	(0.011934)	(0.011992)	(0.011997)
2022– P(Y = 1)	0.050876***	0.034936***	0.032658***	0.032659***	0.035056***	0.034875***	0.034109***	0.030992***
	(0.009339)	(0.007946)	(0.007982)	(0.007977)	(0.007951)	(0.007949)	(0.007973)	(0.007958)
2022– P(Y = 2)	0.016990***	0.012372***	0.011544***	0.011554***	0.012410***	0.012354***	0.012068***	0.011004***
	(0.003011)	(0.002766)	(0.002778)	(0.002779)	(0.002766)	(0.002768)	(0.002775)	(0.002784)
2022– P(Y = 3)	0.012777***	0.010627***	0.009897***	0.009911***	0.010656***	0.010601***	0.010360***	0.009442***
	(0.002211)	(0.002331)	(0.002340)	(0.002342)	(0.002330)	(0.002330)	(0.002339)	(0.002349)
2022– P(Y = 4)	0.005709***	0.005364***	0.004971***	0.004983***	0.005377***	0.005344***	0.005220***	0.004763***
	(0.000978)	(0.001150)	(0.001150)	(0.001152)	(0.001149)	(0.001147)	(0.001152)	(0.001158)
2022– P(Y = 5)	-0.086352***	-0.063298***	-0.059070***	-0.059108***	-0.063499***	-0.063174***	-0.061757***	-0.056202***
	(0.015337)	(0.014096)	(0.014165)	(0.014166)	(0.014099)	(0.014097)	(0.014148)	(0.014173)
2023– P(Y = 1)	0.032421**	0.026172*	0.025729*	0.024746*	0.027009*	0.028257**	0.025696*	0.025614*
	(0.015944)	(0.013773)	(0.013857)	(0.013795)	(0.013871)	(0.013912)	(0.013792)	(0.013886)
2023– P(Y = 2)	0.011379**	0.009453**	0.009239*	0.008913*	0.009736**	0.010160**	0.009266*	0.009206*
	(0.005221)	(0.004756)	(0.004761)	(0.004763)	(0.004772)	(0.004764)	(0.004759)	(0.004775)
2023– P(Y = 3)	0.008872**	0.008257**	0.008026**	0.007762*	0.008490**	0.008830**	0.008084**	0.007982**
	(0.003864)	(0.003990)	(0.003974)	(0.003993)	(0.003991)	(0.003962)	(0.003991)	(0.003978)
2023– P(Y = 4)	0.004298**	0.004322**	0.004151**	0.004034**	0.004429**	0.004576**	0.004218**	0.004119**
	(0.001673)	(0.001916)	(0.001885)	(0.001912)	(0.001903)	(0.001867)	(0.001913)	(0.001883)
2023– P(Y = 5)	-0.056971**	-0.048204**	-0.047145*	-0.045455*	-0.049665**	-0.051822**	-0.047264*	-0.046920*
	(0.026637)	(0.024395)	(0.024438)	(0.024427)	(0.024494)	(0.024458)	(0.024417)	(0.024483)
Negative price sentiment– P(Y = 1)		0.366525***	0.365663***	0.365563***	0.366611***	0.365891***	0.366342***	0.365410***
		(0.009341)	(0.009340)	(0.009341)	(0.009341)	(0.009339)	(0.009341)	(0.009354)
Negative price sentiment– P(Y = 2)		0.067186***	0.067200***	0.067204***	0.067185***	0.067242***	0.067179***	0.067386***
		(0.002266)	(0.002265)	(0.002265)	(0.002266)	(0.002266)	(0.002266)	(0.002272)
Negative price sentiment– P(Y = 3)		0.015111***	0.015233***	0.015211***	0.015112***	0.015218***	0.015144***	0.015323***
		(0.001872)	(0.001868)	(0.001868)	(0.001872)	(0.001869)	(0.001871)	(0.001871)
Negative price sentiment– P(Y = 4)		-0.030873***	-0.030734***	-0.030742***	-0.030873***	-0.030745***	-0.030843***	-0.030709***
		(0.002086)	(0.002084)	(0.002083)	(0.002086)	(0.002084)	(0.002086)	(0.002089)
Negative price sentiment– P(Y = 5)		-0.417949***	-0.417363***	-0.417235***	-0.418036***	-0.417605***	-0.417822***	-0.417410***
		(0.007119)	(0.007128)	(0.007131)	(0.007119)	(0.007125)	(0.007121)	(0.007139)
Positive price sentiment– P(Y = 1)		-0.075611***	-0.075717***	-0.075738***	-0.075587***	-0.075505***	-0.075667***	-0.075527***
		(0.003017)	(0.003017)	(0.003017)	(0.003018)	(0.003021)	(0.003017)	(0.003015)
Positive price sentiment– P(Y = 2)		-0.053100***	-0.053151***	-0.053194***	-0.053073***	-0.052997***	-0.053128***	-0.053150***
		(0.002606)	(0.002605)	(0.002606)	(0.002606)	(0.002608)	(0.002605)	(0.002609)
Positive price sentiment– P(Y = 3)		-0.059780***	-0.059846***	-0.059884***	-0.059747***	-0.059607***	-0.059832***	-0.059799***
		(0.003298)	(0.003297)	(0.003298)	(0.003298)	(0.003298)	(0.003298)	(0.003299)
Positive price sentiment– P(Y = 4)		-0.044778***	-0.044841***	-0.044843***	-0.044760***	-0.044630***	-0.044833***	-0.044824***
		(0.002966)	(0.002965)	(0.002966)	(0.002966)	(0.002964)	(0.002967)	(0.002968)
Positive price sentiment– P(Y = 5)		0.233268***	0.233555***	0.233659***	0.233166***	0.232739***	0.233460***	0.233299***
		(0.010529)	(0.010522)	(0.010523)	(0.010530)	(0.010539)	(0.010525)	(0.010529)
Crowd's average rating– P(Y = 1)			-0.015256***					
			(0.004649)					
Crowd's average rating– P(Y = 2)			-0.005307***					
			(0.001621)					
Crowd's average rating– P(Y = 3)			-0.004494***					
			(0.001373)					
Crowd's average rating– P(Y = 4)			-0.002199***					
			(0.000676)					
Crowd's average rating– P(Y = 5)			0.027255***					
			(0.008298)					
Crowd's share of 5-star ratings– P(Y = 1)				-0.047927***				
				(0.015166)				
Crowd's share of 5-star ratings– P(Y = 2)				-0.016676***				
				(0.005289)				
Crowd's share of 5-star ratings– P(Y = 3)				-0.014117***				
				(0.004478)				
Crowd's share of 5-star ratings– P(Y = 4)				-0.006902***				
				(0.002200)				
Crowd's share of 5-star ratings– P(Y = 5)				0.085623***				
				(0.027073)				
Crowd's share of 4-star ratings– P(Y = 1)					-0.020151			
					(0.028205)			
Crowd's share of 4-star ratings– P(Y = 2)					-0.007006			
					(0.009806)			
Crowd's share of 4-star ratings– P(Y = 3)					-0.005930			

					(0.008299)			
Crowd's share of 4-star ratings– P(Y = 4)					-0.002900			
					(0.004060)			
Crowd's share of 4-star ratings– P(Y = 5)					0.035988			
					(0.050364)			
Crowd's share of 2-star ratings– P(Y = 1)						0.086900***		
						(0.025511)		
Crowd's share of 2-star ratings– P(Y = 2)						0.030237***		
						(0.008901)		
Crowd's share of 2-star ratings– P(Y = 3)						0.025588***		
						(0.007532)		
Crowd's share of 2-star ratings– P(Y = 4)						0.012522***		
						(0.003712)		
Crowd's share of 2-star ratings– P(Y = 5)						-0.155246***		
						(0.045539)		
Crowd's share of 1-star ratings– P(Y = 1)							0.027812	
							(0.020676)	
Crowd's share of 1-star ratings– P(Y = 2)							0.009671	
							(0.007192)	
Crowd's share of 1-star ratings– P(Y = 3)							0.008188	
							(0.006091)	
Crowd's share of 1-star ratings– P(Y = 4)							0.004004	
							(0.002982)	
Crowd's share of 1-star ratings– P(Y = 5)							-0.049676	
							(0.036927)	
Crowd's positive imbalance– P(Y = 1)								-0.057286***
								(0.015200)
Crowd's positive imbalance– P(Y = 2)								-0.019984***
								(0.005317)
Crowd's positive imbalance– P(Y = 3)								-0.016916***
								(0.004503)
Crowd's positive imbalance– P(Y = 4)								-0.008291***
								(0.002226)
Crowd's positive imbalance– P(Y = 5)								0.102477***
								(0.027161)
N	14,565	14,565	14,565	14,565	14,565	14,565	14,565	14,526

* Note: *p < .10; **p < .05; **p < .01. Days of the week dummies, month dummies, and store dummies are excluded for space. Perfectly predicted samples are dropped.