Stand Out from the Millions: Information Friction and Market Congestion on Global E-Commerce Platforms*

Jie Bai
Harvard Kennedy School&NBER

Maggie X. Chen
George Washington University

Jin Liu
New York University

Xiaosheng Mu
Princeton University

Daniel Yi Xu
Duke University&NBER

February 21, 2024

Abstract

We study how information friction affects firm dynamics and market efficiency in global e-commerce. Combining detailed transaction-level platform data with objective quality measures, we first document significant information friction in the online market. We then conduct a randomized experiment that generates exogenous demand and information shocks to help small exporters overcome the information friction and grow. We show experimentally and theoretically that the effectiveness of such demand intervention is undermined by the large number of market participants. We build and estimate an empirical model of the online market to quantify the efficiency implications of and the interaction between information friction and market congestion. Finally, we apply the model to examine policy counterfactuals aimed at promoting SME growth online. Our results highlight that blanket-wide onboarding initiatives can aggravate market congestion, slow down the resolution of the information problem, and hinder the discovery of high-quality firms.

Keywords: global e-commerce, exporter dynamics, quality, information friction, and congestion

JEL classification: F14, L11, O12

*We thank Costas Akolakis, Treb Allen, David Atkin, Christopher Avery, Oriana Bandiera, Abhijit Banerjee, Heski Bar-Isaac, Lauren Bergquist, Davin Chor, Michael Dinerstein, Dave Donaldson, Esther Duflo, Ben Faber, Gordon Hanson, Chang-tai Hsieh, Panle Jia, Ginger Jin, Amit Khandelwal, Asim Khwaja, Pete Klenow, Danielle Li, Rocco Macchiavello, Isabela Manelici, Nicholas Ryan, Meredith Startz, Tianshu Sun, Christopher Snyder, Catherine Thomas, Eric Verhoogen, David Weinstein, Richard Zeckhauser, and seminar and conference participants at Asian Pacific IO Conference, Berkeley Economics, Birmingham University, BREAD/CEPR/STICERD/TCD Conference, Fudan, Harvard Kennedy School, Hong Kong Trade Workshop, IADB, Imperial College London, IPA SME Working Group Meeting, Michigan Economics, Microsoft Research, Northwestern University, NBER Conferences (China Working Group, DEV, ITI, IT/Digitization), Oxford, Singapore Management University, Stanford, Syracuse, University of Chicago, University of Oregon, USC, World Bank, and Yale for helpful comments. We thank Zixu Chen, Chengdai Huang, Haoran Zhang, and Qiang Zheng for excellent research assistance. All errors are our own.

1 Introduction

E-commerce sales have grown tremendously in recent years, reaching $5.8 trillion and 19.5% of total global retail sales in 2023. Within e-commerce, cross-border sales have grown twice as fast as domestic sales, and nearly 70% of online buyers completed a cross-border transaction in 2020.[1] By extending market access beyond geographical boundaries, global e-commerce platforms present a promising avenue for small and medium-sized enterprises (SMEs) in developing countries to enter export markets. Furthermore, online exporting lowers many of the traditional barriers of exporting, including the need to build export relationships and set up distributional channels in destination countries[2]. Given these promises and the large market potential, numerous policy initiatives have been adopted worldwide to foster e-commerce growth (e.g, UNCTAD, 2016, 2021), with a specific target to onboard developing-country SMEs onto e-commerce platforms and allow them to tap into the global market.[3]

Despite the widespread adoption of e-commerce onboarding initiatives worldwide, there is limited empirical evidence regarding the growth outcomes for SMEs in e-commerce *beyond the initial entry point*. While e-commerce offers the potential for SMEs to connect with buyers globally, significant challenges persist in the online marketplace that may impede the growth of SMEs. Importantly, information friction is prevalent in the online market (Tadelis, 2016). Consumers often struggle to accurately assess the quality of sellers and must rely on online reviews as noisy signals. This information problem is likely more pronounced for cross-border transactions. How does information friction affect the growth dynamics of SMEs following their initial onboarding?

To answer this question, we begin by documenting descriptive evidence for the presence of significant information friction in the online market, leveraging detailed platform data and novel objective quality measures. We then evaluate a demand intervention that generates exogenous demand and information shocks to help newly-onboarded SMEs to overcome the information friction and grow. We show experimentally and theoretically that the effectiveness of such

[1]Worldwide E-commerce Forecast 2023, eMarketer.

[2]For example, AliExpress, a leading cross-border e-commerce platform that we study in this project, states on its website (https://sell.aliexpress.com/), "Set up your e-commerce store in a flash, it's easy and free! Millions of shoppers are waiting to visit your store!"

[3]Examples of such initiatives include the Multichannel E-Commerce Platform Program in Singapore (subsidized training and connection), the e-Smart IKM program and Export program with Alibaba in Indonesia, the Global export program with Amazon in Vietnam (government-sponsored training), the e-Commerce Accelerator Program in Australia (financial assistance), and the Pan-African e-Commerce Initiative in Ghana, Kenya and Rwanda (training and setting up new platforms). Almost all of the existing programs focus on the initial onboarding and market entry of the SMEs.

demand intervention is undermined by the large number of market participants, which leads to greater market congestion and makes it harder for high-quality newcomers to accumulate sales and reviews. Next, we build and estimate an empirical model of the online market to quantify the efficiency implications of—and the interaction between—information friction and market congestion. Finally, we apply the model to examine policy counterfactuals aimed at promoting SME growth online. Our results highlight that blanket-wide onboarding initiatives may not be able to generate sustained SME growth, but may instead aggravate market congestion, slow down the resolution of the information problem, and hinder the discovery of high-quality firms. The results put a cautionary tale against existing onboarding programs and suggests alternative policy designs to facilitate the growth of high-quality SME exporters.

Our study is grounded in the context of AliExpress, a world-leading B2C cross-border e-commerce platform owned by Alibaba. We focus on the segment of children's T-shirts, one of the top-selling product categories on the platform, and collect comprehensive data about sellers' detailed product-level characteristics and transaction-level sales records. We complement the platform data with a novel set of objective, multidimensional measures of quality, ranging from detailed product quality metrics to shipping and service quality indicators. These measures are collected by the research team through actual online purchases and direct interactions with the sellers as well as third-party assessments.

Using the newly assembled micro dataset, we document a set of stylized facts about the global online marketplace. First, we show that online ratings are highly noisy and inflated. At the same time, considerable variation in underlying true quality exists, even among products rated at the top, suggesting that a significant amount of information friction remains unresolved in the online market. Second, we show that quality only weakly predicts sales. The "superstars," which we define as the largest seller in each identical-looking product variety, do not necessarily command higher quality compared to the small listings in the same variety group, even when prices are broadly similar. This provides suggestive evidence for the presence of potential market misallocation resulting from information friction. Finally, we examine the growth dynamics of sellers. Transaction-level data reveals that higher past sales predict faster arrival of future sales. This is, in part, driven by existing online ranking and review mechanisms that enable sellers to enhance their visibility and popularity by accumulating sales and (positive) reviews. Intuitively, this process could help high-quality sellers to overcome the information friction, allowing them to stand out over time by accumulating initial sales and reviews on the platform.

The descriptive findings motivate us to conduct a demand intervention to help online SMEs to overcome the information friction and grow. In particular, the experiment generates exogenous demand and information shocks to a set of small exporters via randomly placed online purchase orders and reviews. Consistent with the descriptive evidence, the order treatment does lead to a significantly positive impact on sellers' subsequent sales. However, the magnitude of the estimated average treatment effect is much smaller than the size of the initial purchase. We also do not find any significant treatment effect from the reviews nor any heterogeneous treatment effect based on quality. Interestingly, when we resold the t-shirts offline through a third-party consignment store, we find that quality, observable at the point of transaction in the offline setting, does translate into higher prices and higher sales probability. Taken together, the experimental findings suggest that a genuine demand for quality exists among buyers, but information friction could hamper the growth of high-quality sellers in the online market. While a demand-side intervention, working through existing platform mechanisms, helps to generate sales, its success is rather limited in helping high-quality sellers to overcome the information friction and grow.

Next, we develop a theoretical model to interpret the reduced-form findings and use the model to examine the efficiency properties of the online market. The model incorporates heterogeneity in seller quality and information friction, as well as a process of consumer search and learning through existing online ranking and review mechanisms. Specifically, we extend the classical Polya urn model by incorporating consumer choice and allow the probability of a seller entering into a consumer's consideration set to increase with the seller's cumulative sales.[4] Among sellers in the consideration set, consumers make purchase decisions based on their expected qualities inferred from past reviews. We prove theoretically that having a large number of sellers slows down the resolution of the information problem. Importantly, connecting back to the policy motivation, we show that in a market with a large number of existing sellers, further increasing the number of sellers (for example, through large-scale e-commerce onboarding programs) exacerbates market congestion, making it harder for high-quality sellers to accumulate sales and reviews. This, in turn, hinders the discovery of high-quality sellers and strictly worsens allocation along the path of market evolution.

Finally, building on the theoretical insights, we estimate an empirical model of the online market. We enrich the setup of the theoretical model by incorporating seller heterogeneity in

[4]Following Goeree (2008), we consider the formation of consumer's consideration set to be a "reduced-form" representation of underlying consumer search behavior. Given a fixed size of the consideration set, increasing the number of sellers means that consumers would consider a smaller fraction of the market, which we formally define as "market congestion".

both quality and cost as well as modeling sellers' pricing decisions. Our model estimates imply considerable information friction on AliExpress. Improving the informational signal of reviews leads to a sizable improvement in market allocation and consumer surplus. However, with the large noise in review signals, uncertainty regarding quality resolves very slowly, i.e., only after a seller accumulates a substantial number of orders. Given the amount of market congestion, it takes time for high-quality sellers to accumulate sales and stand out. On the other hand, reducing the number of sellers helps to allocate market share towards high-quality sellers and yields welfare gain despite the loss in product variety, consistent with the theoretical insights. In sum, the analyses demonstrate that information friction, combined with high market congestion, can constitute an important hurdle for the growth of high-quality prospective exporters. Using the experiment as a model-validation exercise, we find quantitatively comparable average treatment effects when we simulate one-time demand shocks in the model.

We end with a model-based evaluation of potential government onboarding programs that aim to bring SMEs online and facilitate the growth of high-quality businesses through e-commerce. Most existing onboarding initiatives seek to onboard SMEs to existing large global e-commerce platforms such as AliExpress. We show that such initiatives may have limited success due to the large information friction and market congestion present in these existing marketplaces. An alternative approach, which becomes increasingly discussed among policy-makers, is to onboard SMEs onto newly created marketplaces, either new platforms or designated market segments on existing platforms. We consider such alternative interventions under different assumptions of consumer traffic. Consistent with the above-discussed interplay between information friction and congestion, we find that creating and promoting such a designated marketplace to host the newly onboarded SMEs can lead to better growth performance and allocative efficiency. The results highlight the policy trade-off between subsidizing SMEs to operate on large marketplaces versus allocating budgets to enhance the visibility of new marketplaces. Last but not least, our results suggest a promising avenue for future policy design aimed at injecting new information into markets. This could be achieved through government or third-party screening and certification, in conjunction with onboarding programs, to effectively identify and promote the growth of high-quality firms.

Related Literature. Our work contributes to several strands of the existing literature. By studying market frictions and firm dynamics in a marketplace with heterogeneous quality, our paper builds on a large literature that documents the important role of quality in determining firm performance and exports (see Verhoogen, 2020 for a recent review). Most of the prior works focus on offline settings with a few exceptions (e.g., Jin and Kato (2006)). We build on a growing body of research that collects detailed measures on quality for specific industries (e.g., Atkin et al., 2017; Macchiavello and Miquel-Florensa, 2019; Bai et al., 2019; Hansman et al., 2020). Leveraging the detailed objective quality measures, we establish large variations in firm-product quality in the online marketplace. However, we find that in contrast to the offline setting, quality plays a less pronounced role in explaining exporter growth and market share distribution in the global e-commerce market. Our paper highlights the role of information friction, in the presence of large market congestion, in slowing down the growth of high-quality sellers. We further model these realistic market frictions and quantify their impacts on market dynamics and efficiency through the lens of a rich empirical model.

Our paper also speaks to the existing literature on information friction in trade and development (Allen, 2014; Macchiavello and Morjaria, 2015; Steinwender, 2018; Startz, 2018) as well as in the online marketplace (Resnick et al., 2006; Hui et al., 2022; Li et al., 2020). We bring in novel data and new sources of variations to document the existence of information friction, theoretically examine its interaction with market congestion, and experimentally evaluate a demand intervention that can potentially help sellers overcome these frictions. While our study focuses on the children's t-shirt segment, the theoretical insight and empirical methodology can be applied to other e-commerce settings.

Relatedly, we contribute to a growing literature on demand-side interventions to facilitate the growth and upgrading of SMEs in developing countries (Bai, 2016; Bold et al., 2022). Our paper is closely related to Atkin et al. (2017), which also studies the impact of foreign demand shocks on exporters, showing that firms respond to these demand shocks by improving quality through learning by doing. In our study, we explore how foreign demand shocks help firms overcome information friction and congestion in the market.[5]

Finally, our study also relates to the existing literature on consumer consideration sets and search (for example, Goeree, 2008; Dinerstein et al., 2018; Ershov, 2022)[6]. Theoretically, we introduce a novel perspective by formalizing how market congestion can obstruct the discovery of high-quality sellers due to slow information resolution. Empirically, we quantify the impact of this mechanism on hindering the growth of high-quality firms and causing misallocation of market share, as supported by our experimental evidence.

[5] In a different setting, Pallais (2014) and Stanton and Thomas (2016) examine information friction in online labor markets and show that information generated from initial hires affects workers' subsequent hiring outcomes. In a similar vein, we show that initial demand generated from purchases affects the subsequent growth of sellers.

[6] We refer interested readers to a recent article by Honka et al. (2019) for a review of the broader literature.

The remainder of the paper is organized as follows. Section 2 describes the empirical setting and data. Section 3 presents a set of stylized facts about online exporters. Section 4 describes the experiment design and main findings. Section 5 develops the theoretical model and derives market efficiency properties. Section 6 builds and estimates an empirical model of the online market. Section 7 performs counterfactual analyses. Section 8 concludes.

2 Empirical Setting and Data

In this section, we introduce the empirical setting of our study—the market of children's T-shirts on AliExpress—and describe the data.

2.1 The Market for Children's T-shirts on AliExpress

AliExpress, a subsidiary of Alibaba, was founded in April 2010 to specialize in international trade. As a global leading platform for cross-border B2C trade, AliExpress serves over 150 million consumers from 220 countries and regions, attracting over 400 million monthly visits.[7] The platform hosts over 100 million products, ranging from clothes and shoes to electronics and home appliances, and 1.1 million active sellers, primarily retailers located in China.[8] Most sellers on the platform are retailers, rather than manufacturers, and source products from factories all over the country to export through the platform. Therefore, quality, in this context, captures sellers' sourcing ability (i.e., ability to source high-quality products from manufacturers) as well as the quality of their marketing and shipping services.[9]

For this study, we focus on the children's T-shirt segment. As the largest textile and garment exporting country in the world, China accounted for 41% of the world's total textile and garment exports in 2022 (WTO, 2023). In the world of e-commerce, textile and apparel amount to 22.6 percent of China's total online retail, including sales on Alibaba's platforms.[10] The growth and efficiency of the online retail market, therefore, matter for upstream manufacturing: in particular, growth of retailers that sell high-quality products in turn benefits their producers.

[7] Sources: https://sell.aliexpress.com/.

[8] During our sample period, AliExpress hosted sellers from mainland China only; starting in 2018, the platform also became available to sellers in Russia, Spain, Italy, Turkey, and France.

[9] While most of the sellers on the e-commerce platform are retailers instead of manufacturers, quality may still vary significantly depending on where the sellers choose to source from—whether high-quality or low-quality factories—and how much quality inspection effort they put in. We document this formally using detailed quality measures that we collect from the study in Section 2.2.

[10] "Online Retail Market Development Report 2022," Ministry of Commerce, China.

The vibrant entry and growth dynamics in the online market also provide an ideal setting for studying exporter dynamics. In addition, the T-shirt product category features well-specified quality dimensions, making it possible to construct *direct* quality measures to study quality-size distributions and allocative efficiency.

Two features of the platform are worth highlighting. First, AliExpress does not require a sign-up fee to set up a store and list a product, thereby essentially eliminating entry and fixed operation costs of exporting and allowing sellers large and small to tap into export markets.[11] While this helps bring many SMEs onto the platform, the low entry barrier can create important congestion on the platform, resulting in an excessive number of sellers and product offerings competing for consumers' attention in the online marketplace. The resulting efficiency implications of the increasing number of market participants are far less clear in the presence of information friction. We examine this key interaction and its implications on firm growth and market allocation in this study.

Second, AliExpress allows us to group product listings into different *varieties*.[12] A single *variety group* (hereafter referred to as a *group*) may contain multiple listings that are sold by different sellers but share an identical product design. This is illustrated in Figure 1. This unique feature allows us to compare listings with the same observable product attributes, thereby controlling for consumers' horizontal taste differences. We leverage this feature in our descriptive and reduced-form analyses below, and further take into account the empirical distribution of variety in estimating the structure model to account for potential gains from variety.

2.2 Data

We collect comprehensive data from the platform, including detailed firm-product-level characteristics and transaction-level sales records. We complement the platform data with objective quality measures obtained from actual purchases, direct interactions with sellers, and third-party assessment. Below, we describe the sample and the key variables used in the analyses.

(1) Store-Listing-Level Data. We scraped nearly the full universe of product listings in the children's T-shirt segment.[13] We collected all the information that a buyer can view on the

[11] AliExpress charges sellers 5-8% of their sales revenue as a commission fee for each successful transaction. Source: https://sell.aliexpress.com/.

[12] Unfortunately, this feature has been disabled since our study period and is currently no longer available to the public.

[13] The scraping was done at the variety group level. The platform allowed users to view the first 99 pages of variety groups with 48 groups per search page.

listings' pages, including total cumulative orders (quantity sold), current prices, discounts (if any), ratings, buyer protection schemes (if any), and detailed product attributes. We further collected information about the stores that carry these products, including the year of opening and other products that the stores carry.

Table 1 summarizes the product-listing-level (Panel A) and store-level (Panel B) characteristics. There are 10,089 product listings in total. The average price is $6.1. Approximately 54% of the listings offer free shipping, and the average price of shipping to the US is $0.63. At the store level, there are 1,291 stores carrying these products. Most exporters are young, with an average age of 1.61 years. The average cumulative sales is 235 with a standard deviation of 970, indicating large performance heterogeneity. We observe similar patterns of performance heterogeneity at the listing level. At a given point in time, more than 35% of the listings have zero sales; the median has 2, whereas the largest listing has 10,517 orders accumulated.

(2) Transaction Records. We take advantage of a unique feature of AliExpress during our sample period that allows us to keep track of a listing's most recent six-month transaction history. For each transaction, we observe information on sales quantities, ratings, and previous buyers' countries of origin. In contrast, most existing e-commerce platforms (e.g., Amazon and eBay) report only customer reviews and the total volume of transactions, not the full transaction history. The availability of real-time transaction records enables us to closely track each product listing's sales activities over time.

(3) Measures of Quality. Finally, we complement the platform data with a rich set of objective quality measures collected through (i) actual purchases of the products, (ii) direct communications with the sellers, and (iii) third-party assessment. We collect quality data for variety groups with at least 100 cumulative sales (aggregated across all listings in the group) to focus on products that are more relevant for consumer choice. This leaves us with 1,258 product listings sold by 636 stores in 133 variety groups, with varying performance heterogeneity (measured in terms of cumulative sales) within each group. Table A.1 summarizes the listing and store characteristics for this sample.

The quality measures cover multiple dimensions, ranging from product to service to shipping quality. To measure product quality, we placed actual orders for children's T-shirts on AliExpress.[14] After receiving and cataloging the orders, we worked with a large local consignment

[14]Measuring product and shipping quality involves actually purchasing the T-shirts. Therefore, we combined this data collection effort with the experiment in which we generated exogenous demand shocks to a randomly selected subset of treated small listings (with fewer than 5 cumulative orders) in the 133 variety groups. Hence, the sample for product quality consists of all treated small listings (with fewer than 5 cumulative orders) in

store of children's clothing in North Carolina to inspect and grade the quality of each T-shirt. The grading was done on a rich set of metrics, following standard criteria used in the textile and garment industry. Specifically, product quality was assessed along eight dimensions: durability, fabric softness, wrinkle test, seams (straightness and neatness), outside stray threads, inside loose stitches, pattern smoothness, and trendiness. Figure A.1 Panel A shows a picture of the grading process. Quality along each dimension was scored on a 1-to-5 scale, with higher numbers denoting higher quality. Most of the quality metrics (with the exception of trendiness) capture vertical quality differentiation. For example, at equal prices, consumers prefer T-shirts with more durable fabric, straighter seams, and fewer stray loose threads. Exploiting the grouping function, we can further compare quality across T-shirts of the exact same design but sold by different sellers. As shown in Panel B of Figure A.1, there exist considerable quality differences both across and within variety groups, depending on which factories the retailers choose to source from and/or how much quality inspection effort they put in.

To measure shipping quality, we recorded the date of each purchase, date of shipment, date of delivery, carrier name, and condition of the package upon arrival. The information is used to construct three measures of shipping quality: (i) the time lag between order and shipping, (ii) the time lag between shipping and delivery, and (iii) whether the package was damaged.

To measure service quality, we visited the homepage of each store and sent a message to the seller via the platform to inquire about a particular product.[15] We rate service quality based on the time it took to receive a reply, in particular, whether the message was replied to within two days (which represents the 70th percentile in reply time). Appendix B.1 provides more details of the quality measurement process.

Panels A and B in Table 2 present summary statistics of the various quality measures. For the empirical analysis, we construct different quality indices by first standardizing the quality measures in each dimension and then averaging them within and across the three dimensions. Panel C in Table 2 summarizes the distribution of the quality indices. Table A.2 decomposes the variation of the overall quality index into that explained by each quality metric.

To cross-validate these objective quality measures, we first examine the relationships be-

the experiment described in Section 4 and their medium-size (with cumulative orders between 6 and 50) and superstar (with the largest number of cumulative orders) peers in the same variety groups. This sampling procedure aimed to achieve two goals: first, it allowed us to obtain product and shipping quality measures for listings with different baseline sales to examine quality-sales relationships; second, it ensured that we have a control group of identical small listings not receiving any purchase order treatment.

[15] To measure service quality, we reached out to all 636 stores in the 133 variety groups. For those with multiple listings included in the 133 groups, we randomly selected one listing to inquire and assign the same service quality score to all listings sold by the same seller.

tween them and the online ratings and find all three quality indices—product, shipping and service—to be positively correlated with the online star ratings and—in the case of shipping and service quality—statistically significant, as shown in Table A.3. For product quality, we further asked the owner of the consignment store to report a bid price and a resell price for each T-shirt, reflecting consumers' willingness to pay. Reassuringly, the objective product quality metrics are strongly correlated with the third party's price evaluations. Last but not least, we show that the measured quality remains consistent for a seller over time. In particular, we sent multiple rounds of messages to the same stores and tracked sellers' replies. Table A.4 shows that a seller's reply speed is highly consistent over time.[16]

3 Stylized Facts about Online Exporters

Using the newly assembled micro dataset, we begin by documenting a set of stylized facts about online exporters. These facts provide suggestive evidence for the presence of information friction and potential market misallocation.

Fact 1. *Online ratings are highly skewed towards the top; however, large quality variation exists among top-rated products.*

We begin by examining how well existing online mechanisms help resolve information friction. Panel A of Figure 2 plots the distribution of the online star rating, the primary signal of quality in the online marketplace. The rating is based on consumer-generated reviews, and ranges from 1 (lowest) to 5 (highest). As shown in Panel A, the online star rating is highly skewed towards the top, with 43.8% of product listings having five stars and 80.9% scoring above 4.7; the median is 4.92. This finding is consistent with prior studies documenting that online reviews are highly inflated and may only serve as noisy signals of quality (Tadelis, 2016).

Panel B plots the distribution of quality for listings with an above-median star rating. We leverage the objective quality measures described in Section 2.2. Focusing on the overall quality index (by standardizing and averaging across detailed individual quality metrics), we

[16]We appreciate this suggestion made by various seminar and conference participants, which led us to revisit the platform in June 2021 and collect 3 rounds of service quality data following the exact same procedure for a new sample of 132 stores with active listings in children's T-shirts (in popular variety groups) at the time. Pooling data across all the 132 stores over 3 rounds, we estimate intracluster correlations as high as 0.5, 0.51, and 0.48 for the 3 quality measures examined in Table A.4, respectively. Regressing the reply behavior measured in the second and third rounds (stacked) on that in the first round yields positive coefficients of 0.614, 0.562, and 0.591, which are highly significant at the 1% level, as shown in Table A.4.

see large variation in quality among highly-rated products, indicating that a substantial amount of information friction remains unresolved in the online market.

Fact 2. *Superstars do not necessarily have the highest quality; quality only weakly predicts sales.*

Next, we exploit the grouping feature described in Section 2.1 that allowed us to group product listings into different identical-looking varieties (controlling for horizontal differences) to examine the relationship between sales performance and quality. On average, a group's superstar, defined as the listing with the highest number of cumulative orders within the group, accounts for about 63.8% of the total sales of the group, whereas the smallest listing only captures 0.27% of the group's market share.

Panel A of Figure 3 compares the quality of the superstars (with the highest sales in the group) and the small listings (with fewer than 5 cumulative orders) in each variety group. Plotting the distribution of the *difference* in the overall quality index between the superstar and the average of the small listings in each group, we find that the superstars in fact have lower quality than the small listings in 45% of the variety groups sampled. In line with this, Panel B looks at how quality predicts sales. We see that the average market share of a listing only weakly increases with quality. The difference is not significant except at the top. These patterns hold even when we restrict the sample to variety groups with relatively small price dispersion/difference across listings as shown in Figure A.2.[17]

These observations suggest potential challenges that high-quality sellers face in gaining market share and indicate potential market misallocation. To fully quantify the degree of market misallocaiton, we rely on a structural model in Section 6 to take into account differences in both quality and price.[18]

Fact 3. *The probability of receiving new orders increases as the seller's total number of cumulative orders increases.*

Lastly, we delve further into the growth dynamics of sellers. Using the transaction-level data, we document dependence of new order arrivals on past orders. Figure 4 plots the empirical probability of receiving any new order in the week following the census data collection against the number of cumulative orders collected in the census. A clear pattern emerges: listings with

[17]Interestingly, we find that superstars do not always charge the lowest price: within a variety group, the listing with the highest sales charges the lowest price only 14% of the time.

[18]Empirically, we observe a positive relationship between price and quality, which corroborates our quality measures but could partly explain the weak relationship between quality and sales.

higher cumulative orders have a higher chance of attracting new orders. In particular, 94.4% of listings with more than 500 cumulative orders receive at least one new order in the following week, whereas the fraction is only 19% for listings with 2 to 5 cumulative orders. Table A.5 regresses the dummy of receiving an order in a given week on the logged past cumulative orders of a product listing, with and without store fixed effects.

This descriptive result is, in part, driven by existing online ranking and review mechanisms that enable sellers to enhance their visibility and popularity by accumulating sales and (positive) reviews, which then speed up the arrival of future sales. Intuitively, this process could help high-quality sellers to overcome the information friction and stand out over time.

4 Experiment and Findings

The descriptive findings motivate us to conduct a demand intervention to help online SMEs to overcome the information friction and grow. In particular, the experiment generates positive demand and information shocks to a set of small exporters through randomized online purchase orders and reviews. We track sellers' performance over time to examine the impact of the treatments on sellers' subsequent growth in the online market.

4.1 Experiment Design

To select the experimental sample, we start with the same 133 variety groups with at least 100 cumulative sales aggregated across all listings within the group (see Section 2.2). Among the 1,258 product listings in the 133 groups, we identify 784 small listings with fewer than 5 orders and randomly assign the 784 small listings to three groups with different order and review treatments: control group C, which receives neither the order nor the review treatment; T1, which receives one order randomly generated by the research team and a star rating; and T2, which, in addition to receiving an order and a star rating, receives a detailed written review on product and shipping quality. Table A.6 summarizes the listing and store characteristics for the experimental sample.

Given that ratings are highly inflated on AliExpress,[19] for all the treatment groups, we leave a five-star rating for the order unless there are obvious quality defects or shipping problems.

[19]Out of the 6,487 reviews that we observe in the transaction data, 85.9% are five stars.

This is to mimic the behavior of actual consumers.[20] To generate the contents of the written reviews on shipping and product quality, we use a latent Dirichlet allocation topic model in natural language processing to analyze past reviews and construct the review messages based on the identified keywords. Appendix B.2 describes the reviews in detail.

The difference between T1 and C identifies the impact of receiving an initial order and a star rating. The difference between T1 and T2 identifies any additional impact of receiving a detailed written review. To allow comparisons across otherwise "identical" listings, we stratify the randomization by variety group. For varieties sold by two small sellers (and other large sellers), we assign 1 to the control and 1 to the treatment. We then pool the latter across variety groups and randomly split them into T1 and T2 with equal probabilities. For varieties sold by more than two small sellers, we assign 1/3 of the small listings to each of C, T1, and T2. This randomization procedure is powered to identify the impact of the order treatment, followed by the impact of the written reviews. In the end, we have 300 listings in C, 258 in T1, and 226 in T2. Table A.7 presents the balance checks and shows that the randomization is balanced across baseline characteristics.

4.2 Results: Treatment Effects of Initial Demand and Information

To examine the effects of order and review treatments on sellers' subsequent growth, we track all listings for 13 weeks after the initial order placement and estimate the following regression:

$$\text{WeeklyOrders}_{it} = \beta_0 + \beta_1 \text{Order}_i + \beta_2 \text{Review}_i \times \text{PostReview}_t + \lambda_t + \nu_{g(i)} + \epsilon_{it} \tag{1}$$

where the dependent variable is the total number of orders (excluding our own order) for listing i in week t.[21] Order is a dummy variable for receiving the order treatment (which equals 1 for T1 and T2). Review is an indicator for receiving additional shipping and product reviews (T2). PostReview is a time dummy variable that equals 1 for the period after the reviews were provided.[22] λ_t and $\nu_{g(i)}$ are week and variety group fixed effects. In addition, all regressions control for baseline sales at the store and the listing level. Results without baseline controls are shown in Table A.8. Standard errors are clustered at the listing level.

[20] We are not able to randomize the star ratings due to ethical considerations. A previous study by Cabral and Hortacsu (2010) shows that receiving one negative review significantly hurts a store's subsequent growth.

[21] We focus on orders instead of revenue since we observe very few price adjustments during the study period. In the 13 weeks following the initial treatment, only 6.5% of the listings experienced any price adjustments.

[22] Most of the orders arrived within the first 7 weeks. We left the online reviews in week 7 after the initial order placement, once we had received the majority of the orders.

Table 3 shows the main experimental findings. Columns (1) and (2) examine sales to all destinations, and Columns (3) to (6) look at sales to English-speaking countries and to the United States, respectively. Overall, we see that an exogenously generated initial order has a significantly positive impact on subsequent orders. Table A.9 shows that the order treatment effect is unlikely to be explained by endogenous supply-side responses, in terms of pricing, shipping offering, and observable advertising efforts. Consistent with Fact 3 in Section 3, the impact is likely driven by demand-side forces mediated through existing online mechanisms. First, receiving an initial order and a star rating generates a positive information signal that reduces quality uncertainty. This could be especially valuable for small new listings that lack a sales and review history. In addition, receiving an order may enhance a listing's visibility on the platform through the online ranking algorithm. Table A.10 provides suggestive evidence that the order treatment leads to a short-term boost in a listing's ranking. In light of this discussion, in the theoretical and empirical models, we incorporate both the information signaling channel and the visibility channel of initial demand.

The positive order treatment effect suggests that the demand intervention could potentially help sellers overcome the information friction and grow by accumulating sales. However, we find that its effectiveness is rather limited. To quantify the magnitude, Table 4 takes cumulative sales measured at the endline, netting out our own order, and estimates an average treatment effect ranging from 0.1 to 0.25. That is, 1 order generated by the research team leads to an additional 0.1 to 0.25 orders. The magnitude is much smaller than the size of the initial treatment, which explains why individual sellers would not replicate the order treatment themselves and suggests that the market friction cannot be easily overcome by individual sellers' private efforts.[23]

Last but not least, we do not find any significant impact of the written reviews. One potential explanation is that reviews matter only when a seller's listing is discovered by consumers, which is a rare event for small businesses due to their low visibility as discussed above. The findings suggest that the online review mechanism may not function effectively to resolve the information friction in the presence of large market congestion. Consistent with this, we do not find any heterogeneous treatment effects based on quality, as shown in Table A.11.[24]

This discussion echoes the earlier stylized fact that quality does not strongly predict sales

[23]In addition, the cost of manipulating orders on AliExpress (an exclusively cross-border platform) is fairly significant and greater than that on domestic platforms. It requires recruiting people overseas and gaining access to a foreign address, foreign bank account, and foreign IP address. If a buyer account or credit card is found to repeatedly place orders on listings carried by the same store, the account is at risk of being suspended.

[24]Here, we interact the treatment variable with service quality and listing ratings because product quality and shipping quality are not measured for the control-group listings.

performance in the online market. To further corroborate this point and demonstrate the role of information friction in online markets, we conduct a follow-up exercise, apart from the online experiment, in which we resell the purchased t-shirts offline to examine how quality influences sales outcomes when it becomes more observable at the point of transaction as in the offline setting. To do this, we worked with our partner children's consignment store (which conducted the quality grading) and displayed the t-shirts for sale, as shown in Figure A.3. Table A.12 shows that product quality strongly influences the price at resale (Column 1), and conditioning on price, higher quality is associated with higher probability of being sold (Columns 2 and 3). Although the estimate is less precise, partly due to the limited time window of the exercise[25], the magnitude is economically meaningful: on average 18% of the t-shirts were sold during the two-month period, which implies that a one standard deviation increase in product quality increases the sales probability by 24%.

Taken together, the evidence from the experiment and the offline reselling exercise suggest that a genuine demand for quality exists among buyers, but information friction could hamper the growth of high-quality sellers in the online market. While a demand-side intervention, working through existing platform mechanisms, could help to generate growth, its success is rather limited in helping high-quality sellers to overcome the information friction and grow.

5 Theory

We now develop a theoretical model that formalizes the role of information friction, in the presence of a large number of market participants, on firm growth dynamics and market efficiency. The model incorporates key features of the online market, including heterogeneity in seller quality and information friction, as well as a process of consumer search and learning through existing online ranking and review mechanisms. We use the model to investigate the impact of increasing the number of market participants, connecting to the policy motivation of large-scale onboarding programs, and show that such an increase may exacerbate market congestion, further slowing down the resolution of the information problem and worsening market allocation.

[25]Unfortunately, the store owner had to shut down the store for an extended period for personal reasons. Therefore, we are only able to capture sales outcomes within a two-month window.

5.1 Model Setup

Consider $N \geq 2$ sellers on a platform, whose true qualities $\{q^i\}_{i=1}^N$ are learned over time through past purchases and reviews. Consumers hold a common prior belief that $q^i \sim \mathcal{N}(0, 1)$ are i.i.d. standard normally distributed with prior mean $\widehat{q_0^i} = 0$. One consumer comes to the market in each period, purchases from some seller i, and leaves a noisy review, which serves as a signal about q^i. We model the formation of consumers' consideration set, which involves a random sampling of a small subset of sellers. The sampling probability that a seller appears in the consideration set is governed by the seller's visibility, which is the sum of some initial visibility parameter and total past sales. Among those sellers in the consideration set, the consumer then chooses/purchases from a specific seller with a logit probability that depends on the consumer's belief about its quality relative to other sellers' expected qualities. This corresponds to the choice probability of a consumer who faces random utility shocks, as we describe in more detail later when introducing our structural model.

Below we lay out the formal details of the model. Suppose that at the end of period $t \geq 0$ the cumulative sales of each seller i are s_t^i and consumers' common posterior mean of q^i is $\widehat{q_t^i}$. Then, in period $t+1$, the following occurs:

1. **Sampling Procedure:** A consumer arrives at the platform and samples K sellers $i_1, \ldots, i_K$ with replacement.[26] The probability of sampling seller i is proportional to a power function of the seller's visibility $v_t^i = v_0 + s_t^i$, where $v_0 > 0$ is a parameter that represents the sellers' common initial visibility level. Specifically, the sampling probability is modeled as
$$\frac{(v_t^i)^\lambda}{\sum_{j=1}^N (v_t^j)^\lambda}.$$
Both v_0 and $\lambda > 0$ govern the extent to which an additional sale boosts a seller's relative visibility. A smaller v_0 and a larger λ both imply a stronger effect of past sales on the probability that a seller enters future consumers' consideration sets. However, the effect of a smaller v_0 is most salient for early sales, whereas the effect of a larger λ is more persistent.

2. **Choice Procedure:** After forming the sample of K sellers, the consumer chooses to

[26]We make the assumption of sampling with replacement for clarity of exposition. In our empirical application of the model, the number of sellers N is substantially larger than K. We show in Table A.13 that the alternative procedure of sampling without replacement generates nearly identical quantitative predictions.

purchase from a particular seller i_k in this sample, with probability

$$\frac{e^{\widehat{q_t^{i_k}}}}{\sum_{\ell=1}^{K} e^{\widehat{q_t^{i_\ell}}}}.$$

This is the logit choice probability computed from the expected qualities of the sellers in the sample. For the chosen seller i_k, its cumulative sales $s_{t+1}^{i_k}$ and visibility level $v_{t+1}^{i_k}$ both increase by 1 from their period t values. All other sellers' sales and visibility are unchanged.

3. **Review and Belief Updating:** The consumer who purchases from seller i_k in period $t+1$ produces a publicly observed review of its quality. This review/signal takes the form $z_{t+1} = q^{i_k} + \zeta_{t+1}$ with an independent normal noise term $\zeta_{t+1} \sim \mathcal{N}(0, \sigma^2)$, where the parameter $\sigma \geq 0$ captures the degree of information friction in the market. The posterior mean of q^{i_k} at the end of period $t+1$ is given by

$$\widehat{q_{t+1}^{i_k}} = \frac{z_{t+1}/\sigma^2 + (1 + s_t^{i_k}/\sigma^2) \cdot \widehat{q_t^{i_k}}}{1 + s_{t+1}^{i_k}/\sigma^2}. \tag{2}$$

   This familiar Bayesian updating formula represents a weighted average of the review z_{t+1} and the belief $\widehat{q_t^{i_k}}$ before this review, with weights proportional to their respective precision levels $1/\sigma^2$ and $1 + s_t^{i_k}/\sigma^2$. Equivalently, if we let $\bar{z}_{t+1}^{i_k}$ be the average of the past $s_{t+1}^{i_k}$ reviews about seller i_k's quality q^{i_k}, up to and including period $t+1$, then $\widehat{q_{t+1}^{i_k}}$ is also equal to $\frac{\bar{z}_{t+1}^{i_k} \cdot s_{t+1}^{i_k}/\sigma^2}{1+s_{t+1}^{i_k}/\sigma^2}$.

The above fully describes the dynamics of our model, whose primitive parameters are $N, K, v_0, \lambda, \sigma$. It is evident that the information problem only resolves slowly over time as sellers accumulate sales and reviews on the platform.

5.2 Discussion of the Model

Two important remarks are in order. First, our model can be seen as a generalization of the classic Polya urn model, which corresponds to $\lambda = 1$ (sampling probability directly proportional to visibility) and $K = 1$ (consumers do not choose within the consideration set). The main distinction of our model is that with sample size $K \geq 2$, we focus on consumer choice based on heterogeneous seller qualities. Thus, higher-quality sellers are more likely to be chosen, conditioning on entering into the consumer's consideration set. This departure from the classical model leads to fundamentally different market outcomes.

A second remark is that we have presented the model in a way that is closest to our structural estimation in Section 6. However, the theoretical analysis applies to broader cases than the specific functional forms above. In particular, we can generalize the sampling procedure to make it depend on past average review as well. Our theoretical results continue to hold as long as sellers with higher reviews are at least weakly favored by the sampling procedure.[27] Empirically, we also report the robustness check results for such a more general sampling procedure in Appendix D.5.

Below, we present the main proposition and discuss the key economic intuitions. Complete proofs are provided in Appendix C.

5.3 Impact of Increasing the Number of Sellers

Our theoretical analysis reconnects with the overarching policy motivation, focusing on a crucial comparative aspect related to the number of sellers. E-commerce has the advantage of lowering entry barriers for exporting, thereby enabling a significant number of SMEs to venture online. However, with an increasing number of sellers, the process of consumer sampling can become congested. Consequently, this congestion may impede the resolution of information friction and the recognition and rise of high-quality sellers. In what follows, we study the effect of increasing the number of sellers N on market evolution outcomes.

Proposition 1. *Given any set of parameters K, v_0, λ, σ. Then, for every positive integer T, there exists $\underline{N}(T)$ such that whenever $N \geq \underline{N}(T)$, the expected quality received by the consumer in each of the periods $2 \sim T$ strictly decreases with N.*[28]

Thus, when there are already many sellers in the market, allocation worsens as the number of sellers further increases. Intuitively, there are two underlying channels. First, the presence of more sellers dampens the positive impact of one additional order on a seller's future probability of being sampled. While this force applies to all sellers, the effect is most relevant for high-quality sellers, who are favored by consumer choice. As a result, it takes longer for high-quality

[27]Given our proof of the results below, there are other straightforward generalizations. For example, it is not necessary that choice probabilities follow the precise logit formula; all we need is that every seller in the sample is chosen with a positive probability that increases with its expected quality. In addition, the review signals need not be normally distributed; we just require a standard consistency condition that with infinite signal observations, posterior expected qualities almost surely converge to the truth.

[28]The expected quality in period 1 is always zero, as in the prior belief.

sellers to accumulate demand and stand out. In addition, the presence of more sellers reduces the number of orders and review signals that each seller can obtain on average. Thus, it also takes longer for the informational uncertainty to be resolved and high-quality sellers to be discovered. In Section 6.5, we perform counterfactual analysis to quantify how market congestion affects the resolution of the information problem, and how that impacts market share allocation and consumer welfare.

6 An Empirical Model of the Online Market

We build on the theoretical model in Section 5 to estimate an empirical model of the online market to quantitatively assess the role of information friction as well as its interaction with market congestion and to examine counterfactual policies to promote high-quality SME growth online. The demand side closely follows the setup of the theoretical model; however, we extend the model to accommodate several empirical features of the online platform. Following our stylized facts in Section 3, we let product listings differ both vertically in their expected quality and horizontally based on their variety group. As a result, our empirical model allows for potential "gains from variety" as we increase the number of product listings, even if the intrinsic qualities of the newly added products are low.

On the supply side, we further incorporate seller heterogeneity in both quality and cost and model sellers' pricing decisions. We structurally estimate the model to fit the key data moments and evaluate the model's ability to rationalize the non-targeted observational moments and experimental findings.

6.1 Demand

Sampling. Following the theoretical setup in Section 5.1, consumers randomly sample K seller listings with replacement upon their arrival.[29] We allow for heterogeneity in the size of consumers' consideration set by assuming that K follows a positive Poisson distribution. Given K, the probability of each seller listing i being drawn depends on its visibility, v_t^i. As described in Section 5.1, $v_t^i = v_0 + s_t^i$; i.e., the visibility of seller listing i depends on the initial visibility parameter v_0 and cumulative sales s_t^i, reflecting the fact that products with larger cumulative

[29]Our model abstracts away from multiple listings within a store and treats each listing as an independent selling entity. This simplification does not capture across-product spillovers within a store, which are likely to matter for large sellers but be relatively less relevant for small sellers. Table A.5 shows that the demand-accumulation force is salient even with store fixed effects, i.e., at the listing level within a store.

sales often appear in more pronounced positions on the platform[30].Fix any ordered sample of sellers $(i_1, i_2, \ldots, i_K)$ of size K. The probability that this sample is considered by the consumer is given by $\prod_{k=1}^{K} R_t^{i_k}$, where we use $R_t^i = \frac{(v_t^i)^\lambda}{\sum_j (v_t^j)^\lambda}$ to denote seller i's relative visibility, moderated by the λ-power function. As discussed in Section 5.2, v_0 has a more salient impact on early sales, while λ exhibits a more persistent impact in the longer term. Section 5.2 also shows that Proposition 1 is robust to the functional form of this visibility function and can be extended to include additional variables such as reviews (see our Appendix D.5).

Beliefs and Learning. Buyers do not directly observe quality at the point of transaction but observe imperfect signals based on past reviews. Prior beliefs and the belief updating process again follow the description in Section 5.1. In particular, we assume that prior beliefs follow a standard normal distribution $q^i \sim \mathcal{N}(0, 1)$. Empirically, we standardize our quality measures to be consistent with this assumption.

The consumers' common posterior expectation of each seller listing i's quality, denoted by $\widehat{q_t^i}$, follows the Bayesian updating rule as described in Equation (2). From the discussion there, we see that the expected quality $\widehat{q_t^i}$ at time t can also be written as $\frac{\bar{z}_t^i \cdot s_t^i/\sigma^2}{1+s_t^i/\sigma^2}$, which depends on $\bar{z}_t^i$ (i's rating, or average past review) and s_t^i (i's cumulative sales). The importance of the rating ($\bar{z}_t^i$) relative to the prior belief is determined by s_t^i/σ^2 (cumulative sales adjusted by the precision of the review signals).

Purchase and Review. We extend the baseline logit demand framework described in Section 5.1 to include variety groups, prices, and an outside option of nonpurchase with mean utility zero. Listings on the platform are both vertically and horizontally differentiated. For product listings in the same variety group g (i.e., t-shirts with identical design and observable attributes), we assume that consumers draw a common idiosyncratic preference shock ε_g for all listings in the group. The preference shock ε_g thus captures the value of horizontal differentiation.

Consumers' perceived utility of purchasing from listing i in the consideration set can be written as a function of the posterior expected quality $\widehat{q_t^i}$ and price p_t^i:

$$U_t^i = \beta + \widehat{q^i}(\bar{z}_t^i, s_t^i) - \gamma p_t^i + \varepsilon_g, \quad \forall i \in g$$

[30] A strand of the marketing literature examines how online ranking algorithms interact with consumer search and leverage detailed consumer browsing data in online marketplaces (e.g., De los Santos and Koulayev (2017); Chen and Yao (2017); Ursu (2018)). In the absence of granular data on consumer behavior, we abstract away from the exact formation process of consumers' consideration set, and focus instead on the impact of market congestion resulting from increasing the number of sellers, holding fixed the consumers' consideration set.

where ε_g represents an idiosyncratic preference shock with an i.i.d. type-I extreme value distribution for variety g. β and γ are the constant and the price coefficient. Note that in this formulation, if multiple listings from the same variety group get sampled, consumers strictly prefer the listing with the highest expected price-adjusted quality $\widehat{q^i}(\bar{z}_t^i, s_t^i) - \gamma p_t^i$. We could have further added seller-listing specific preference shock ε_i but decided to abstract from it for two reasons. First, a large fraction of variety groups contain only one listing. Second, consumers' random sampling can also explain market share dispersion across listings within each variety group even in the absence of idiosyncratic taste shocks.

6.2 Supply

On the supply side, we extend the baseline setup in Section 5.1 to incorporate seller heterogeneity in cost that can be correlated with quality. This is to account for the well-documented fact that higher-quality product listing might incur higher production and logistic costs. Each seller's pair (c^i, q^i) is drawn from a distribution upon the firm's entry to the online platform. We denote by ρ the correlation between c^i and q^i. To avoid further complicating our model, we assume that neither individual sellers nor consumers are sophisticated enough to dissect this population correlation of c and q. This assumption limits the possibility of using product price as a "signal" for unobserved quality.

Price Adjustment. Since the consumer's sampling process depends on each seller's cumulative orders, one might naturally think that sellers would have an incentive to compete for future demand through dynamic pricing. However, in our data, we observe very infrequent price adjustments.[31] More importantly, we do not observe systematic patterns of price increases as sellers grow their cumulative orders either.

As a result, to model pricing behavior, we assume that each seller has an exogenous probability of adjusting its price after a certain period of time. The frequency is directly matched to the empirical frequency of price adjustment. When a seller adjusts its price, it *does* recognize that it will be competing with a small set of rivals if they end up in the consumer's consideration set. We use D_i to denote the perceived demand of seller i, which is the probability of the seller getting sampled and eventually being chosen by the consumer. Thus D_i depends on the

[31] In our study sample with $1,258$ listings, there were only 142 price adjustments during the 13-week post-treatment periods. We also find little empirical evidence of life-cycle price dynamics for sellers, in particular, for those with higher measured quality. The lack of price movement is consistent with the results documented in Fitzgerald et al. (2020).

complete set of public information $\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}$, including the prices, ratings, and cumulative sales of all sellers at the time of a price adjustment. For product listing $i = i_1 \in g$, its perceived demand depends on *all* possible combinations of rivals $i_2, ..., i_K$:

$$D_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = K \sum_{i_2,...,i_K} \prod_{k=2}^{K} R_t^{i_k} \cdot share^i, \tag{3}$$

where $share^i$ is the probability of consumer purchasing listing $i = i_1$.

We next adjust for the probability that sampled listings can be from the same variety group. Define a subset of listings $i_1, i_2, .., i_{\bar{K}}, \bar{K} \leq K$ where we keep only listings of highest price-adjusted quality from each variety group in the sample. If listing i remains in this subset, the standard logit share applies

$$share^i = \frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{\bar{K}} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]}$$

where $\widehat{q^i}$ is a shorthand for the expected quality $\widehat{q^i}(\bar{z}^i, s^i)$. Otherwise, $share^i = 0$, i.e., there is another listing from the same variety group as i and dominates i in terms of expected price-adjusted quality. Given the demand function D_i, seller i solves the following problem:

$$\max_{p_i} D_i \cdot (p_i - c_i),$$

where the first-order condition reads

$$p_i - c_i = -\frac{D_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s})}{\partial D_i / \partial p_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s})}. \tag{4}$$

Given the additive structure of D_i, we can easily define the key piece of demand elasticity:

$$\frac{\partial D_i}{\partial p_i}(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = -K\gamma \sum_{i_2,...,i_K} \prod_{k=2}^{K} R_t^{i_k} share^i \times (1 - share^i).$$

This formula makes it clear that similar to a standard discrete choice model, a seller's own elasticity is decreasing in its probability of being chosen, conditioning on being considered by the consumer. However, this strategic consideration now also depends on the relative visibility $R_t^{i_k}$ of all its potential rivals.

Entry. To close the model, a large number of potential listings can be introduced to each variety group by paying a variety group-specific entry cost. The cost covers initial sourcing and logistic efforts for carrying the variety online. Upon entry, each listing obtains a random draw of quality q and cost c. Sellers of the listings then set their initial prices as specified above. Given the free entry condition, one can recover the group-specific entry cost that justifies the discounted payoff flow of an average entrant in the data[32].

6.3 Model Estimation

6.3.1 Parametrization and Identification

Our model has seven structural parameters: $\{K, v_0, \lambda, \sigma, \beta, \gamma, \rho\}$. The consumer demand depends on the size of the consideration set K, the initial visibility parameter v_0, the power parameter of the sampling function λ, the review signal noise σ, and the constant and price coefficient in mean utility, β and γ. On the supply side, to allow for flexible correlation between each seller-listing's quality q and cost c, we use a Gaussian copula to model the dependence of their respective marginal distributions. The dependence is governed by parameter ρ.

Despite the richness of our data on each seller-listing's online sales history, the data provide relatively little information on the variation in their unobserved cost over time. This makes it challenging to identify the consumer price elasticity parameter γ. Thus, we start by calibrating γ to an average price elasticity of 6.7 (in line with the estimates in Broda and Weinstein (2006) for apparel) and calibrate β to match the market share of the outside option.[33] Another important parameter of the model is the size of consumers' consideration set K. Prior studies have found that consumers effectively consider a surprisingly small number of alternatives, usually between 2 to 5, before making a purchase decision (Shocker et al., 1991; Roberts and Lattin, 1997).[34] Therefore, in our baseline estimate, we assume that K follows a positive Poisson distribution with mean 2. Section 6.6 performs robustness checks with different values of γ and K.

The rest of the structural parameters $\{v_0, \lambda, \sigma, \rho\}$ are estimated using the Method of Simulated Moments. We use the following data moments:

[32] As a result, the variation in the number of sellers, as shown in our comparative statics in Table 8, can be attributed to a corresponding change in entry costs.

[33] The Payers Inc. (2020) estimates that AliExpress's market share for its largest market, Russia, is approximately 58%. To be conservative, we impose an outside market share of 50% in our estimation.

[34] Studies consistently find that in online marketplaces, the vast majority of consumers search very little and thus end up consider a very small subset of sellers (e.g., Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Wildenbeest, 2011).

1. The distribution of cumulative sales for the sellers;
2. The dependence of new orders on cumulative orders;
3. The conditional distribution of cumulative orders for each measured quality segment;
4. The regression coefficient of log price and the measured quality.

We simulate our model from the start until the average listings' cumulative orders reach the level in our data (31 per listing). We initiate our simulation directly from the empirical distribution of listings across variety groups. Specifically, as reported in table A.14, we simulate 10,089 listings distributed across 3,244 variety groups (the average number of listings per group is 3.11, with variations across groups). We will show later that our estimated model also matches well the cross-group sales concentration despite not explicitly targeting those moments.

All the moments are jointly determined by the structural parameters in our model. However, some data moments are more informative about a specific parameter than others. The distribution of cumulative sales is tightly related to v_0 and λ. Intuitively, a small v_0 and a large λ amplify the effect of accumulating initial orders on the seller's subsequent growth, leading to higher market concentration over time. In addition, the dependence of a listing's new order on cumulative orders provides another channel disciplining v_0 separate from λ. Conditioning on v_0 and λ, the correlation between a listing's cumulative orders and measured quality identifies the review signal noise σ. If reviews were very precise, then higher-quality sellers would grow their listing rapidly once they ended up in a consumer's consideration set. In contrast, a larger σ results in a flattened relationship between quality and cumulative orders. Finally, a competing force that could result in a low correlation between cumulative orders and quality is the cost-quality dependence ρ. Hence, we also require our simulated data to be consistent with the observed correlation between price and our measured quality.

We bootstrap the weighting matrix using our data sample. We describe the detailed simulation and estimation procedures in Appendix D.

6.3.2 Estimation Results

Table 5 presents the parameter estimates with standard errors. The parameter v_0 that governs the initial visibility is estimated at 0.26 and λ at 0.97. To interpret the magnitudes, consider the initial stage of a market where one seller makes its first sales while all other sellers have zero sales; the visibility of the former increases by 4.6 times relative to that of the latter. The review noise σ is estimated at 5.39. This result implies that the standard deviation of the posterior belief is reduced by only 3.3% after one order is made (recall that the standard deviation of the prior belief for quality is 1). Combined, the estimates suggest that while sellers can achieve growth by accumulating sales and reviews, the latter are very noisy signals of quality; as a result, the uncertainty about each seller's quality is resolved very slowly, i.e., only after a substantial number of orders have accumulated.

On the supply side, the estimate for ρ is 0.48. Given the empirical marginal distribution of costs and the standard normal quality distribution, this translates into a coefficient of correlation between quality and cost of 0.482. Table 6 demonstrates how well our model matches the moments. Our model is over-identified. With essentially four parameters, we can match the market concentration, the dependence of new orders on cumulative orders, the correlations between price and quality, and the cumulative orders versus quality relationship very well.

6.4 External Validations of the Model

We now evaluate our model's ability to rationalize the untargeted patterns of order arrivals documented in both the observational and experimental data. Figure 5 reports the model's predicted probability of receiving a new order within a week for seller-listings of different cumulative sales. In line with Fact 3 in Figure 4, the probability rises steeply with past cumulative sales. For sellers with more than 100 sales, almost surely (90%) they will receive an additional order in the following week in both the model and data. In contrast, for the sellers with fewer than 5 cumulative sales, the chance is less than 20%. The results indicate that our modeling of the demand structure, in particular the sampling probability, despite its simplicity, captures the salient features of order arrivals across sellers with a broad range of cumulative sales.[35]

Next, more importantly, we show that our model is able to rationalize the experimental findings in Section 4. Table 7 presents the model-predicted treatment effects for various one-time demand shocks as the fraction of treated sellers and the number of purchase orders vary. Recall that in our experiment, 4% of the sellers were treated and they each received 1 order. Since the overall market is growing, we conduct the treatment in our model at the point when the number of average cumulative orders per seller is the same as that in the data, and we

[35]In addition, Table A.15 reports the share of cumulative orders accounted for by listings in the top product groups. Our model simulation is initiated with the number of listings observed for each variety group in the data but we did not target the cross-group cumulative sales distribution in our estimation. We find that our model also matches the share of cumulative orders in top groups well, although it slightly under-predicted the concentration at the very top.

simulate the market forward for a number of periods that matches the overall growth in sales between the baseline and endline.[36] In our baseline experiment simulation ($P = 4\%$, $O = 1$), the model predicts an average treatment effect of 0.129, which is quantitatively comparable to the experimentally estimated average treatment effect between 0.1 and 0.25 as shown in Table 4. Finally, using the model, we can simulate large demand shocks. Table 7 shows that when the number of orders increases from 1 to 2 and to 5, the average treatment effect scales up proportionately. However, notice that the size of the effect is always lower than that of the initial treatment, suggesting that the market friction may not be easily overcome by sellers' private efforts of accumulating orders. Contrasting this with the earlier discussion on the salient impact of accumulating sales at the initial stage of the market (where all sellers have zero sales), we see that this impact, in terms of facilitating the growth of newcomers, is considerably more muted in a mature market congested with many established incumbents.

6.5 Impact of Reducing Information Friction and Congestion

Using the estimated model, we first quantify the distortion due to information friction. In particular, we consider a hypothetical scenario where review signals are perfect—meaning a seller's true quality is fully disclosed with the first review. In the empirical model, sellers differ in both quality and cost. Therefore, to summarize the market allocation outcomes, we construct a *cost-adjusted* quality measure[37] and examine the distribution of market shares for the top sellers using this metric.

Panel A of Table 8 summarizes the impact of improving the informational signal of the reviews on allocative efficiency, using our baseline estimates in Table 5. Columns 1 to 3 show that sellers with higher quality and lower cost gain higher market shares: the cumulative market share of sellers in the top 10% in terms of cost-adjusted quality increases by 70% ($= 0.9/0.53 - 1$) when the review noise σ decreases from 5.39 to 0. Market shares for the top 25% and 33% also increase. As a result of the improved allocation, the expected consumer surplus increases by 57% ($= 1.092/0.694 - 1$) as shown in Column 4. This takes into account the fact that higher-quality sellers charge higher prices; there is sizable gain in consumer surplus due to improved allocative efficiency.[38]

[36]In our experiment, we evaluate the impact after 13 weeks of the treatment (during which period total market orders grew by 41.9%). This number guides our choice of the number of post-treatment periods in the model to evaluate the result.

[37]The cost-adjusted quality of listing i is defined as $q_i - \hat{\gamma} c_i$, where $\hat{\gamma}$ is the baseline estimate for γ.

[38]Without information friction, consumer surplus can be computed with the standard log sum formula.

Next, we examine how market congestion affects the resolution of the information problem. Specifically, we consider the impact of reducing the number of sellers, N, holding the baseline level of informational uncertainty (at $\sigma = 5.39$). Conceptually, the reduction in the number of sellers is analogous to raising entry costs or the costs of maintaining active product listings on the platform. We experiment with two ways of reducing the number of sellers, either randomly or by targeting niche variety groups that feature only one listing in the group. The latter entails a bigger loss of product variety.

Guided by our theoretical model, we first examine the expected quality of the chosen seller over time. Panel A of Figure 6 shows that the expected sample quality, weighted by the choice probabilities, improves at a faster rate when N is reduced (for the case of random removal). The results are consistent with Proposition 1: reducing the number of sellers allows high-quality sellers to be discovered faster. Over time, sellers with higher quality receive higher visibility, as shown in Panel B of Figure 6.

Panel B of Table 8 summarizes the impact of reducing the number of sellers on allocative efficiency. We see that market share allocates towards higher-quality sellers compared to the baseline case in Panel A. As a result of the improved allocation, the expected consumer surplus increases by 7% ($= 0.73/0.69 - 1$) as shown in Column 4. The last column reports the loss of varieties due to the removal of the listings. Despite that, given the limited capacity of consumer consideration set, the gain in allocative efficiency dominates the loss from varieties, thereby raising the average consumer surplus on net. In Table A.16, we further show the allocative efficiency gain remains larger than the variety loss even when the size of consumer consideration set is substantially expanded to $K = 50$.[39]

Finally, we perform two extensions of the baseline model and examine the robustness of the main findings above. First, we examine the possibility that, in addition to reviews, consumers could use sellers' cumulative sales as an additional signal for quality. Appendix D.4 formalizes such a learning procedure. The results from estimating the extended model, as reported in Table A.17, are qualitatively and quantitatively similar to the main findings in Table 8. Second, as we discussed in our theory Section 5.2, the implication of having fewer sellers on market allocation applies broadly to alternative sampling procedures. In Table A.18, we re-calibrate the model

However, with information friction, consumer surplus takes a more complicated form because beliefs under which purchasing decisions are made differ from the truth. We follow the procedure developed in Leggett (2002) to compute the consumer surplus with belief adjustment. Details are provided in Appendix D.3.

[39]When $K = 500$, reducing the number of sellers from $10,000$ to $5,000$ starts to generate lower average consumer surplus, corroborating the conventional wisdom of "loss of variety". We also show that, when $K = 500$, this loss is especially substantial when we remove listings belonging to niche variety groups.

parameters and show that market allocation and consumer surplus similarly improve with a reduction in the number of sellers when we allow the sampling probability to depend on both past sales and reviews, by generalizing the sampling weights to $(v_0 + s_i)^{\lambda} \cdot \exp(\zeta \bar{z}_i)$. Appendix D.5 discusses the procedure and results in greater detail.

6.6 Additional Robustness Checks

Table A.13 reports estimated parameter values under alternative consideration set sizes and price elasticities. Table A.19 examines the robustness of the counterfactual analyses under the different parameter values. For comparison, Panel A reproduces the baseline ($\sigma = 5.39, N = 10,000$) and counterfactual ($\sigma = 5.39, N = 5,000$) outcomes of reducing N using the parameter estimates in Table 5. Panel B shows the outcomes under parameter values re-estimated with a larger consideration set size K drawn from a positive Poisson distribution with mean 5. We see that reducing the number of sellers has a similar positive impact on market share allocation and consumer surplus under larger values of K. In Panels C and D, we perform two robustness checks under different parameter values of γ that correspond to different price elasticities (4 and 10, respectively). Conditional on fitting the same set of data moments, we find that the market allocation and average consumer surplus are very similar to those in the baseline case (where the price elasticity is 6.7). The key counterfactual analyses also remain robust.

7 Policy Discussion

We now use the model to evaluate different onboarding initiatives by governments in developing countries that aim to bring SMEs online and facilitate the growth of high-quality businesses through e-commerce. In particular, we focus on the government's decision between partnering with large existing platforms (for example, the e-Smart IKM program and Export program with Alibaba in Indonesia, the Global export program with Amazon in Vietnam) versus creating new, designated marketplaces to host the newly onboarded SMEs (for example, the Pan-African e-Commerce Initiative in Ghana, Kenya and Rwanda).

Most existing onboarding initiatives seek to onboard SMEs to existing marketplaces on global e-commerce platforms, such as AliExpress. Panel A of Table 9 performs a model-based evaluation of such a program. We start with the simulated market configuration at the end of our sample period to imitate a mature global e-commerce marketplace. We then add another $1,000$ sellers into the market and simulate the market forward by another 6 months. These new sellers do not pay for the sunk cost of entry. This is motivated by the fact that most existing policy initiatives cover the costs of initial on-boarding and training for the SMEs. Columns 1 and 2 show that the newly onboarded sellers accumulate 126 orders in total and earn a total profits of $121 over a period of 6 months. The low overall performance reflects the new sellers' low visibility due to lack of sales and reviews, especially for a mature e-commerce marketplace that already has many established incumbents. Furthermore, Columns 3-5 show the limited success of such an onboarding program in selecting high-quality SMEs to grow: of the 126 orders made to the newly onboarded sellers, the share captured by the top 10% sellers in terms of cost-adjusted quality is only 19%, higher than randomly assigned but significantly lower than the 53% among the incumbents as shown in Table 8.[40]

With that, we next consider an alternative onboarding program increasingly discussed among policy makers that brings targeted SMEs onto newly created marketplaces (either new platforms or designated sub-sites of existing platforms).[41] In general, new marketplaces would not be able to attract as many consumers as existing large ones. Panel B of Table 9 experiments with a few scenarios of consumer traffic, ranging from 0.1% to 1% of AliExpress in terms of potential consumer arrivals. Even with 0.1% consumer traffic, we see an improvement in overall performance and market allocation among these newly onboarded sellers. The allocation further improves with the amount of traffic on the new marketplace. With 1% consumer traffic, the market share for the top 10% sellers increases from 19% to 27%. The results highlight the trade-off between allocating budget to enhance the visibility of new marketplaces versus subsidizing SMEs to operate on large existing ones.

Finally, considering the second onboarding program as a viable policy alternative, the results in Table 8 highlight one important policy lesson in designing such a new marketplace: onboarding too many sellers, for example through blanket-wide training and entry subsidies, can aggravate market congestion, slow down the resolution of the information problem, and result in market misallocation. Indeed as shown in Panel C of Table 9, when we scale down the onboarding from 1000 new sellers to 500, market allocation improves, consistent with the previous theoretical and structural findings.

[40] In both Table 8 and Table 9, we hold the period of simulations to be the same.

[41] For example, the Pan-African e-Commerce Initiative is exploring the development of a regional Business-to-Business platform to digitize and onboard East Africa's leather value chain.

8 Conclusion

In this paper, we study how information friction, in the presence of market congestion, affect the growth dynamics of small exporters and market efficiency of global e-commerce. Our study focuses on the children's t-shirt segment. While the exact degree of distortion arising from information friction and market congestion may vary across products and markets, the theoretical insight and empirical methodology can be applied to study other e-commerce settings

One key policy takeaway is that, despite the increasing emphasis on the opportunities provided by global e-commerce, universal blanket-wide onboarding initiatives may not be effective at promoting firm growth and achieving allocative efficiency. Our paper speaks to the need for more effective policies to help SMEs, especially high-quality ones, overcome market frictions beyond the initial entry point. Our finding that information friction interacts with market congestion points to a few fruitful directions for future exploration: for example, combining onboarding with screening and certification that inject information to the market may help facilitate the growth of high-quality sellers and improve overall market efficiency.

We believe that some of the economic insights generalize beyond e-commerce to broader market settings. It is well understood that there can be excessive entry when firms do not internalize their business-stealing from competitors (Mankiw and Whinston, 1986). Our paper shows that with the presence of market congestion, such business-stealing extends beyond simple price competition, when sellers compete for customer attention. We further show that the business-stealing effect is particularly costly when there exists large information friction, which prevent the best firms from being discovered and worsen market allocation. We believe that the broad lessons can be applied to other settings beyond global e-commerce that feature market congestion and information friction.

References

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Atkin, David, Amit K Khandelwal, and Adam Osman, "Exporting and firm performance: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

Bai, Jie, "Melons as lemons: Asymmetric information, consumer learning and seller reputation," *Working paper*, 2016.

_ , Panle Jia Barwick, Shengmao Cao, and Shanjun Li, "Quid pro quo, knowledge spillover and industrial quality upgrading," *Working paper*, 2019.

Bold, Tessa, Selene Ghisolfi, Frances Nsonzi, and Jakob Svensson, "Market Access and Quality Upgrading: Evidence from Four Field Experiments," *American Economic Review*, 2022, *112* (8), 2518–2552.

Broda, Christian and David E Weinstein, "Globalization and the gains from variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Cabral, Luis and Ali Hortacsu, "The dynamics of seller reputation: Evidence from eBay," *The Journal of Industrial Economics*, 2010, *58* (1), 54–78.

Chen, Yuxin and Song Yao, "Sequential search with refinement: Model and application with click-stream data," *Management Science*, 2017, *63* (12), 4345–4365.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer price search and platform design in internet commerce," *American Economic Review*, 2018, *108* (7), 1820–59.

Ershov, Daniel, "Variety-Based Congestion in Online Markets: Evidence from Mobile Apps," Technical Report, University College London 2022.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi, "How exporters grow," *Working paper*, 2020.

Fredriksson, Torbjörn UNCTAD, "Unlocking the potential of e-commerce for developing countries," *Report*, 2016.

Goeree, Michelle Sovinsky, "Limited information and advertising in the US personal computer industry," *Econometrica*, 2008, *76* (5), 1017–1074.

Hansman, Christopher, Jonas Hjort, Gianmarco León-Ciliotta, and Matthieu Teachout, "Vertical integration, supplier behavior, and quality upgrading among exporters," *Journal of Political Economy*, 2020, *128* (9), 3570–3625.

Hong, Han and Matthew Shum, "Using price distributions to estimate search costs," *The RAND Journal of Economics*, 2006, *37* (2), 257–275.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest, "Empirical search and consideration sets," *Handbook of the Economics of Marketing*, 2019, *1* (2), 193–257.

Hui, Xiang, Ginger Zhe Jin, and Meng Liu, "Designing Quality Certificates: Insights from eBay," Technical Report, National Bureau of Economic Research 2022.

Jin, Ginger Zhe and Andrew Kato, "Price, quality, and reputation: Evidence from an online field experiment," *The RAND Journal of Economics*, 2006, *37* (4), 983–1005.

Leggett, Christopher G, "Environmental valuation with imperfect information the case of the random utility model," *Environmental and Resource Economics*, 2002, *23* (3), 343–355.

Li, Lingfang, Steven Tadelis, and Xiaolan Zhou, "Buying reputation as a signal of quality: Evidence from an online marketplace," *The RAND Journal of Economics*, 2020, *51* (4), 965–988.

los Santos, Babur De and Sergei Koulayev, "Optimizing click-through in online rankings with endogenous search refinement," *Marketing Science*, 2017, *36* (4), 542–564.

Macchiavello, Rocco and Ameet Morjaria, "The value of relationships: Evidence from a supply shock to Kenyan rose exports," *American Economic Review*, 2015, *105* (9), 2911–45.

_ **and Josepa Miquel-Florensa**, "Buyer-driven upgrading in gvcs: The sustainable quality program in Colombia," *CEPR Discussion Paper No. DP13935*, 2019.

Mankiw, N Gregory and Michael D Whinston, "Free entry and social inefficiency," *The RAND Journal of Economics*, 1986, pp. 48–58.

Moraga-González, José Luis and Matthijs R Wildenbeest, "Maximum likelihood estimation of search costs," *European Economic Review*, 2008, *52* (5), 820–848.

Pallais, Amanda, "Inefficient hiring in entry-level labor markets," *American Economic Review*, 2014, *104* (11), 3565–99.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood, "The value of reputation on eBay: A controlled experiment," *Experimental economics*, 2006, *9*, 79–101.

Roberts, John H and James M Lattin, "Consideration: Review of research and prospects for future insights," *Journal of Marketing Research*, 1997, *34* (3), 406–410.

Shocker, Allan D, Moshe Ben-Akiva, Bruno Boccara, and Prakash Nedungadi, "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions," *Marketing Letters*, 1991, *2* (3), 181–197.

Stanton, Christopher T and Catherine Thomas, "Landing the first job: The value of intermediaries in online hiring," *The Review of Economic Studies*, 2016, *83* (2), 810–854.

Startz, Meredith, "The value of face-to-face: Search and contracting problems in Nigerian trade," *Working paper*, 2018.

Steinwender, Claudia, "Real effects of information frictions: When the states and the kingdom became united," *American Economic Review*, 2018, *108* (3), 657–96.

Tadelis, Steven, "Reputation and feedback systems in online platform markets," *Annual Review of Economics*, 2016, *8*, 321–340.

The Payers Inc., "Cross-border payments and ecommerce report 2020-2021," 2020.

UNCTAD, "COVID-19 and e-commerce: a global review," *Report*, 2021.

Ursu, Raluca M, "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 2018, *37* (4), 530–552.

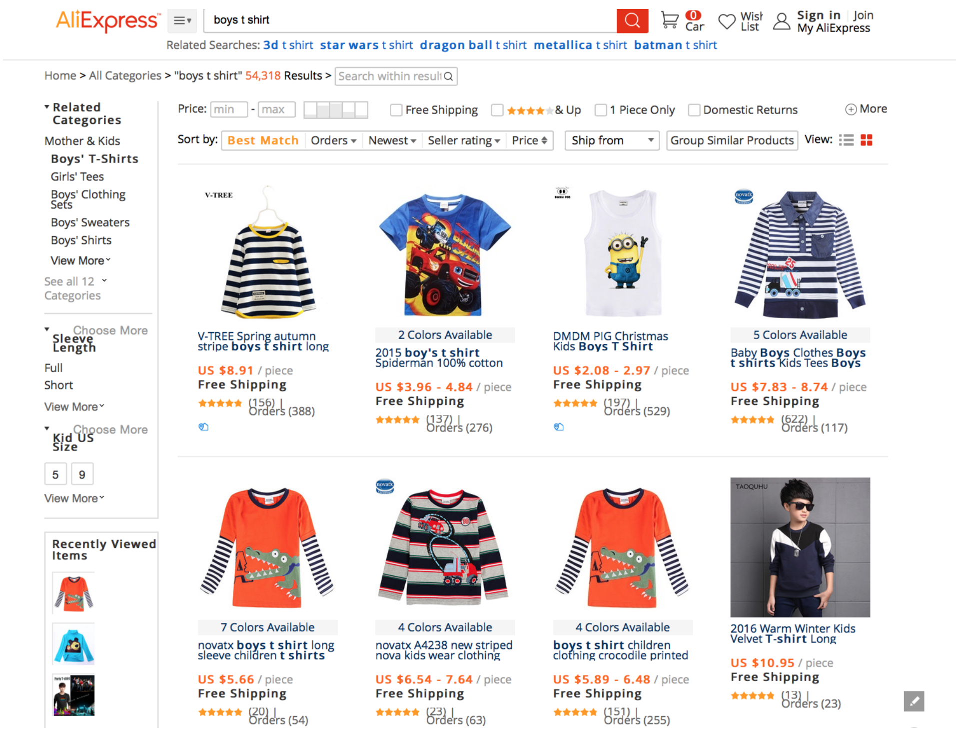
Verhoogen, Eric, "Firm-level upgrading in developing countries," *Working paper*, 2020.

Wildenbeest, Matthijs R, "An empirical model of search with vertically differentiated products," *The RAND Journal of Economics*, 2011, *42* (4), 729–757.

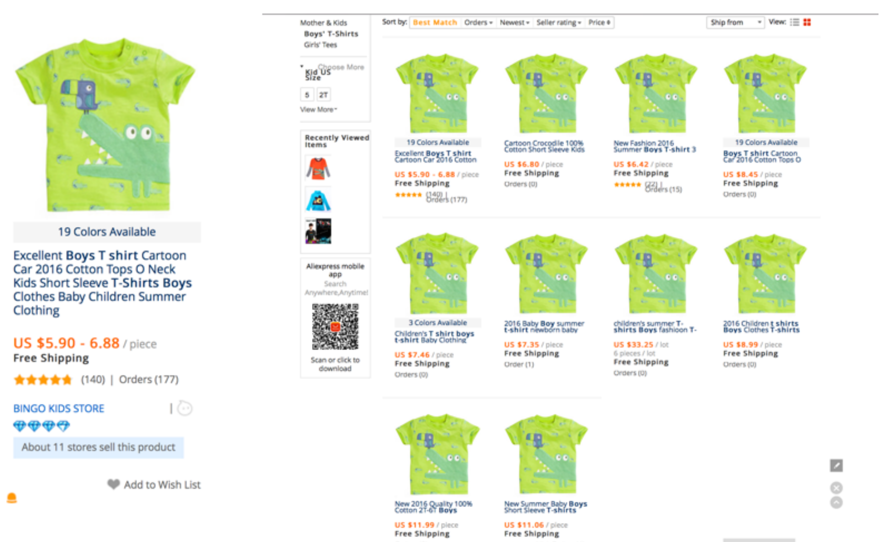
WTO, World Trade Organization, "World trade statistical review 2023," 2023.

Figure 1: AliExpress: Search Results with and without Grouping

Panel A. Search Results without the Grouping Function



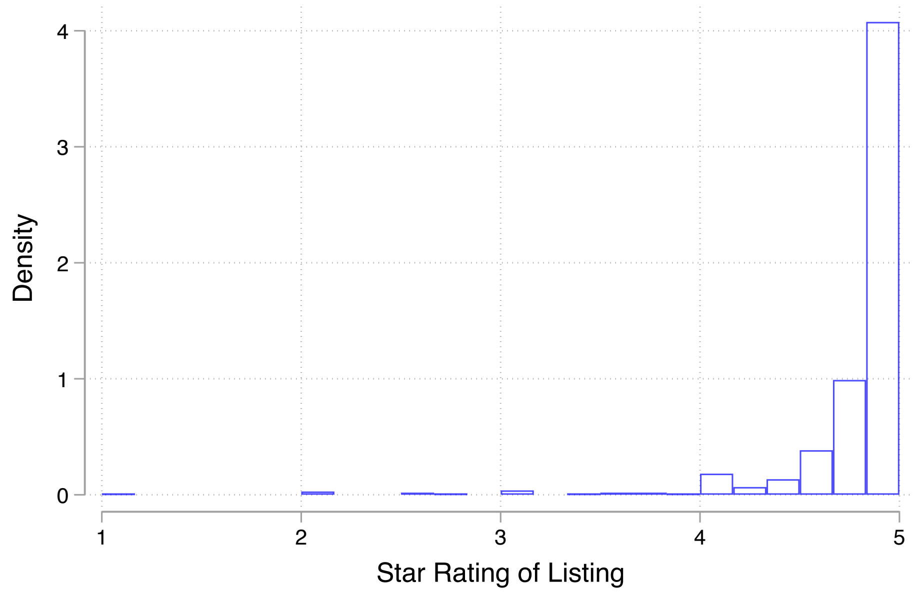
Panel B. Search Results with the Grouping Function



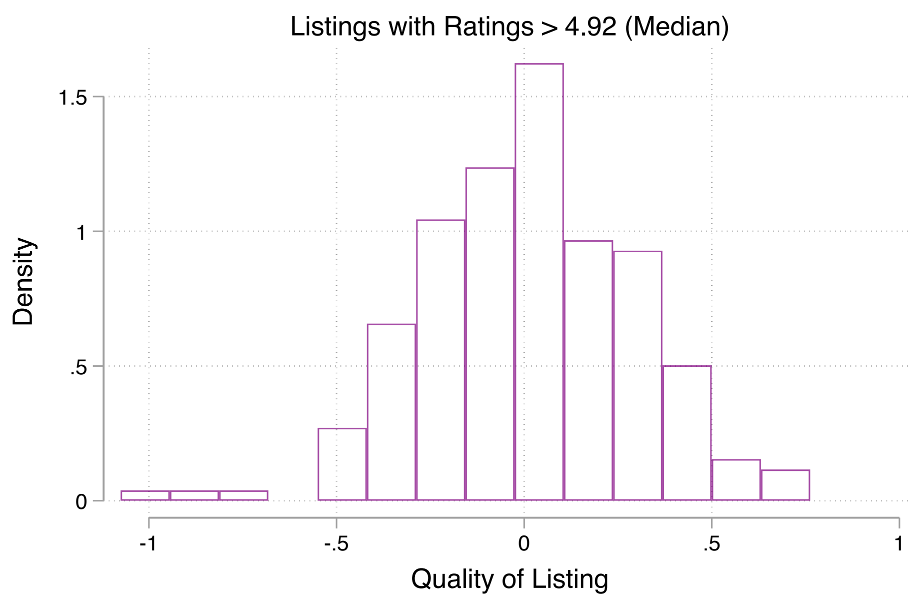
Note: This figure presents examples of search results on AliExpress. Panel A displays the search results of using "children's T-shirts" as keywords without applying the grouping function. Panel B displays the same search results with the grouping function applied.

Figure 2: Information Friction in the Online Marketplace

Panel A. Distribution of Online Star Rating



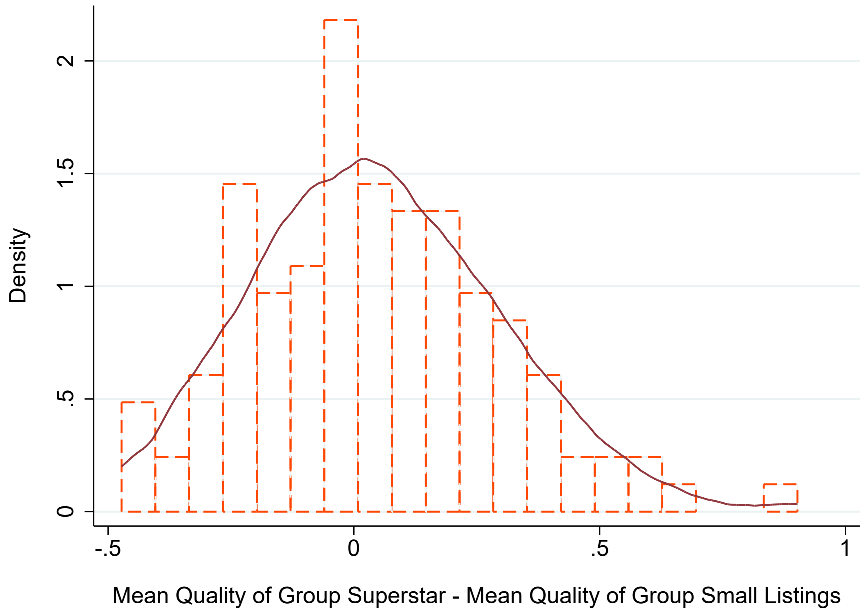
Panel B. Quality Distribution among Top-Rated Listings



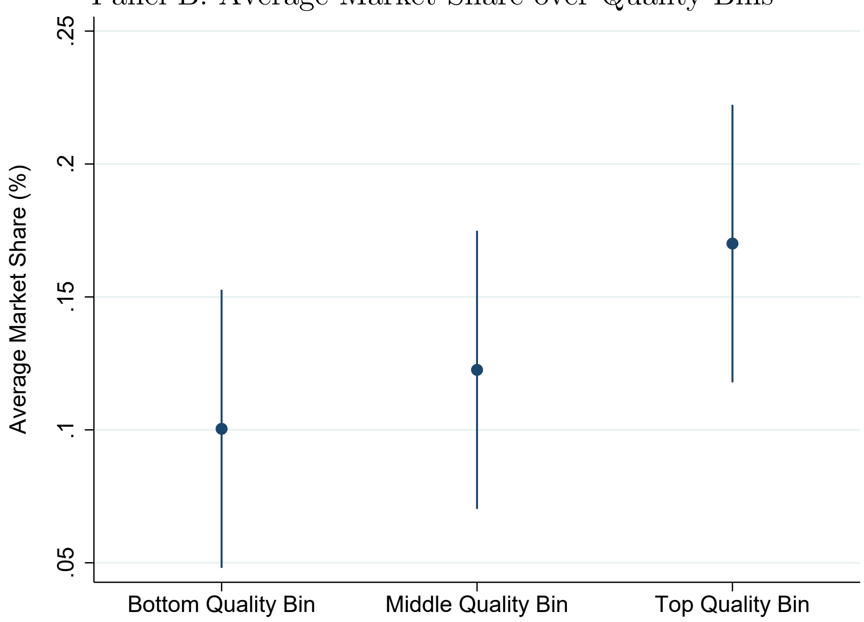
Note: This figure describes the existence of information friction in the online marketplace. Panel A plots the distribution of the online star rating, based on consumer-generated reviews, across all listings. Panel B plots the distribution of quality among listings with above-median star rating. Quality is measured in terms of the overall quality index (see Section 2.2 for details on construction of the quality index).

Figure 3: Quality and Sales Performance

Panel A. Quality Comparison between Group Superstar and Small Listings

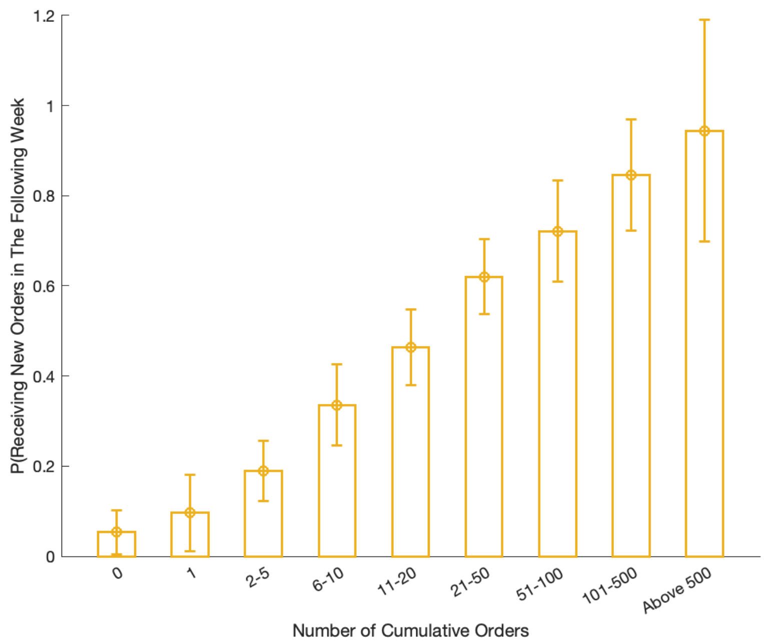


Panel B. Average Market Share over Quality Bins



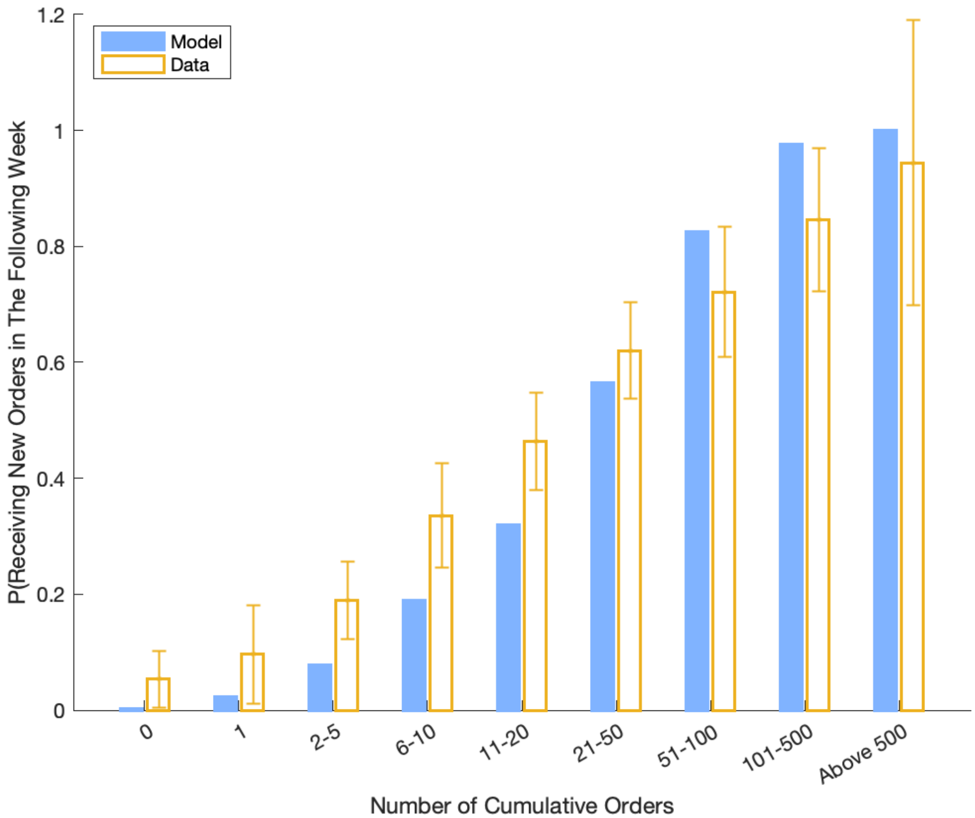
Note: This figure describes the relationship between listings' quality and sales performance. Panel A plots the distribution of the quality gap between group superstars and small listings in the group. Quality is measured in terms of the overall quality index (see Section 2.2 for details on construction of the quality index). The group superstar is defined as the listing with the largest number of cumulative orders in a group. Small listings are defined as those with fewer than 5 cumulative orders. Panel B plots the regression coefficients and the 95% confidence intervals from regressing the listings' market shares based on cumulative orders on the quality bins that they belong to.

Figure 4: Dependence of New Order Arrivals on Cumulative Orders



Note: The x-axis plots the number of cumulative orders for listings in the census sample obtained on May 21, 2018. The y-axis plots the empirical probability of getting at least one new order in the following week by size group. Smoothed 95% confidence intervals are included.

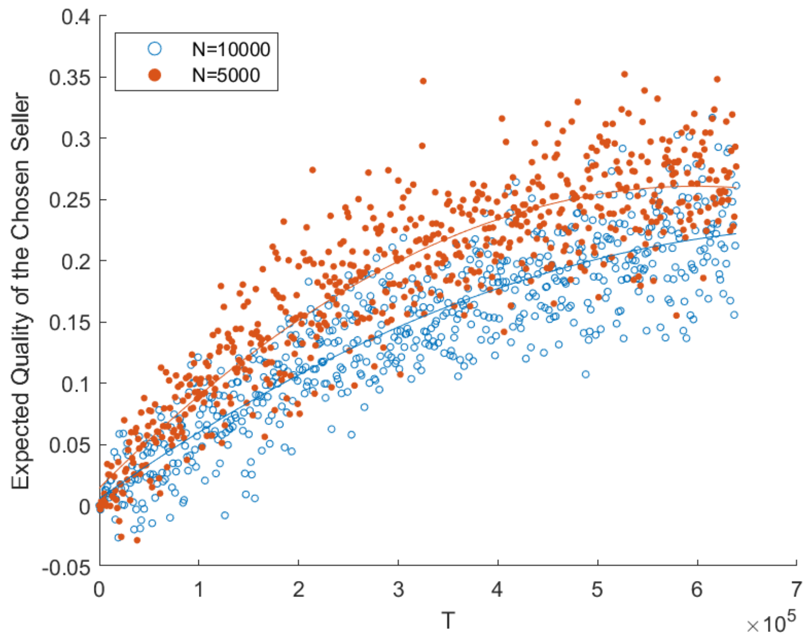
Figure 5: Model Validation - Order Arrival Pattern



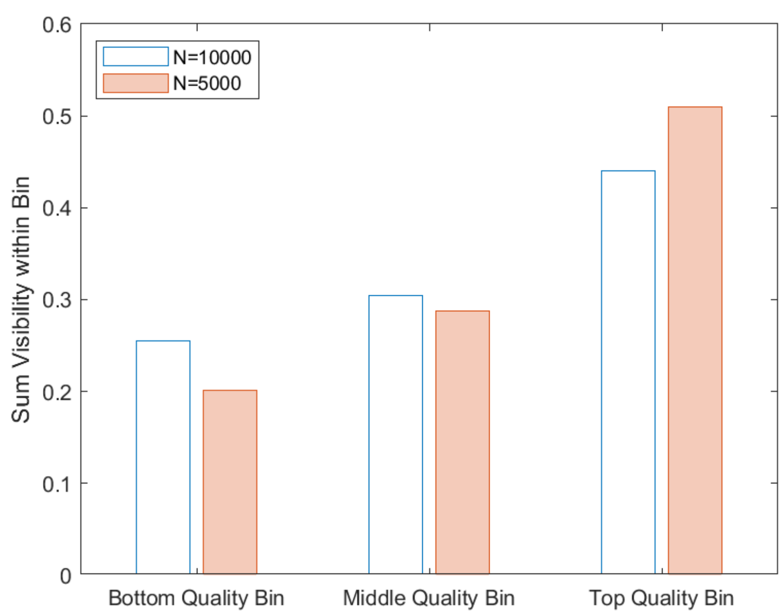
Note: This figure compares the simulated order arrival pattern in the model against its empirical counterpart. In the simulation, we first let the market reach the baseline size and then forward simulate for another week. We record the baseline number of cumulative orders for each listing, as well as whether they receive new orders in the following week. Simulations are based on our baseline parameter estimates in Table 5.

Figure 6: Model Simulation

Panel A. Simulated Expected Quality of the Chosen Seller over Time



Panel B. Simulated Visibility Distribution over Quality



Note: This figure shows the simulation results from our counterfactual exercise of randomly reducing the number of listings by a half. Panel A plots the expected quality of the chosen seller over time. Panel B plots the distribution of visibility over quality bins in the last simulation period. Data are simulated based on parameter estimates in Table 5.

Table 1: Summary Statistics of the Children's T-Shirt Market on AliExpress

	Observations	Mean	Std Dev	Median	5th Pctile	95th Pctile
		Panel A. Listing Level				
Price	10089	6.14	8.46	5	2.78	11.59
Orders	10089	31.07	189.19	2	0	110
Revenue	10089	163.7	891.68	9	0	636.4
Total Feedback	10089	19.69	127	1	0	67
Rating	5050	96.66	7.4	100	82.9	100
Free Shipping Indicator	10089	.54	.5	1	0	1
Shipping Cost to US	10089	.63	1.44	0	0	2.18
		Panel B. Store Level				
Age	1291	1.61	1.77	1	0	5
T-shirts Orders	1291	235.2	969.81	23	0	1076
T-shirts Revenue	1291	1232.05	4786.45	132.47	0	5649.28
Shop Rating	1218	4.73	.13	4.7	4.5	4.9

Note: This table reports the summary statistics for the children's T-shirt market on AliExpress based on the store-listing-level data described in Section 2.2. Panel A reports the summary statistics at the listing level. Panel B reports the summary statistics at the store level for stores carrying these listings. Price, revenue, and shipping cost to US are measured in US dollars. Total feedback for a listing is the number of reviews it has received in the past. Rating ranges from 0 to 100 and reflects the rate of positive feedback. Shop Rating is the average star score received by the store, ranging from 0 to 5. Listing and store-level ratings are only available for those with reviews.

Table 2: Summary Statistics of the Quality Measures

	Observations	Mean	Std Dev	Median
Panel A: Product Quality				
NoObviousQualityDefect	763	.93	.26	1
Durability	763	2.64	.8	3
MaterialSoftness	763	3.21	.67	3
WrinkleTest	763	3.08	.39	3
SeamStraight	763	4.23	.44	4
OutsideString	763	2.83	1.55	3
InsideString	763	.77	1.17	0
PatternSmoothness	763	3.44	1.54	4
Trendiness	763	3.14	1.36	3
Panel B: Service and Shipping Quality				
BuyShipTimeLag	819	3.67	3.24	3
ShipDeliveryTimeLag	789	12.97	4.13	12
PackageDamage	789	0	.05	0
ReplyWithinTwoDays	1258	.69	.46	1
Panel C: Quality Indices				
ProductQualityIndex	763	0	.41	-.01
ShippingQualityIndex	789	.04	.43	.12
ServiceQualityIndex	1258	0	1	.67
OverallQualityIndex	763	.01	.29	.01

Notes: This table reports the summary statistics of the various quality measures described in Section 2.2. Quality data are collected for the 133 variety groups with at least 100 cumulative sales (aggregated across all listings in the group). To measure product and shipping quality, we placed orders for 826 listings, consisting of all all treated small listings (with fewer than 5 cumulative orders) in these variety groups (as described in Section 4) and their medium-size (with cumulative orders between 6 and 50) and superstar (with the largest number of cumulative orders) peers in the same groups. Among the 826 purchase orders we placed, 819 were shipped and 789 were delivered. Due to storage and transportation issues, we managed to grade the product quality for 763 of the 789 T-shirts delivered. For service quality, we reached out to all 636 stores in the 133 variety groups. For those with multiple listings included in the 133 groups, we randomly selected one listing to inquire and assign the same service quality score to all listings sold by the same seller. Panel C reports the aggregate quality indices constructed by standardizing the scores of the individual quality metrics and taking their average within each and across all three quality dimensions.

Table 3: Treatment Effects of Order and Review

	All Destinations		English-speaking Countries		United States	
	(1)	(2)	(3)	(4)	(5)	(6)
Order	0.023	0.028**	0.015***	0.017***	0.017***	0.018***
	(0.020)	(0.014)	(0.005)	(0.005)	(0.003)	(0.003)
ReviewXPostReview	0.003	-0.018	0.021	0.015	0.017	0.014
	(0.023)	(0.028)	(0.019)	(0.018)	(0.017)	(0.016)
Observations	10192	10192	10192	10192	10192	10192
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the treatment effects of the experimentally generated orders and reviews. The dependent variable is the weekly number of orders made to different destinations, calculated using the transaction data, for the 784 listings in the experimental sample over 13 weeks. The baseline controls include the baseline total number of cumulative orders of the store and of the particular product listing. Order is a dummy variable that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. Review is a dummy that equals one for all products in T2, where we place one order and leave a written review on shipping and product quality. PostReview is a dummy that equals one for the weeks after the reviews were given. Standard errors clustered at the listing level are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table 4: Average Treatment Effects Measured at the Endline

	All Destinations	English-speaking Countries	United States
	(1)	(2)	(3)
Order	0.096	0.186**	0.245***
	(0.309)	(0.093)	(0.063)
Review	0.444	0.078	-0.006
	(0.331)	(0.100)	(0.068)
Observations	784	784	784
Baseline Controls	Yes	Yes	Yes

Note: This table reports the average treatment effects of order and review treatment. The dependent variable is the endline number of cumulative orders net of the experimentally generated one, calculated using the transaction data collected in August 2018. Order is a dummy variable that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. Review is a dummy that equals one for all products in T2, where we place one order and leave a written review on shipping and product quality. The baseline controls include the baseline total number of cumulative orders of the store and of the particular product listing. Column (1) reports the average treatment effect measured by the number of orders that the listing receives from all destinations. In contrast, Column (2) and (3) consider only orders from English-speaking countries and the United States, respectively. Standard errors are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table 5: Estimated Parameters of the Empirical Model

Parameters	v_0	σ	ρ	λ
Value	0.260	5.392	0.481	0.974
S.E.	(0.025)	(0.083)	(0.018)	(0.004)

Note: This table reports our parameter estimates for the structural model described in Section 6. v_0 governs the initial visibility; σ is the review noise; ρ is the parameter that maps to the correlation between cost and quality; and λ is the power parameter in the visibility function. Standard errors are reported in parentheses.

Table 6: Matching Moments

Moments	**Data**	**Model**
1. Market share distribution (λ)		
Top 1% cumulative revenue share	0.304	0.347
Top 5% cumulative revenue share	0.608	0.612
Top 10% cumulative revenue share	0.745	0.744
Top 25% cumulative revenue share	0.898	0.897
Top 50% cumulative revenue share	0.974	0.974
2. Dependence of new order on cumulative orders (v_0)	0.103	0.135
3. Quality and sales relationship (σ)		
Cumulative orders share: Top 1/3 quality bin	0.434	0.449
Cumulative orders share: Middle 1/3 quality bin	0.311	0.302
4. Reg. coef. of log price and quality (ρ)	0.125	0.132

Note: This table reports the data moments and the model-predicted moments evaluated at the parameter estimates reported in Table 5.

Table 7: Model Validation Using the Experiment

Percent of Sellers Purchased	Size of Purchase	Average Effect on Sales: Treated - Control
P	O	$\Delta M = 41.9\%$
4	**1**	**0.130**
4	2	0.249
4	5	0.639

Note: This table shows the simulated treatment effect based on the estimated model. The first two columns are the coverage and size of the treatment, and the last column reports the increase in cumulative orders averaged over treated sellers measured at the point after the total number of orders in the market increases by 41.9% (to mimic the actual experiment setting).

Table 8: Model Simulated Impact of Reducing Information Friction and Congestion

	Share for Top 10% Adj-Quality (1)	Share for Top 25% Adj-Quality (2)	Share for Top 33% Adj-Quality (3)	Average Consumer Surplus (4)	Variety Loss (%) (5)
			Panel A: with 10000 sellers		
$\sigma = 5.39$	0.53	0.79	0.87	0.694	0
$\sigma = 0$	0.90	0.97	0.99	1.092	0
			Panel B: With 5000 sellers and $\sigma = 5.39$		
by removing random listings	0.65	0.84	0.91	0.733	39
by removing listings from niche groups	0.70	0.83	0.90	0.733	93

Note: This table reports the results of several counterfactual exercises based on the estimates reported in Table 5. Panel A compares market outcomes when the review noise σ is reduced from the baseline estimate 5.39 to 0. Panel B considers reducing the number of product listings from 10,000 (default) to 5,000, either randomly or targeting niche variety groups, with the baseline estimated level of review noise. Section 6.5 describes the counterfactual exercises in more detail.

Table 9: Policy Counterfactuals

	Total Number of Orders (1)	Total Profits (2)	Total Share for Top 10% Adj-Quality (3)	Total Share for Top 25% Adj-Quality (4)	Total Share for Top 33% Adj-Quality (5)
Panel A: Onboarding to a Large Existing Marketplace					
$\sigma = 5.39$, 1000 Seller-Listings	127	122	0.20	0.46	0.59
Panel B: Onboarding to a New Marketplace					
$\sigma = 5.39$, 1000 Seller-Listings					
0.1% Traffic	223	185	0.21	0.48	0.62
0.5% Traffic	1340	1114	0.23	0.53	0.67
1.0% Traffic	2846	2369	0.27	0.57	0.71
Panel C: Onboarding to a New Marketplace					
$\sigma = 5.39$, 500 Seller-Listings					
0.1% Traffic	243	206	0.20	0.48	0.63
0.5% Traffic	1416	1206	0.26	0.56	0.71
1.0% Traffic	2937	2504	0.32	0.62	0.75

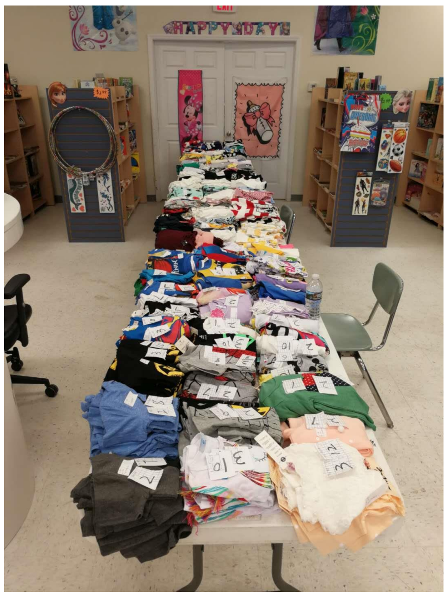
Note: Panel A reports the performance of 1000 new sellers when they are onboarded to a large existing marketplace, at the baseline estimated level of review noise ($\sigma = 5.39$). Panel B reports the performance of the same 1000 sellers when they are onboarded to a new marketplace, holding the same informational uncertainty, but under different assumptions of consumer traffic relative to the existing marketplace. Panel C reports the performance of onboarding 500 sellers to a new marketplace.

Appendices. For Online Publication Only

A Figures and Tables

Figure A.1: Product Quality Measurement

Panel A. Inspection and Grading

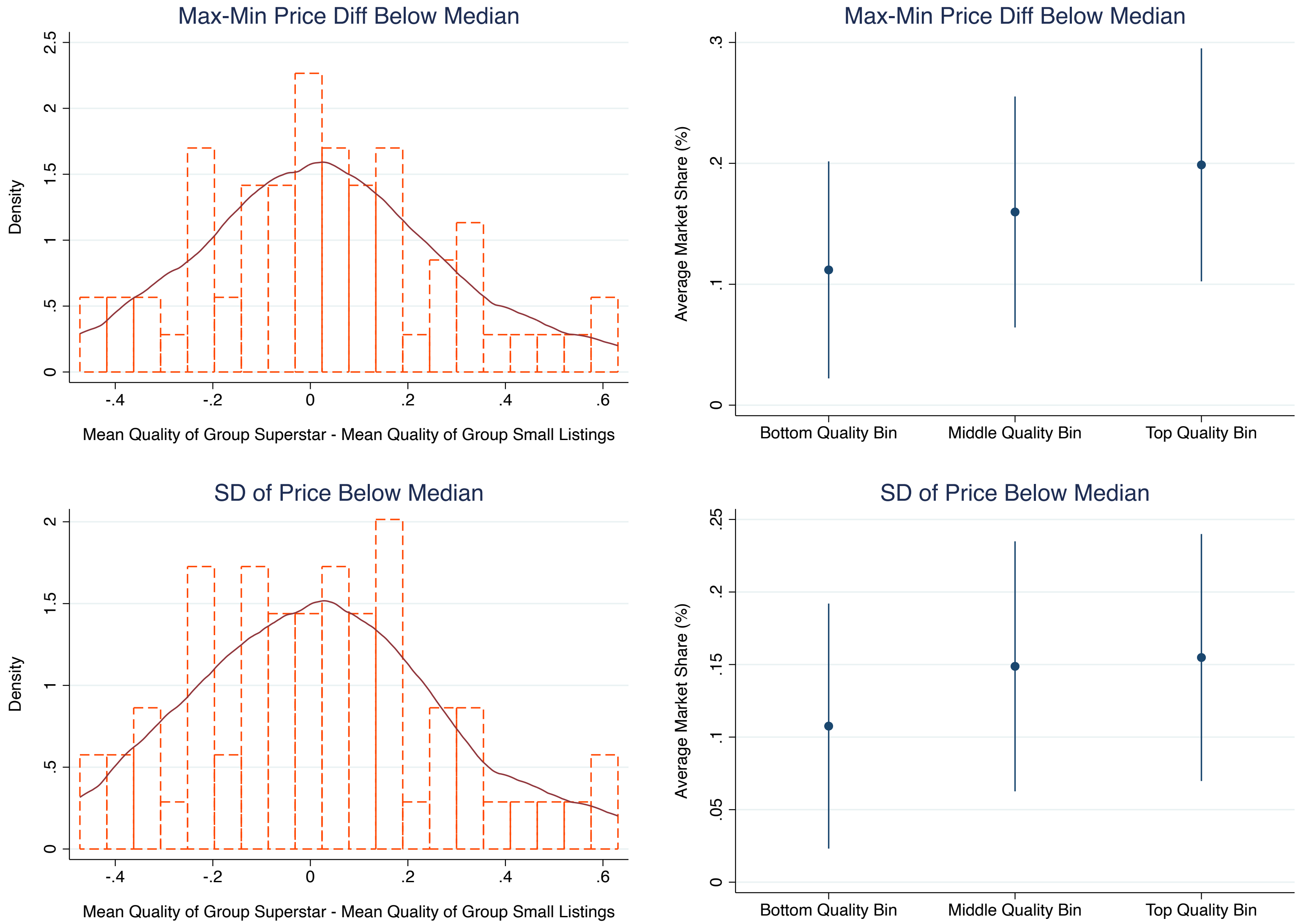


Quality Metrics:

- Obvious Quality Defect (dummy)
- Fabric/Materials (1-5 Rating) :
  - ✓ Durability/ Strength(tightly woven?)
  - ✓ Softness
  - ✓ Wrinkle test
- Seam (1-5 Rating):
  - ✓ Straight and neat (e.g. armpit)
  - ✓ Outside stray threads
  - ✓ Inside multiple unnecessary/loose stitches
- Pattern Printing (1-5 Rating):
  - ✓ Smoothness
  - ✓ Trendiness (subjective)

Panel B. Variation in Scores

Figure A.2: Quality and Sales Performance: Groups with Limited Price Dispersion



Note: This figure describes the relationship between listings' quality and sales performance for groups that have limited price dispersion. Plots in the top row are limited to groups with maximum within-group price differences that are below the median. Plots in the bottom row are limited to groups with standard deviation of prices below the median. Quality is measured in terms of the overall quality index (see Section 2.2 for details on construction of the quality index). The group superstar is defined as the listing with the largest number of cumulative orders in a group. Small listings are defined as those with fewer than 5 cumulative orders. Plots in the right column show the regression coefficients and the 95% confidence intervals from regressing the listings' market shares based on cumulative orders on the quality bins that they belong to.

Figure A.3: Offline Reselling of the T-shirts

Note: We worked with our partner, the children's clothing consignment store at North Carolina, to resell the t-shirts bought from AliExpress between July 11 2019 and September 13 2019. The figure displays a subset of the t-shirts at resale.

Table A.1: Summary Statistics of the Quality Sample

	Observations	Mean	Std Dev	Median	5th Pctile	95th Pctile
			Panel A. Listing Level			
Price	1258	5.3	3.09	4.37	2.8	9.86
Orders	1258	44.6	159.41	2	0	221
Revenue	1258	203.21	712.13	7.59	0	1090.95
Total Feedback	1258	27.42	111.99	1	0	141
Rating	624	96.04	12.84	100	84.43	100
Free Shipping Indicator	1258	.48	.5	0	0	1
Shipping Cost to US	1258	.67	.9	.21	0	2.36
			Panel B. Store Level			
Age	627	1.29	1.69	0	0	5
T-shirts Orders	636	88.22	246.51	4	0	532
T-shirts Revenue	636	401.96	1126.02	16.1	0	2253.91
Store Rating	597	4.72	.15	4.7	4.5	4.9

Note: This table reports the summary statistics for the sample of listings and stores for which we have collected quality measures. Panel A reports the summary statistics at the listing level. Panel B reports the summary statistics at the store level for stores carrying these listings. Price, revenue, and shipping cost to US are measured in US dollars. Total feedback for a listing is the number of reviews it has received in the past. Rating ranges from 0 to 100 and reflects the rate of positive feedback. Shop Rating is the average star score received by the store, ranging from 0 to 5. Listing and store-level ratings are only available for those with reviews.

Table A.2: Decomposition of the Overall Quality Index

Quality Metrics	Explained R^2
OverallQualityIndex	**100**
ProductQualityIndex	**76.0**
NoObviousQualityDefect	9.3
Durability	13.5
MaterialSoftness	8.8
WrinkleTest	7.1
SeamsSraight	6.6
OutsideString	8.3
InsideString	8.4
PatternSmoothness	9.7
Trendiness	4.3
ShippingQualityIndex	**18.2**
BuyShipTimeLag	3.4
NoPackageDamage	8.0
ShipDeliveryTimeLag	6.8
ServiceQualityIndex	**5.8**
ReplyWithinTwoDays	5.8

Note: This table decomposes the variation in the overall quality index into that explained by each individual quality subindex and metric. For the subindices (i.e., ProductQualityIndex, ServiceQualityIndex, and ShippingQualityIndex), the Shapley value is reported. For other metrics, the Owen value is reported.

Table A.3: Correlation between Quality and Online Rating

	Dependent: Star Rating					
	(1)	(2)	(3)	(4)	(5)	(6)
ProductQualityIndex	0.030	0.170				
	(0.048)	(0.114)				
ShippingQualityIndex			0.081*	0.098*		
			(0.044)	(0.056)		
ServiceQualityIndex					0.034**	0.036*
					(0.017)	(0.020)
Constant	4.802***	4.804***	4.794***	4.793***	4.793***	4.793***
	(0.019)	(0.020)	(0.019)	(0.020)	(0.016)	(0.017)
Observations	408	408	421	421	624	624
Rsquare	0.001	0.316	0.008	0.318	0.006	0.210
Group FE	No	Yes	No	Yes	No	Yes

Note: This table presents the results from regressing listings' star ratings on the three quality indices. The number of observations in each column reflects the number of listings with non-missing quality indices and star rating. Standard errors are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.4: Consistency of Service Quality Over Time

	Measured 2nd and 3rd rounds		
	Reply (dummy)	Reply within 2 days (dummy)	Hours to reply
Measured in 1st round	(1)	(2)	(3)
Reply (dummy)	0.614***		
	(0.058)		
Reply within 2 days (dummy)		0.562***	
		(0.059)	
Hours to reply			0.591***
			(0.060)
Constant	0.231***	0.259***	26.769***
	(0.052)	(0.052)	(4.094)
Observations	264	264	264

Note: This table presents the results from regressing the seller's reply behavior in the second and third rounds (stacked) on its behavior in the first round. The data consist of the 132 stores visited in June 2021, for which we collected three rounds of service quality data. Standard errors are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.5: Dependence of New Order Arrival on Cumulative Orders

Dummy=1 if having an order in the following week	(1)	(2)
Log Orders	0.092***	0.102***
	(0.001)	(0.002)
Observations	15096	15096
Store FE	No	Yes

Note: This table reports the results from regressing a dummy variable that equals one for listings that receive orders in the following week on the log number of cumulative orders in the current week. The data consists of a weekly panel of 1258 listings over 12 weeks (corresponding to the intervention period from May to August 2018 as described in Section 4). The weekly panel is constructed based on the six-month transaction data described in Section 2.2.

Table A.6: Summary Statistics of the Experiment Sample

	Observations	Mean	Std Dev	Median	5th Pctile	95th Pctile
			Panel A. Listing Level			
Price	784	5.7	3.56	4.75	2.87	10.44
Orders	784	.82	1.22	0	0	4
Revenue	784	3.91	6.37	0	0	16.98
Total Feedback	784	.49	1.49	0	0	3
Rating	167	94.11	22.34	100	50	100
Free Shipping Indicator	784	.48	.5	0	0	1
Shipping Cost to US	784	.72	.97	.21	0	2.69
			Panel B. Store Level			
Age	468	1.18	1.66	0	0	5
T-shirts Orders	477	1.35	1.85	1	0	5
T-shirts Revenue	477	6.43	9.48	3.04	0	26.9
Store Rating	439	4.71	.17	4.7	4.4	4.9

Note: This table reports the summary statistics for our experiment sample. Panel A reports the summary statistics at the listing level. Panel B reports the summary statistics at the store level for stores carrying these listings. Price, revenue, and shipping cost to US are measured in US dollars. Total feedback for a listing is the number of reviews it has received in the past. Rating ranges from 0 to 100 and reflects the rate of positive feedback. Shop Rating is the average star score received by the store, ranging from 0 to 5. Listing and store-level ratings are only available for those with reviews.

Table A.7: Balance Check

	(1) Control mean/(sd)	(2) T1 mean/(sd)	(3) T2 mean/(sd)	(4) T1-Control b/(se)	(5) T2-Control b/(se)	(6) T2-T1 b/(se)	(7) Joint Test F/(p)
Price After Discount	5.94	5.47	5.64	-0.47	-0.30	0.17	1.14
	(4.11)	(2.57)	(3.73)	(0.30)	(0.35)	(0.29)	(0.29)
Cumulative Orders	0.91	0.73	0.82	-0.18*	-0.09	0.09	0.88
	(1.27)	(1.19)	(1.20)	(0.10)	(0.11)	(0.11)	(0.35)
Total Feedback	0.46	0.38	0.65	-0.08	0.20	0.28*	1.94
	(1.22)	(1.37)	(1.88)	(0.11)	(0.14)	(0.15)	(0.16)
Positive Rating Rate	0.95	0.96	0.91	0.01	-0.04	-0.05	0.79
	(0.21)	(0.17)	(0.28)	(0.04)	(0.04)	(0.05)	(0.37)
Free Shipping Dummy	0.50	0.45	0.49	-0.05	-0.01	0.04	0.16
	(0.50)	(0.50)	(0.50)	(0.04)	(0.04)	(0.05)	(0.69)
Shipping Price	0.73	0.75	0.67	0.02	-0.05	-0.07	0.34
	(1.00)	(0.89)	(1.02)	(0.08)	(0.09)	(0.09)	(0.56)

Note: This table performs the balance check of the randomization. Columns (1)-(3) report the mean and standard deviation of the variables for each treatment group. Columns (4)-(6) show the difference between any two groups and the standard error of the difference. Column (7) performs the joint F test. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.8: Treatment Effects of Order and Review: Without Baseline Controls

	All Destinations		English-speaking		United States	
Order	0.022	0.027*	0.014**	0.016***	0.016***	0.017***
	(0.020)	(0.015)	(0.005)	(0.005)	(0.003)	(0.003)
ReviewXPostReview	0.006	-0.015	0.022	0.016	0.018	0.014
	(0.024)	(0.028)	(0.019)	(0.018)	(0.017)	(0.016)
Observations	10192	10192	10192	10192	10192	10192
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	No	No	No	No	No	No

Note: This table reports the treatment effects of the experimentally generated orders and reviews without baseline controls. Standard errors clustered at the listing level are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.9: Seller Responses after Treatments

	Panel A: Pricing Behavior							
	AdjustPrice		CutPrice		RaisePrice		ΔLogPrice	
Order	0.022	0.003	0.033	0.008	0.022	0.007	-0.003	-0.005
	(0.035)	(0.044)	(0.038)	(0.046)	(0.036)	(0.045)	(0.009)	(0.011)
Observations	716	456	716	456	716	456	716	456
Group FE	No	Yes	No	Yes	No	Yes	No	Yes
	Panel B: Shipping Costs							
	AdjustShippingCost		CutShippingCost		RaiseShippingCost		ΔLogShippingCost	
Order	-0.010	-0.040	0.010	-0.029	-0.027	-0.040	-0.022	-0.008
	(0.034)	(0.037)	(0.030)	(0.034)	(0.029)	(0.032)	(0.039)	(0.044)
Observations	715	456	715	456	715	456	715	456
Group FE	No	Yes	No	Yes	No	Yes	No	Yes
	Panel C: Product Description and Introduction of New Listings							
	ChangeTitle		ChangeDescription		ChangeNumPictures		LogNewListings	
Order	0.001	0.000	-0.008	-0.008	-0.004	-0.008	-0.096	-0.089
	(0.011)	(0.011)	(0.019)	(0.019)	(0.014)	(0.014)	(0.080)	(0.067)
Observations	764	764	784	784	763	763	758	758
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: This table presents regression results on sellers' responses after treatments. AdjustPrice is a dummy that equals one for listings experiencing price changes of more than 5% within the 13 weeks after the initial order placement. CutPrice, RaisePrice, AdjustShippingCost, CutShipingCost, RaiseShippingCost are dummy variables defined in a similar way. ChangeTitle is a dummy that equals one for listings that experienced title updates. ChangeDescription is a dummy that equals one for listings that experienced description updates. Descriptions include website pictures and textual information about pattern type, material, fit, gender, sleeve length, collar, clothing length, item type, and color. HaveNewListings is dummy that equals one for a listing if the associated store introduces new listings within the 13 weeks after the initial order placement; LogNewListings is the log number of those new listings. The number of observations in each column reflects the non-missing observations of the dependent variable. Standard errors are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.10: Treatment Effects on Ranking

	(1)	(2)
	Enter First 15 Pages	
OrderXMonth1	0.004*	0.003*
	(0.002)	(0.002)
OrderXMonth2	0.002	0.002
	(0.002)	(0.001)
OrderXMonth3	-0.000	-0.000
	(0.002)	(0.002)
OrderXMonth4	0.002	0.002
	(0.002)	(0.002)
Observations	10192	10192
Group FE	No	Yes
Week FE	Yes	Yes
Baseline Controls	Yes	Yes

Note: This table reports the treatment effects of the experimentally generated orders and reviews on a listing's ranking. The dependent variable is a dummy variable that equals one if the listing enters the first 15 pages of the search results (without grouping). The baseline controls include the baseline total number of cumulative orders of the store and of the particular product listing. Order is a dummy variable that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. MonthX is a dummy variable that equals one for the X-th month after the initial order placement. Standard errors clustered at the listing level are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.11: Heterogeneous Treatment Effects Based on Quality

	(1)	(2)	(3)	(4)
Order	0.301	0.327*	-0.489	-0.676
	(0.236)	(0.180)	(0.624)	(0.509)
OrderXServiceQualityIndex	-0.157	-0.075		
	(0.228)	(0.203)		
ServiceQualityIndex	0.290**	0.093		
	(0.119)	(0.125)		
OrderXStarRating			-0.715	-0.691
			(0.816)	(0.755)
StarRating			-0.122	0.202
			(0.392)	(0.428)
Observations	784	784	307	307
Baseline Controls	Yes	Yes	Yes	Yes
Group FE	No	Yes	No	Yes

Note: This table reports the heterogeneous treatment effects of the experimentally generated orders based on quality measures. The dependent variable is the endline number of cumulative orders net of the experimentally generated one. The fewer numbers of observations in Columns (3) and (4) reflect the fact that some listings have not received any rating. The baseline controls include the baseline total number of cumulative orders of the store and of the particular product listing. Order is a dummy variable that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. Standard errors are in the parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.12: Offline Reselling Outcomes

	Price	Sold within 2 months	
	(1)	(2)	(3)
ProductQualityIndex	0.522**	0.044	0.044
	(0.239)	(0.045)	(0.046)
Resell Price			0.001
			(0.009)
Observations	430	430	430

Note: This table reports the outcomes of reselling the t-shirts by the children's consignment store we worked with in North Carolina. The dependent variable in Column 1 is the reselling price set by our partner. The dependent variable in Columns 2 and 3 is a dummy variable that indicates whether a t-shirt was sold between July 11 2019 and Septmber 13 2019. ProductQualityIndex is the standardized product quality rating (see Section 2.2 for details). Standard errors are in parentheses. *** indicates significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.13: Robustness: Structural Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	Data	Baseline	Sampling Without Replacement	Alternative Sample Size	Smaller Price Elasticity	Larger Price Elasticity
A. Parameter Estimates						
Initial Visibility v_0		0.26	0.26	0.25	0.25	0.25
Review Noisiness σ		5.39	5.39	7.01	6.03	5.80
Quality-Cost Correlation ρ		0.48	0.48	0.45	0.45	0.45
Power of the visibility function λ		0.97	0.97	0.94	0.96	0.96
B. Simulated vs. Data Moments						
Market share distribution						
Top 1% cumulative revenue share	0.304	0.344	0.340	0.346	0.327	0.345
Top 5% cumulative revenue share	0.608	0.612	0.608	0.604	0.596	0.606
Top 10% cumulative revenue share	0.745	0.739	0.737	0.732	0.728	0.734
Top 25% cumulative revenue share	0.898	0.894	0.894	0.891	0.890	0.893
Top 50% cumulative revenue share	0.974	0.974	0.973	0.973	0.973	0.973
Dependence of new order on cumulative orders	0.102	0.135	0.135	0.137	0.137	0.136
Quality and sales relationship						
Cumulative orders share: Top 1/3 quality bin	0.434	0.443	0.443	0.422	0.435	0.458
Cumulative orders share: Middle 1/3 quality bin	0.311	0.309	0.305	0.312	0.310	0.298
Reg. coef. of log price and quality	0.125	0.132	0.132	0.124	0.131	0.112

Note: This table reports the parameter estimates and model fitness in the baseline and alternative model calibrations. Column (2) reproduces the baseline results in Tables 5 and 6. Column (3) simulates the model moments under the baseline parameter values based on sampling without replacement for the formation of consumers' consideration set. Column (4) assumes that the consumer sample size K follows a positive Poisson distribution with mean 5. Column (5) assumes that the price elasticity is 4 or, equivalently, that γ is 0.75. Finally, Column (6) sets the price elasticity to 10, where γ equals 1.89.

Table A.14: Distribution of Group Size

	Total No. of Listings	Ave No. Listings
Top 1% groups	2,146	65.03
Top 5% groups	4,651	28.71
Top 10% groups	6,063	18.71
Top 25% groups	7,599	9.38
Top 50% groups	8,468	5.22
All groups	100,89	3.11

Note: This table reports the total and average numbers of listings for the largest 1%, 5%, 10%, 25%, and 50% of groups.

Table A.15: Across-Group Sales Distribution

Total share of cumulative orders by top groups	**Data**	**Model**
Top 1% groups	27.47	22.84
Top 5% groups	56.14	49.05
Top 10% groups	70.59	65.05
Top 25% groups	87.53	88.62
Top 50% groups	96.56	99.16

Note: This table reports the total market share of the largest groups in terms of cumulative orders. Model simulations are based on the parameter estimates reported in Table 5.

Table A.16: Model Simulated Impact of Reducing the Number of Sellers

	Share for Top 10% Adj-Quality (1)	Share for Top 25% Adj-Quality (2)	Share for Top 33% Adj-Quality (3)	Average Consumer Surplus (4)	Variety Loss (%) (5)
Panel A: With $\sigma = 5.39, K = 50$, Random Removal					
10000 Seller-Listings	0.68	0.88	0.93	3.463	0
5000 Seller-Listings	0.76	0.91	0.95	3.491	39
Panel B: With $\sigma = 5.39, K = 500$, Random Removal					
10000 Seller-Listings	0.61	0.85	0.92	5.424	0
5000 Seller-Listings	0.69	0.87	0.93	5.360	39
Panel C: With $\sigma = 5.39, K = 500$, Removing Niche Variety Groups					
10000 Seller-Listings	0.61	0.85	0.92	5.424	0
5000 Seller-Listings	0.82	0.94	0.96	4.992	93

Note: This table reports the results of several counterfactual exercises based on the estimates reported in Table 5. Panel A compares market outcomes when the number of product listings is reduced from 10,000 (default) to 5,000 randomly, with the size of consumer consideration set (K) being 50 and the baseline estimated level of review noise held fixed. Panel B makes the same comparison but with K being 500. In addition to setting K to 500, Panel C further considers a counterfactual scenario where the removal of 5000 seller listings are from the niche variety groups with the fewest product listings. Section 6.5 describes the counterfactual exercises in more detail.

Table A.17: Robustness: Quality Inference Using Cumulative Orders and Reviews

Re-estimated Parameters	
v_0	0.268
λ	1.000
σ	5.402
ρ	0.482
Moments	
Top 1% cumulative revenue share	0.351
Top 5% cumulative revenue share	0.611
Top 10% cumulative revenue share	0.736
Top 25% cumulative revenue share	0.893
Top 50% cumulative revenue share	0.973
Dependence of new order on cumulative orders	0.132
Cumulative orders share: Top 1/3 quality bin	0.461
Cumulative orders share: Middle 1/3 quality bin	0.296
Reg. coef. of log price and quality	0.132
Baseline: $\sigma = 5.39, N = 10,000$	
Cumulative orders share: Top 10% adj-quality	0.537
Cumulative orders share: Top 25% adj-quality	0.789
Cumulative orders Share: Top 33% adj-quality	0.872
Average consumer surplus	0.709
Counterfactual: $\sigma = 5.39, N = 5,000$	
Cumulative orders share: Top 10% adj-quality	0.633
Cumulative orders share: Top 25% adj-quality	0.854
Cumulative orders Share: Top 33% adj-quality	0.915
Average consumer surplus	0.762

Note: This table considers an alternative version of the model where consumers use both cumulative orders and reviews to infer listing qualities, as described in Appendix D.4.

Table A.18: Robustness: Visibility Function Based on Cumulative Orders and Reviews

	$\zeta = 0.1$	$\zeta = 0.2$
Re-calibrated Parameters		
v_0	0.260	0.260
λ	0.900	0.800
σ	8.000	15.000
ρ	0.481	0.450
Moments		
Top 1% cumulative revenue share	0.340	0.314
Top 5% cumulative revenue share	0.608	0.611
Top 10% cumulative revenue share	0.740	0.761
Top 25% cumulative revenue share	0.899	0.927
Top 50% cumulative revenue share	0.976	0.985
Dependence of new order on cumulative orders	0.138	0.144
Cumulative orders share: Top 1/3 quality bin	0.459	0.444
Cumulative orders share: Middle 1/3 quality bin	0.292	0.296
Reg. coef. of log price and quality	0.131	0.120
Baseline: $\sigma = 5.39, N = 10,000$		
Cumulative orders share: Top 10% adj-quality	0.527	0.471
Cumulative orders share: Top 25% adj-quality	0.778	0.720
Cumulative orders Share: Top 33% adj-quality	0.863	0.835
Average consumer surplus	0.690	0.631
Counterfactual: $\sigma = 5.39, N = 5,000$		
Cumulative orders share: Top 10% adj-quality	0.608	0.530
Cumulative orders share: Top 25% adj-quality	0.829	0.761
Cumulative orders Share: Top 33% adj-quality	0.897	0.864
Average consumer surplus	0.736	0.668

Note: This table considers alternative sampling weights, where the probability that seller i enters the consumer consideration set depends on both its cumulative orders and its past reviews, as described in Appendix D.5.

Table A.19: Counterfactual Robustness

	Share for Top 10% Adj-Quality (1)	Share for Top 25% Adj-Quality (2)	Share for Top 33% Adj-Quality (3)	Average Consumer Surplus (4)
	Panel A (Baseline): $K \sim$ Positive Poisson(2), Price Elasticity=6.7			
10000 Seller-Listings	0.53	0.79	0.87	0.69
5000 Seller-Listings	0.61	0.84	0.91	0.74
	Panel B: $K \sim$ Positive Poisson(5)			
10000 Seller-Listings	0.52	0.77	0.87	0.67
5000 Seller-Listings	0.61	0.83	0.90	0.74
	Panel C: Price Elasticity=4 ($\gamma = 0.75$)			
10000 Seller-Listings	0.54	0.78	0.88	0.72
5000 Seller-Listings	0.62	0.84	0.91	0.76
	Panel D: Price Elasticity=10 ($\gamma = 1.89$)			
10000 Seller-Listings	0.53	0.77	0.87	0.71
5000 Seller-Listings	0.61	0.83	0.91	0.75

Note: This table displays the robustness of the model's predicted effect of reducing the number of listings to different values of K and γ. Panel A shows the benchmark effect when the number of product listings is reduced from 10,000 (default) to 5,000. Panel B makes the same comparison, but with K following a positive Poisson distribution with a mean of 5. Panel C and D make the same comparison when the price elasticity is 4 and 10, respectively. Section 6.5 describes the counterfactual exercises in more detail.

B Data and the Experiment

B.1 Additional Details on Measuring Product and Service Quality

Product Quality. We worked with a large local consignment store of children's clothing in North Carolina to inspect and grade the quality of each T-shirt. The owner has over 30 years of experience in the retail clothing business. Each T-shirt was given an anonymous identification number, and the owner was asked to grade the T-shirt on 8 quality dimensions, following standard grading criteria used in the textile and garment industry, as shown in Panel A of Figure A.1. In addition, the examiner was asked to price each T-shirt based on her willingness to pay and willingness to sell, respectively. T-shirts within the same variety were grouped together for assessment to make sure that the grading could capture within-variety variations. The examiner conducted two rounds of evaluation that took place several weeks apart to ensure consistency in grading.

Service Quality. To measure service quality, the following message was sent to sellers via the platform:

"Hi, I am wondering if you could help me choose a size that fits my kid, who is 5 years old, 45 lbs and about 4 feet. I would also like to know a bit more about the quality of the T-shirt. Are the colors as shown in the picture? Will it fade after washing? What is the material content, by the way? Does it contain 100% cotton? The order is a little urgent; how soon can you send the good? Would it be possible to expedite the shipping, and how much would that cost? Thanks in advance!"

B.2 Review Treatment

To generate the content of the reviews, we use the latent Dirichlet allocation topic model in natural language processing to analyze past reviews and construct the messages based on the identified keywords. Specifically, the following reviews were provided (randomly) to listings in T2:

Product Quality:

- "Great shirt! Soft, dense material, quality is good; color matching the picture exactly, and I am happy with the design; no problem after washing. My kid really likes it. Thank you!"
- "Well-made shirt. It was true to size. The material was very soft and smooth. My kid really likes the design. I am overall satisfied with it."
- "This shirt is nice and as seen in the photo. It fits my kid pretty well. The material is quite sturdy and colorfast after washing."

Shipping Quality:

- "The shipping was pretty good. Package arrived within the estimated amount of time and appeared intact on my porch."
- "I am pleased with the shipping. It was fast and easily trackable online. The delivery was right on time, and the package appeared without any scratches."
- "Fast delivery and convenient pickup, everything is smooth, shirt came in a neat package, not wrinkled. Thank you!"

We left positive reviews on all listings unless there were obvious quality defects or shipping problems, in which case no review was provided.

C Theory Appendix: Proof of Proposition 1

Since the full proof below is technically involved, we first explain the simple intuition here. The key observation is that for a consumer to obtain higher quality than expected under her prior, it is necessary that she samples a seller chosen in previous periods so as to benefit from past reviews. In light of this feature, expected quality is related to the probability of re-sampling a seller. In early periods, this probability is proportional to $\frac{1}{N}$ and thus decreases with the number N of sellers.

Turning to the formal proof, it suffices to study the expected quality in period T. We compute this expectation by adopting the subjective perspective, which involves averaging across different histories the *belief* in period $T-1$ about the seller's quality in period T. By the law of iterated expectations, this average is indeed the ex ante expected quality.

Notice that each possible history of the first T periods can be described by the following:

- the sample $(i_t^1, \ldots, i_t^K)$ in each period $1 \leq t \leq T$;
- an index $k(t) \in \{1, \ldots, K\}$ in each period $1 \leq t \leq T$ describing which of the K sellers is chosen out of the sample;
- conditionally independent signal realizations z_t about the true quality of seller $i_t^{k(t)}$ chosen in each period t, where $z_t = q^{i_t^{k(t)}} + \mathcal{N}(0, \sigma^2)$.

These variables, which we denote by $\mathcal{H}$, are sufficient to pin down the evolution of sales $\{s_t^i\}$ and beliefs $\widehat{q}_t^i$. It turns out to be convenient to ignore the last variables $k(T)$ and z_T and compute the expected quality in period T conditional on what happens *before* a choice is made in period T. Thus, in what follows, when we refer to a "history," we exclude $k(T)$ and z_T.

Crucially, the likelihood of any such history can be explicitly written as the following product:

$$L(\mathcal{H}) = \left(\prod_{t=1}^{T}\prod_{k=1}^{K} \frac{(v_0 + s_{t-1}^{i_t^k})^\lambda}{\sum_{j=1}^{N}(v_0 + s_{t-1}^{j})^\lambda}\right) \cdot \left(\prod_{t=1}^{T-1} \frac{exp(\widehat{q_{t-1}^{i_t^{k(t)}}})}{\sum_{k=1}^{K} exp(\widehat{q_{t-1}^{i_t^k}})}\right) \cdot l\left(z_1, \ldots, z_{T-1} \mid \{i_t^{k(t)}\}_{t \leq T-1}\right). \quad \text{(C.1)}$$

The first multiplicative factor above captures the probability of generating each K-sample (based on initial visibility and sales). The second factor is the probability of choosing the particular seller out of the sample (based on the logit rule applied to beliefs). The last factor, $l(z_1, \ldots, z_{t-1} \mid \cdot)$, represents the probability of seeing the signal realizations z_t (based on the normal prior and signals). The product of these factors is the likelihood of a given history, which is the weight that we use to average across different histories.

Note also that given $\mathcal{H}$, the believed quality of the seller chosen in period T is completely determined by the sample $(i_T^1, \ldots, i_T^K)$ in period T and the beliefs about these sellers at the end of period $T-1$. This believed quality can be written as

$$f(\mathcal{H}) = \sum_{j=1}^{K} \widehat{q_{T-1}^{i_T^j}} \cdot \frac{exp(\widehat{q_{T-1}^{i_T^j}})}{\sum_{k=1}^{K} exp(\widehat{q_{T-1}^{i_T^k}})}. \quad \text{(C.2)}$$

Hence, the ex ante expected quality in period T can be computed as the integral

$$\int f(\mathcal{H}) \cdot L(\mathcal{H})\, d\mathcal{H}.$$

Below, we decompose this integral into 3 parts, corresponding to 3 different kinds of histories $\mathcal{H}$:

(1) First, consider any history $\mathcal{H}$ where all sellers sampled in period T have *not* been previously chosen (i.e., $i_T^j \neq i_t^{k(t)}$ for all j and all $t < T$). In this case, all these sellers are believed to have expected quality 0, just as in the prior. It follows that $f(\mathcal{H}) = 0$, and so we can ignore such histories in computing the above integral.

(2) We then consider histories where all $K \cdot (T-1)$ sellers sampled before period T are distinct but there is a unique seller sampled in period T that coincides with a previously chosen seller. The other $K-1$ sellers sampled in period T are all distinct from the previously sampled $K \cdot (T-1)$ and distinct from each other.

Ignoring the signals for the moment, the total likelihood/probability of generating samples of this form is

$$\frac{K(T-1)(v_0+1)^\lambda \left(\prod_{s=0}^{KT-2}(N-s)(v_0)^\lambda\right)}{\prod_{t=1}^{T}((N-t+1)(v_0)^\lambda + (t-1)(v_0+1)^\lambda)^K}.$$

To understand this expression, note that for the sample in period 1 to consist of distinct sellers, we can arbitrarily draw i_1^1 but can only draw i_1^2 with probability $\frac{(N-1)(v_0)^\lambda}{N(v_0)^\lambda}$, i_1^3 with probability $\frac{(N-2)(v_0)^\lambda}{N(v_0)^\lambda}$ and so on. Similar conditional probabilities apply to all sampled sellers before period T and to all but one of the sellers sampled in period T. The remaining term $\frac{K(T-1)(v_0+1)^\lambda}{(N-T+1)(v_0)^\lambda+(T-1)(v_0+1)^\lambda}$ in the above expression is the probability of the only seller sampled in period T that repeats a previously chosen seller—K here is the possible positions of this seller in the period T sample, $T-1$ is the number of previously chosen sellers that can be repeated, and v_0+1 is the visibility of a previously chosen seller.[42]

We now take into account the signals before period T. Only one of those signals is relevant for what happens in period T, and that is the signal about the particular seller i that is repeated in the period T sample. This signal $z = q^i + \mathcal{N}(0, \sigma^2)$ leads to the belief $\widehat{q^i_{T-1}} = \frac{z}{1+\sigma^2}$ by Bayes's rule. Since $q^i \sim \mathcal{N}(0,1)$, it is easy to see that the ex ante distribution of the belief $\widehat{q^i_{T-1}}$ is normal with mean 0 and variance $\frac{1}{1+\sigma^2}$. For the remaining $K-1$ sellers sampled in period T, their beliefs are zero, as in the prior.

Thus, given the samples and the belief $\widehat{q} = \widehat{q^i_{T-1}}$ about the special seller i, the believed quality in period T can be computed as $f(\mathcal{H}) = \widehat{q} \cdot \frac{exp(\widehat{q})}{K-1+exp(\widehat{q})}$. Integrating over $\widehat{q}$, we obtain that given any collection of samples in the first T periods that repeat only one seller (in period T), the believed quality in period T is

$$\eta = \mathbb{E}\left[\widehat{q} \cdot \frac{exp(\widehat{q})}{K-1+exp(\widehat{q})} \mid \widehat{q} \sim \mathcal{N}(0, \frac{1}{1+\sigma^2})\right] > 0.$$

This is positive because $\widehat{q} \cdot \frac{exp(\widehat{q})}{K-1+exp(\widehat{q})} + (-\widehat{q}) \cdot \frac{exp((-\widehat{q}))}{K-1+exp((-\widehat{q}))} > 0$ whenever $\widehat{q} \neq 0$.

To summarize, for samples in the first T periods that have the "repeat only once" property, their contribution to the expected quality in period T is

$$\eta \cdot \frac{K(T-1)(v_0+1)^\lambda \left(\prod_{s=0}^{KT-2}(N-s)(v_0)^\lambda\right)}{\prod_{t=1}^{T}((N-t+1)(v_0)^\lambda + (t-1)(v_0+1)^\lambda)^K}.$$

The specific expression does not matter; what is important is that we can rewrite this contribution as

$$\frac{P(N)}{Q(N)}$$

for some polynomials P and Q with positive leading coefficients and degrees $KT-1$ and KT, respectively. This ratio formalizes the intuitive idea that the probability of repeating one seller

[42] In this probability calculation, we do not worry about $k(t)$, the positions of the previously chosen sellers. This is without loss because we assume that the sellers sampled before period T are all distinct and thus completely symmetric.

in the samples is on the order of $\frac{1}{N}$.

(3) In all remaining histories, the KT sellers sampled in the first T periods represent at most $KT - 2$ distinct sellers (i.e., there are at least two repetitions in the samples). We show that the contribution of these histories to the period T expected quality can be written as a finite sum of ratios

$$\sum_m \frac{R_m(N)}{S_m(N)},$$

where each $R_m(N)$ is a polynomial with degree at most $KT-2$ and each $S_m(N)$ is a polynomial with degree KT. Again, the broad intuition is that the probability of "repeating twice" is on the order of $\frac{1}{N^2}$.

More formally, let us consider a generic "collection" of samples $\{i_t^j\}_{1\leq j\leq K, 1\leq t\leq T}$ and chosen seller positions $\{k(t)\}_{1\leq t\leq T-1}$, representing a set of histories in which the signal realizations are random. If we permute the labeling of all N sellers, then the indices in the samples are relabeled accordingly. However, due to ex ante symmetry, the resulting set of histories makes the the same contribution to the period T expected quality as the original set. Thus, to compute the total contribution of all possible "collections," we just need to compute the contributions of different collections that cannot be relabeled into each other and then do a weighted sum with weights given by the number of relabelings associated with each collection.

The benefit of this approach is that modulo relabeling, we are essentially concerned with the patterns of repetition among KT sampled sellers. The number of such patterns depends on K, T but not on N, and so does the number of collections that cannot be relabeled into each other (the latter number is K^{T-1} times larger since a collection also specifies chosen sellers). On the other hand, for any fixed collection in which the samples represent $d \leq KT - 2$ sellers, the number of possible relabelings is simply $\prod_{s=0}^{d-1}(N - s)$, which is a polynomial of degree at most $KT - 2$. Thus, if we could show that the contribution of any fixed collection can be written as $\frac{c}{S(N)}$ for some constant c and some polynomial $S(N)$ with degree KT, then the weighted sum of such contributions would have the desired form $\sum_m \frac{R_m(N)}{S_m(N)}$ with $deg(S_m) = KT$ and $deg(R_m) \leq KT - 2$.

Now, for a fixed collection, we know that the sales evolution has been determined. Thus, the first part on the RHS of (C.1) is fixed and has the form $\frac{c_1}{S(N)}$ for some constant c_1 and some polynomial $S(N)$ with degree KT. Using (C.1) and (C.2), we can write the contribution of this collection as

$$\frac{c_1}{S(N)} \cdot \int_{z_1,\ldots,z_{T-1}} \left(\prod_{t=1}^{T-1} \frac{exp(\widehat{q_{t-1}^{i_t^{k(t)}}})}{\sum_{k=1}^{K} exp(\widehat{q_{t-1}^{i_t^k}})} \right) \cdot l\left(z_1,\ldots,z_{T-1}\right) \cdot \left(\sum_{j=1}^{K} \widehat{q_{T-1}^{i_T^j}} \cdot \frac{exp(\widehat{q_{T-1}^{i_T^j}})}{\sum_{k=1}^{K} exp(\widehat{q_{T-1}^{i_T^k}})} \right),$$

where we recall that the beliefs $\widehat{q_t^i}$ can be expressed in terms of the signal realizations z_t. The integral above is another finite constant c_2 independent of N, as we desire to show.[43]

We now put together the 3 kinds of histories studied above to prove Proposition 1. The previous analysis allows us to deduce that the expected quality in period T can be written as

$$\frac{P(N)}{Q(N)} + \sum_m \frac{R_m(N)}{S_m(N)}.$$

Simple calculus shows that the derivative of $\frac{P(N)}{Q(N)}$ with respect to N is negative for large N and on the order of $\frac{1}{N^2}$ (just like the derivative of $\frac{1}{N}$). In contrast, the derivative of each $\frac{R_m(N)}{S_m(N)}$ may be positive or negative but, in either case, is at most on the order of $\frac{1}{N^3}$. Thus, the derivative of the overall sum $\frac{P(N)}{Q(N)} + \sum_m \frac{R_m(N)}{S_m(N)}$ is also negative for large N. In words, when N is sufficiently large, the expected quality in early periods decreases with N, completing the proof.

D Details of the Structural Estimation

D.1 Procedures for Computing Simulated Moments

For each set of structural parameters, we conduct the following procedure to compute the related simulated moments.

Recover Marginal Cost. In the first step, We use the data empirical distributions of price, review, and cumulative orders to recover the distribution of costs, F_c, relying on the set of first-order conditions from the sellers' static pricing problem that is described in Section 6.2. We simulate demand $D_i(\mathbf{p}, \mathbf{r}, \mathbf{s})$ and the demand derivative $\frac{\partial D_i}{\partial p_i}(\mathbf{p}, \mathbf{r}, \mathbf{s})$ based on Equation (3).

[43] To see that the integral is finite, we interpret it as the expectation of the following function of beliefs:

$$\left(\prod_{t=1}^{T-1} \frac{exp(\widehat{q_{t-1}^{i_t^{k(t)}}})}{\sum_{k=1}^{K} exp(\widehat{q_{t-1}^{i_t^k}})} \right) \times \left(\sum_{j=1}^{K} \widehat{q_{T-1}^{i_T^j}} \cdot \frac{exp(\widehat{q_{T-1}^{i_T^j}})}{\sum_{k=1}^{K} exp(\widehat{q_{T-1}^{i_T^k}})} \right).$$

This function is bounded in absolute value by $\sum_{j=1}^{K} |\widehat{q_{T-1}^{i_T^j}}|$, which has a finite expectation because the beliefs $\widehat{q_{T-1}^{i_T^j}}$ all have a normal distribution. Thus, the dominated function has a finite expectation as well.

Initialize Sellers in the Market. We initialize the market by setting the cumulative orders of sellers at 0 and the visibility of sellers at $v_0 = s_0 > 0$. In addition to the marginal distribution of costs F_c obtained in step 1 and the standard normal marginal distribution of quality, we use the Gaussian copula to model their dependence. Specifically, we draw the tuple (q, c) for each seller according to the following formula:

1. Draw a vector $\mathbf{Z}$ from the multivariate standard normal distribution with correlation ρ,
$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$
2. Calculate the standard normal CDF of $\mathbf{Z}$:
$$U_1 = \Phi(Z_1), \quad U_2 = \Phi(Z_2).$$
3. Transform the CDF to quality and cost values using their marginal distributions:
$$c_{\text{draw}} = F_C^{-1}(U_1), \quad \text{q}_{\text{draw}} = \Phi^{-1}(U_2) = \Phi^{-1}(\Phi(Z_2)) = Z_2.$$

After drawing the cost and quality for each seller, we solve their static pricing problem to set the initial prices. Finally, we randomly assign sellers to variety groups, making sure that the number of variety groups as well as the distribution of the number of product listings in each group match the data.

Simulate Order and Review. In each period, we use the weighted sampling without replacement to generate the consumer's sample of size K. Based on its average reviews, we calculate each sampled seller's expected quality and the expected utility of purchasing. Since sellers in the same variety group share the same logit shock, only the seller with the highest inclusive value may be chosen when multiple sellers from the same variety group get sampled. Therefore, we form a subset of the sampled listings by removing dominated product listings. Then, we simulate the purchasing decision based on standard logit probability, the realized experience for the consumer, and the review that he or she leaves. At the end of each period, we update the cumulative orders and the average review for the seller that has made a new sale. In addition, we allow the sellers to update their prices by solving the static pricing problem at the frequency that matches the observed frequency of price adjustment.

Simulate Moments. Starting from the initialized market, we simulate the arrival of $T = 620,000$ consumers so that the total number of fulfilled orders matches that in the data. We use the endline simulated data to calculate the distribution of cumulative revenue for the sellers, the regression coefficient of log price and quality, and the share of cumulative orders accounted for by high-quality sellers. We also simulate an additional ΔT periods from the endline to compute the dependence of sellers' new order arrival on cumulative orders.

D.2 Weighting Matrix and Objective Function

We bootstrap our data sample moments 1,000 times and construct the weighting matrix W. The objective function used for optimization is

$$Q(\theta) = (g_0 - \gamma_m(\theta))' \, W \, (g_0 - \gamma_m(\theta)) \, ,$$

where g_0 is the data moments vector, $\gamma_m(\theta)$ is the simulated moments vector based on $m = 100$ simulations, and $\theta = (v_0, \sigma, \rho, \lambda)$ is the vector of parameters.

D.3 Consumer Surplus Calculations

Without information frictions, the consumer surplus (in dollars) can be computed using the standard log sum formula

$$E(CS) = \log\left(\sum_{k=1}^{K} \exp\left(\widehat{q^{i_k}} - \gamma p^{i_k}\right)\right),$$

where $(i_1, i_2, \ldots, i_K)$ is the consumer's realized consideration set.

With information frictions, consumer surplus takes a more complicated form because the beliefs under which purchasing decisions are made are different from the truth. Leggett (2002) develops a solution to this problem for type-I extreme value random utility errors. In particular, the adjusted formula for consumer surplus realized from a consideration set $(i_1, i_2, \ldots, i_K)$ is

$$E(CS) = \log\left(\sum_{k=1}^{K} \exp\left(\widehat{q^{i_k}} - \gamma p^{i_k}\right)\right) + \sum_{k=1}^{K} \tilde{\pi}_{i_k}\left(q^{i_k} - \widehat{q^{i_k}}\right),$$

where

$$\tilde{\pi}_{i_k} = \frac{\exp\left(\widehat{q^{i_k}} - \gamma p^{i_k}\right)}{1 + \sum_{k=1}^{K} \exp\left(\widehat{q^{i_k}} - \gamma p^{i_k}\right)}.$$

The second term in the consumer surplus formula takes into account the fact that purchasing decisions are made under the current beliefs $(\widehat{q^1}, \widehat{q^2}, \cdots, \widehat{q^N})$ whereas the true underlying quality is $(q^1, q^2, \ldots, q^N)$.

D.4 Alternative Specification of Learning

We examine the possibility that, in addition to the reviews, consumers could use each seller's cumulative sales as an additional signal to infer her fundamental quality. To facilitate consumer learning from cumulative sales, we assume that the empirical distribution of cumulative sales conditional on seller quality is revealed to consumers after every T_0 periods of market evolution. We assume that this empirical distribution can be approximated by a normal distribution with mean $\alpha_0 + \alpha_1 q^i$ such that

$$s_t^i \sim N(\alpha_0 + \alpha_1 q^i, \alpha_1^2 \sigma_s^2).$$

While consumers do not observe true quality q^i, they understand the parameters $(\alpha_0, \alpha_1, \sigma_s^2)$ that govern this relationship. Note this is obviously a strong assumption, but it helps to provide an "upper bound" for the potential importance of s_t^i as a separate signal of quality q^i.

Taking into account both the consumer reviews $\bar{z}_t^i$ and the additional signal $\frac{s_t^i - \alpha_0}{\alpha_1} \sim N(q^i, \sigma_s^2)$, we can define the posterior mean quality of each seller i as

$$\hat{q}^i(\bar{z}_t^i, s_t^i) = \frac{\bar{z}_t^i \cdot s_t^i/\sigma^2 + \frac{s_t^i - \alpha_0}{\alpha_1 \sigma_s^2}}{1 + s_t^i/\sigma^2 + 1/\sigma_s^2}.$$

Since we are not adding any new parameters, we can re-estimate the model using the same procedure. In Table A.17, we report the new parameter estimates and model moments. In addition, we show results, using the new estimates, for the counterfactual exercise that randomly removes 5,000 listings from the original 10,000 listings. They are remarkably similar to the baseline results in Table 8.

D.5 Alternative Specification of Sampling Probability

In our baseline model estimation, we focused on each seller's cumulative sales as the predominant factor that determines visibility and the sampling probability. We made this modeling choice to align with our RCT findings and the novel theoretical results.

In Table A.18, we show that our core theoretical mechanism carries through if we generalize our sampling weight to include additional observed seller characteristics such as reviews. In particular, we assume that each seller i is sampled based on both its cumulative sales s_i and average reviews $\bar{z}_i$; i.e., the sampling weight is now $(v_0 + s_i)^\lambda \cdot \exp(\zeta \bar{z}_i)$. Our baseline model essentially assumes $\zeta = 0$. We investigate the cases of $\zeta = 0.1$ and $\zeta = 0.2$ and calibrate the rest of the parameters to match the same set of data moments. Table A.18 shows that we need a lower value of λ and a larger σ than those in our baseline model when reviews also enter the sampling weights. This is intuitive, as reviews bring additional information and could speed up the transition of market allocation towards high-quality sellers. To rationalize the same concentration and allocation observed in the data, the model needs to weaken the role played by cumulative sales as well as the information content of each additional signal.

Nevertheless, we still find substantial improvement in allocation and average consumer surplus by reducing the number of sellers from $10,000$ to $5,000$. This indicates that our central theoretical mechanism still plays an important role under an alternative specification of sampling probabilities. The key policy lessons discussed in Section 7 remain robust.