Stand Out from the Millions: Market Congestion and Information Friction 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

September 15, 2025

Abstract

We investigate how market congestion and information friction affect firm dynamics and market efficiency in global e-commerce. Observational data and self-collected quality measures from AliExpress suggest significant demand frictions and potential misallocation in the online market. A randomized experiment that offers new exporters exogenous demand and information shocks demonstrates the limited ability of existing platform mechanisms to help small sellers overcome the demand frictions. We show theoretically and quantitatively that having a large number of market participants undermines the functioning of existing online mechanisms and hinders the discovery of high-quality sellers. Policy counterfactuals highlight that blanket-wide onboarding initiatives can aggravate market congestion, slow down the resolution of the information problem, and result in market misallocation.

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

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), 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 $4.9 trillion and 19% of total global retail sales in 2021. 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]

While e-commerce potentially exposes prospective exporters to buyers around the world, important frictions remain beyond the initial entry point. First, it has been well-documented that consumers consider a very small subset of all products available in the online market. A rising number of sellers can then result in market congestion as sellers compete for limited consumer attention. Second, information friction is prevalent in the online market. Consumers often cannot perfectly assess the underlying quality of the sellers and have to rely on online reviews as noisy signals of the true quality.[4] Our study examines how these demand frictions jointly shape firm dynamics and market efficiency in global e-commerce, and how these frictions influence the effectiveness of government onboarding initiatives. Our paper shows that market congestion can undermine the functioning of the platform in resolving information friction and hinder the discovery of high-quality sellers. As a result, blanket-wide onboarding may not be able to generate sustained SME growth, and can in fact exacerbate market congestion, slow down the resolution of the information problem, and worsen market allocation.

[1] `https://www.emarketer.com/content/global-ecommerce-forecast-2022`

[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 on-boarding and market entry of the SMEs.

[4] Tadelis (2016) provides a review of the literature on online review mechanisms.

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 operating in this segment, including detailed seller-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 assessment.

We begin by documenting a set of stylized facts about the global online marketplace. First, using objective measures of seller quality, we show that quality only weakly predicts sales. The "superstars," which we define as the largest seller in each product variety, do not necessarily command higher quality compared to the small listings in the same variety group. This provides suggestive evidence for the presence of important market frictions and potential market misallocation. Next, we examine the role of the platform in potentially helping high-quality sellers overcome these frictions and reducing market misallocation. Using transaction-level observational data, we document that past sales strongly predict the arrival of future sales, consistent with the common perception that accumulating sales and reviews helps boost a firm's visibility via the online ranking and review mechanisms. Such a *demand reinforcement* force can potentially allow high-quality exporters to overcome the frictions and grow by accumulating initial sales and (positive) reviews on the platform, which then generate more future sales and reviews. Firms' growth dynamics and evolution of market share allocation would hinge on the strength of the reinforcement force.

However, it is difficult to identify the strength of the demand reinforcement force through observation data alone. The positive relationship between past and future sales could be driven by unobserved seller effort: for example, larger sellers may spend more on advertising and display, which directly leads to more future sales. To address the identification challenge, we conduct an experiment that generates exogenous demand and information shocks to a set of small exporters via randomly placed online purchase orders and reviews. We find that the order treatment does lead to a significantly positive impact on sellers' subsequent sales, causally establishing the demand reinforcement channel. That said, although the demand reinforcement mechanism exists, we find that its strength is rather weak. In particular, the size of the estimated average treatment effect is much smaller than the size of the initial treatment. We also do not find any significant treatment effect from the reviews nor any heterogeneous

treatment effect based on quality. Taken together, the experimental findings highlight the limited success of the existing platform mechanism in helping sellers, especially high-quality new sellers, to overcome the market frictions and grow.

Motivated by the reduced-form evidence, we develop a theoretical model that incorporate key features of the online market—importantly the demand frictions and the demand reinforcement mechanism, and use the model to examine the roles of market congestion and information friction in determining market dynamics and efficiency properties. The model extends the classical Polya urn model by incorporating consumer choice and seller heterogeneity in quality. In particular, we model the probability of a seller entering into a consumer's consideration set as (proportional to) a power function of the seller's cumulative sales. This is a reduced-form way of capturing the idea that sellers with larger past sales and reviews receive higher visibility through the online ranking and review mechanisms.[5] Among the sellers within the consideration set, consumers make purchase decisions based on their expected qualities inferred from observable past reviews. We prove theoretically that when there are already many sellers in the market, further increasing the number of sellers (for example, through large-scale e-commerce onboarding initiatives) weakens the demand reinforcement force, therefore making it harder for high-quality sellers to accumulate sales and reviews. This leads to strictly worse allocation along the path of market evolution.

Building on the theoretical insights, we estimate a rich empirical model of the online market. We closely follow the setup of the theoretical model while incorporating product varieties, seller-side heterogeneity in both quality and cost and model sellers' pricing decisions. The estimates imply considerable market congestion and information friction on AliExpress. While demand reinforcement plays a role—striking a sale improves a seller's likelihood of entering into the consideration set—it takes time for high-quality sellers to accumulate sales, given the amount of market congestion. With the large estimated noise in review signals, uncertainty regarding quality resolves very slowly over time, i.e., only after a seller accumulates a substantial number of orders. Combined, the estimates highlight that market congestion, interacting with information friction, can constitute an important hurdle for the growth of high-quality prospective exporters. Using the experiment as a model-validation benchmark, we find quantitatively comparable average treatment effects when we simulate one-time demand shocks through the lens of the

[5]Following Goeree (2008), we consider the formation of consumer's consideration set to be a "reduced-form" representation of the 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. This is what we define as "market congestion".

model. The fact that our structural model matches well the experimental findings enhances its credibility for counterfactual analyses.

Using the estimated model, we first examine the impact of reducing the number of sellers on market allocation.[6] Consistent with our theoretical results, we find that doing so helps mitigate market congestion and allows high-quality sellers to be discovered faster, thereby improving allocative efficiency. Importantly, the improvement is especially strong when there exists large information friction. These comparative statics are robust to alternative parameter estimates and consumer consideration set formation.

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 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 market congestion present on 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 and 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 existing large marketplaces versus allocating budget to enhance the visibility of new marketplaces. Considering the latter as a viable policy alternative, our theoretical and empirical results highlight one important policy lesson in designing such a new marketplace: onboarding too many sellers can aggravate market congestion, slow down the resolution of the information problem, and result in greater market misallocation. On the other hand, alongside an onboarding initiative, injecting information to the market, such as through government or third-party screening and certification, can significant improve firm peformance and market outcomes.

Related Literature. Our work contributes to several strands of the existing literature. By studying demand frictions 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 market allocation (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

[6]Varying the number of market participants serves as an empirical counterpart to the key comparative statics result in the theoretical model. This is analogous to raising entry costs or the costs of maintaining active listings on the platform.

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) and establish large variations in firm-product quality in the online marketplace. However, we find that 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 market congestion and information friction in explaining the disintegration of the customer accumulation process and sellers' fundamental quality. We further quantify the scope of market misallocation through the lens of a rich empirical model.

Relatedly, 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 (Hui et al., 2022; Li et al., 2020). Theoretically, we formalize the process of consumers' consideration set formation and learning and derive the efficiency implications for short- and long-run market outcomes, highlighting the important interplay between the demand-side frictions and the number of market participants. Empirically, we bring in new sources of variations to first experimentally identify a demand-reinforcement force that can potentially help sellers overcome these frictions. We then formally model these realistic market frictions and quantify their impacts on market dynamics and efficiency. Methodologically, 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 improve firm visibility and help them overcome demand frictions in the market.

Third, our study also relates to the existing literature on consumer consideration sets (for example, Goeree, 2008; Honka et al., 2017; Dinerstein et al., 2018).[7] We estimate a reduced-form function that captures how cumulative sales boosts a listing's likelihood of being considered, as directly motivated by our experimental evidence.[8] We further incorporate a learning process, as enabled by the online review mechanism, to examine the interaction between market congestion and information friction.[9]

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

[8]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.

[9]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 past purchases affects subsequent growth of sellers.

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 190 countries and regions, attracting over 200 million monthly visits.[10] 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, are listed on the platform.[11] 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.[12]

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 over a third of the world's total textile and garment exports in 2019 (**?**). In the world of e-commerce, textile and apparel amount to 20 percent of China's total online retail, including sales on Alibaba's platforms.[13] The growth and efficiency of the online retail market therefore matters for upstream manufacturing: in

[10] Sources: `https://sell.aliexpress.com/`.

[11] 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.

[12] 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.

[13] "E-Commerce of Textile and Apparel," China Commercial Circulation Association of Textile and Apparel, 2019.

particular, growth of retailers that sell high-quality products in turn benefits their producers. 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.[14] 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 market congestion and information friction. The nature of this tradeoff is the key question that we seek to examine in this study.

Second, AliExpress allows us to group product listings into different *varieties*.[15] 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 first 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. Furthermore, we take the empirical variety groups distribution seriously in building the structure model. Doing so allows us to quantify the welfare tradeoff between the gains from varieties and the efficiency loss due to market congestion resulting from an increasing number of sellers in the marketplace.

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.

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

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

(1) Store-Listing-Level Data. We scraped nearly the full universe of product listings in the children's T-shirt segment in May 2018.[16] We collected all the information that a buyer can view on the 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, and 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 the 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) in order 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. All together, the measures cover multiple dimensions of quality, ranging from product to service to shipping quality.

To measure product quality, we placed actual orders for children's T-shirts on AliExpress.[17]

[16]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.

[17]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

After receiving and cataloging the orders, 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 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 placement 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.[18] 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

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 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.

[18]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.

detailed 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 individual quality metric.

To cross-validate these objective quality measures, we first examine the relationships between 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 (willingness to pay) and a resell price for each T-shirt. Reassuringly, the objective product quality metrics are strongly correlated with the price evaluations. Last but not least, to corroborate the measures of service quality, 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.[19]

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 of the presence of sizable demand frictions and market misallocation and, at the same time, point to the role of existing online mechanisms that can potentially help sellers overcome these frictions and grow.

Fact 1. *Sales performance varies within identical-looking variety groups.*

First, we exploit the grouping feature described in Section 2.1 that allowed us to group product listings into different identical-looking varieties to examine how sales performance varies within a single variety group. As shown in Figure **??**, we see that sales are concentrated at the top within each group. The 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; the top 25% capture nearly all sales (90.3%).

[19]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.

In light of the great amount of heterogeneity in sales performance, even within identical-looking variety groups, a key question to ask is who gets to become the superstars. In particular, do superstars command higher quality than non-superstars? And in general how much can quality predict a seller's sales performance in the online marketplace?

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

To examine these, we compare the quality of the superstar listings and small listings in each variety group using the objective quality measures described in Section 2.2. A superstar is defined as the listing with the highest sales in the group, and small listings are those with fewer than 5 cumulative orders. Panel A of Figure 3 plots the distribution of the difference in the overall quality index between the group superstar and the average of the small listings in each group. We observe a substantial fraction below zero: superstars actually 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 observations indicate the difficulties that high-quality sellers face in gaining market share. This indicates potential market misallocation. That said, this evidence is only suggestive because we have to take into account price differences.[20] To quantify the degree of misallocation, we rely on a structural model in Section 6.

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

Finally, we delve further into the growth dynamics and examine how superstars emerge. Using the census of children's T-shirt listings collected in May 2018 combined with 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 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

[20]Interestingly, we find that superstars do not always charge the lowest price, either: within a variety group, the listing with the highest sales charges the lowest price only 14% of the time. On the other hand, we do observe a positive relationship between price and quality, which corroborates our quality measures but could mean that this relatively flat relationship between quality and sales may be partly driven by price.

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.

The descriptive result is consistent with the common perception that accumulating sales and reviews helps boost a firm's visibility through existing online ranking and review algorithms, which then speeds up the arrival of future sales. The question is whether such demand reinforcement can allow new sellers, especially those of high quality, to overcome the demand friction and grow. Empirically, it is difficult to quantify the strength of the demand reinforcement using observational data due to unobserved supply-side actions. For example, it could be that sellers with higher past sales are taking more costly actions, such as paying for advertising or participating in promotion events organized by the platform, which directly attract more future sales. To overcome this identification challenge, we conduct an experiment in which we generates exogenous demand shocks to a set of small sellers via randomly placed online orders and reviews. The next section describes the experiment design and presents the main findings.

4 Experiment and Findings

To establish the strength of the demand reinforcement force, we experimentally enhance new sellers' visibility through positive online order and review treatments.

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 review on product and shipping quality.

Given that ratings are highly inflated on AliExpress,[21] for all the treatment groups, we leave a five-star rating for the order unless there are obvious quality defects or shipping problems. This is to mimic the behavior of actual consumers. To generate the contents of the shipping and

[21]Out of the 6,487 reviews that we observe over the six-month window from March to August 2018 in the transaction data, 85.9% are five stars.

product reviews, 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 orders. The difference between T1 and T2 identifies any additional impact of receiving reviews. 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 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: Effects of Demand and Information Shocks

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:

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

where the dependent variable is the total number of orders (excluding our own order) for listing i in week t.[22] 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. The specification leverages the panel structure of our data since the reviews were given only upon receipt of the orders. 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 these baseline controls are shown in Table A.8. Standard errors are clustered at the listing level.

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

[22]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.

United States, respectively. Overall, we see that the order treatment leads to a significantly positive impact on subsequent orders. This establishes the existence of demand reinforcement in the online market: exogenously receiving an order increases the speed of arrival of future orders. Tables A.10 and A.9 show that the order effect is likely to be mediated by a short-term boost in a listing's ranking and unlikely to be driven by endogenous supply-side responses.

Although a demand reinforcement force exists in the online market, we find its strength to be rather weak for small sellers as in the experimental sample. 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 demand frictions cannot be easily overcome by individual sellers' private efforts.[23]

Last but not least, we do not find any significant treatment effect of the reviews. There are two possible explanations: first, online reviews serve as only noisy signals of quality. Second, 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. The findings suggest that the online review system may not function effectively 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] The demand reinforcement force, while present, is not effective for helping high-quality sellers stand out from the market. This result echoes the earlier stylized fact that quality does not strongly predict sales performance in the market.

Taken together, the experimental findings highlight the limited success of existing platform ranking and review mechanisms in helping sellers, especially newly entered small businesses, to overcome the demand frictions and grow, and its limited success in aligning market allocation with seller's fundamental quality.

[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.

5 Theory

We now develop a theoretical model that formalizes the demand frictions and the role of demand reinforcement (via the ranking and reivew mechanisms) in the online market. We first characterize the market evolution process and then use the model to investigate the impact of increasing the number of market participants on firm dynamics and market efficiency.

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.[25] 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 seller's common initial visibility level. v_0 modulates the strength of demand reinforcement as it affects how much an additional sale boosts a seller's relative visibility.

[25] 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.16 that the alternative procedure of sampling without replacement generates nearly identical quantitative predictions.

Specifically, the sampling probability is modeled as

$$\frac{(v_t^i)^\lambda}{\sum_{j=1}^N (v_t^j)^\lambda}.$$

The exponent $\lambda > 0$ is another key parameter that moderates the strength of demand reinforcement.[26]

2. **Choice Procedure:** After forming the sample of K sellers, the consumer chooses to 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. If we let $\bar{z}_{t+1}^{i_k}$ be the average of the past $s_{t+1}^{i_k}$ reviews about seller i's quality q^{i_k}, up to and including period $t+1$, then the posterior mean of q^{i_k} at the end of period $t+1$ is given by

$$\widehat{q_{t+1}^{i_k}} = \frac{\bar{z}_{t+1}^{i_k} \cdot s_{t+1}^{i_k}/\sigma^2}{1 + s_{t+1}^{i_k}/\sigma^2}. \tag{2}$$

This familiar Bayesian updating formula represents the weighted average of the prior mean 0 and the past average review $\bar{z}_{t+1}^{i_k}$, with weights given by their respective precision levels 1 and $s_{t+1}^{i_k}/\sigma^2$.

The above fully describes the dynamics of our model, whose primitive parameters are $N, K, v_0, \lambda, \sigma$.

[26]Note that 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—as we show in Section 5.4, it is the value of λ that determines the long-run market outcome.

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 and favored by the demand reinforcement force. This departure from the classical model leads to fundamentally different market outcomes.[27]

A second remark is that we have presented the model in a way that is closest to our structural estimation. However, the theoretical analysis applies beyond the above specific functional forms. In particular, we can generalize the sampling procedure to make it depend on past reviews as well—for example, seller i is sampled with a probability proportional to $(v_t^i)^\lambda \cdot f(\bar{z}_t^i)$ for some positive function f. Our theoretical results continue to hold as long as f is increasing, so that higher-quality sellers are at least weakly favored by the sampling procedure.[28] Empirically, we also report the robustness check results of this sampling procedure in Appendix D.6.

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

5.3 Short-Run Market Outcomes

We begin by studying the short-run dynamics of the market, followed by a discussion of the long-run efficiency properties in the next subsection. We tie our theoretical analysis back with the big-picture policy motivation to focus attention on a key comparative static with respect to the number of sellers: while e-commerce lowers the entry barrier of exporting and brings many SMEs online, the presence of a large number of sellers can exacerbate congestion, given the limited size of consumers' consideration set. In what follows, we study the effect of increasing

[27]It is well known that in the classic Polya urn model, sellers' long-run market shares follow a full-support Dirichlet distribution, which is an inefficient outcome. In contrast, Proposition 2 below shows that consumer choice ($K \geq 2$) combined with suitable demand reinforcement ($\lambda = 1$) can achieve long-run efficiency. Moreover, Proposition 2 shows that the long-run market outcome is qualitatively different if we either strengthen or weaken demand reinforcement by adjusting the parameter λ. This provides additional flexibility for our model predictions to match observed data.

[28]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.

the number of sellers N on short-run market evolution.

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.*[29]

Thus, when there are already many sellers in the market, allocation worsens in the early periods as the number of sellers further increases. This result formalizes a key countervailing force to onboarding initiatives, where the entry of more sellers congests the sampling process and slows down the rise of high-quality sellers. 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 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.

The second channel above further suggests a potential interaction effect between the number of market participants and information friction. In Section 6.5, we perform counterfactual analysis to quantify the magnitude of this interaction effect and the first-order effect of a change in the number of sellers on market dynamics and welfare.

5.4 Long-Run Market Outcomes

We conclude the theoretical analysis with an examination of the long-run market outcomes. Our stylized facts document a weak relationship between quality and seller performance measured in cumulative sales. The question is then the following: in this environment, can efficiency be achieved in the long run and, if so, under what conditions? To formally define efficiency, we fix a profile of true qualities $q^1 > q^2 > \cdots > q^N$ and focus on the market share of the highest-quality seller 1. We say that the market is efficient in the long run if conditional on having the highest quality, seller 1's fraction of total sales $\frac{s_t^1}{t}$ converges in probability to 1 as $t \to \infty$. The following result shows that interestingly, long-run efficiency subtly hinges on the strength of demand reinforcement, as captured by the parameter λ:

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

Proposition 2. *Conditional on seller* 1 *having the unique highest quality, the long-run market outcome is*

1. *efficient if* $\lambda = 1$. *In this case, the convergence* $\frac{s_t^1}{t} \to 1$ *holds almost surely.*

2. *inefficient if* $\lambda > 1$. *In this case, every seller* i *has a positive probability of having all the sales, so that seller* 1 *may have zero market share in the long run.*

3. *inefficient if* $\lambda < 1$. *In this case, the market share of every seller* i *is bounded away from zero, so that seller* 1 *cannot occupy the entire market.*

When $\lambda > 1$, the demand reinforcement force is so strong that initial luck plays an excessively large role—every seller, not necessarily the one with highest quality, may be lucky in obtaining early sales and continue to be sampled and chosen in every period. This leads to persistent misallocation. On the other hand, when $\lambda < 1$, reinforcement is not strong enough for any seller to completely dominate the market—as soon as a seller's market share comes close to one, the probability that it will be sampled in the next period is not high enough to further increase its market share. The case of $\lambda = 1$ turns out to be just the right amount of reinforcement to guarantee that the highest-quality seller can not only overcome initial luck factors but also increase its market share all the way to the efficient benchmark.

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 demand frictions and the strength of demand reinforcement to overcome these frictions. The demand side closely follows the setup of the theoretical model. 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 sellers with replacement upon their arrival.[30] We allow for heterogeneity in the size of consumers'

[30]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

consideration set by assuming that K follows a positive Poisson distribution. Given K, the probability of each seller 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 i depends on the initial visibility parameter v_0 and cumulative sales s_t^i, reflecting the fact that products sold by larger sellers often appear in more pronounced positions on the platform. 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.

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 i's quality, denoted by $\widehat{q_t^i}$, follows the Bayesian updating rule as described in Equation (2). From there, we see that the expected quality $\widehat{q_t^i}$ at time t can be written as a function $\widehat{q^i}(\bar{z}_t^i, s_t^i)$, which depends on $\bar{z}_t^i$ (seller i's rating, or average past review) and s_t^i (seller 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 noisiness of the review signal).

Purchase and Review. We extend the baseline logit demand framework described in Section 5.1 to include prices and an outside option of nonpurchase with mean utility zero. Consumers' perceived utility of purchasing from seller 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:

$$U_t^i = \beta + \widehat{q^i}(\bar{z}_t^i, s_t^i) - \gamma p_t^i + \varepsilon_i,$$

where ε_i represents an idiosyncratic preference shock with an i.i.d. type-I extreme value distribution. β and γ are the constant and the price coefficient.

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. We also specify a pricing strategy that

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.

approximates the observed data. Each seller's pair (c^i, q^i) is drawn from a distribution upon the firm's entry on the online platform. We denote by ρ the correlation between c^i and q^i. However, 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.

As a result, 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 that it appears in the sample and is chosen by the consumer. Thus D_i depends on the rich set of public information $\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}$, which includes the prices, ratings, and cumulative sales of all sellers at the time of a price adjustment. For seller $i = i_1$, its perceived demand depends on all possible combinations of rivals $i_2, ..., i_K$:

$$D_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = K \sum_{i_2,...,i_K} \prod_{k=2}^{K} R_t^{i_k} \cdot \frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]}, \tag{3}$$

where $\widehat{q^i}$ is a shorthand for the expected quality $\widehat{q^i}(\bar{z}^i, s^i)$.

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

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

[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).

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

$$\frac{\partial D_i}{\partial p_i}(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = -K\gamma \sum_{i_2,\ldots,i_K} \prod_{k=2}^{K} R_t^{i_k} \left(\frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]} \right) \times \left(1 - \frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]} \right).$$

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. Sellers enter at the same time by paying a lump-sum entry cost. Upon entry, each seller obtains a random draw of quality q and cost c. Sellers then set their initial prices accordingly. We can recover the entry cost from the standard free entry condition by computing the discounted future payoff of the average entrant.

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 strength of reinforcement 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'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 sellers' online sales history, the data provide relatively little information on the variation in their cost over time. Thus, we start by calibrating γ to the average price elasticity of 6.7 (in line with the estimates in Broda and Weinstein (2006)) and calibrate β to match the market share of the outside option.[32] Another key 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).[33] Therefore, in

[32]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.

[33]Studies consistently find that in online marketplaces, the vast majority of consumers search very little

our baseline estimate, we assume that K follows a positive Poisson distribution with mean 2. Section 6.6 performs robustness checks with different parameter 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:

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 sellers' average cumulative orders reach the level in our data (31 per listing).

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 the initial visibility parameter v_0 and the strength of demand reinforcement λ. Intuitively, a small initial visibility v_0 increases the relative importance of early orders in a seller's life cycle. As a result, the amplification effect of cumulative orders is jointly determined by v_0 and λ. A smaller v_0 or a larger λ increases the concentration of the market distribution at the top. In addition, the dependence of a seller'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 seller's cumulative orders and measured quality identifies the review signal noise σ. If reviews were very precise, then higher-quality sellers would grow their orders 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.

---

and thus end up consider a very small subset of sellers (e.g., Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Wildenbeest, 2011).

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 the demand reinforcement parameter 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. Another way to interpret the magnitude of our estimated λ is to borrow the insight from Proposition 2 and contrast the estimated value with the case when $\lambda = 1$.[34] We find that if we set λ to 1, the top 1% sellers' cumulative sales share would increase from 34% to 43%. In addition, the top 1/3 quality bin would account for 49% of the orders (instead of 44% in our baseline). The seemingly small difference between our estimated $\hat{\lambda}$ of 0.97 versus 1 generates a nontrivial difference in market allocation.

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). Overall, our estimate suggests that reviews are very noisy signals of sellers' quality and that the uncertainty about each seller's quality is resolved very slowly, i.e., only after a substantial number of orders have accumulated. This indicates that the review mechanism takes time to play a role even if a seller emerges in a consumer's consideration set and successfully makes a sale. As a result, the information friction undermines the functioning of demand reinforcement in terms of aligning cumulative sales with sellers' fundamental quality. Finally, the estimate for ρ is 0.48. Given the empirical marginal distribution of costs and the standard normal quality distribution, this estimate 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 are able to 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 Model Validations

In this section, we evaluate our model's ability to rationalize the untargeted patterns of order arrivals documented in both the observational and experimental data.

[34] We use $\lambda = 1$ as a benchmark since, in the case of no endogenous pricing, Proposition 2 proves that the market achieves efficient allocation in the long run.

Figure 5 reports the model's predicted probability of receiving a new order within a week for sellers of different cumulative sales. Similar to Fact 3 in Figure 4, the probability rises steeply with past cumulative sales. For sellers with more than 100 past 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 our modeling of demand reinforcement, despite its simple structure, accounts for the salient features of order arrivals across sellers with a broad range of past cumulative sales.

Next, we show that our model's out-sample-prediction also fits 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 simulate the market forward for a number of periods that matches the overall growth in sales between the baseline and endline.[35] 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 effect size is always lower than the size of the initial treatment, which indicates that the demand frictions may not be easily overcome by sellers' private efforts of accumulating orders. Combining this with the previous discussion of demand reinforcement, we see that while such a force exists, its strength is most salient in the initial stage of the market; however, in a mature market congested by many established incumbents, it plays a limited role in facilitating the growth of newcomers.

6.5 Impact of Reducing the Number of Sellers

Using the estimated model, we examine the impact of reducing the number of sellers, N. This change is analogous to raising entry costs or the costs of maintaining active listings on the platform. Guided by our theoretical model, we first examine the expected quality of the chosen

[35] 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.

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 by half. The results are consistent with Proposition 1 and the related discussion: 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.

Table 8 summarizes the impact of reducing the number of sellers on allocative efficiency. 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[36] and examine the distribution of market shares for the top sellers using this metric. 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 15% ($= 0.61/0.53 - 1$) when the number of sellers is reduced from $10,000$ to $5,000$. Market shares for the top 25% and 33% also increase. As a result of the improved allocation, the expected consumer surplus increases by 7% ($= 0.74/0.69 - 1$) as shown in Column 4.[37]

Next, we examine the potential interaction effect between the number of sellers and information frictions. As discussed in Section 5.3, the efficiency implications of reducing the number of sellers may depend on the degree of information friction. To illustrate this, Panel B of Table 8 examines the impact of reducing N when σ is small (i.e., with more precise review signals). We see that the improvement in allocation from reducing N significantly diminishes. In other words, the market congestion is less harmful when quality can be revealed relatively quickly from purchases and reviews. On the other hand, the noisier the review signal is (the larger is information friction), the greater is the negative effect of increasing N.

Finally, 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 Appendix D.6, we show that market allocation and consumer surplus similarly improve with a reduction in the number of sellers when we generalize our sampling weights to $(v_0 + s_i)^\lambda \cdot \exp(\zeta \bar{z}_i)$.

[36]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.

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

6.6 Robustness Checks

Table A.16 reports estimated parameter values under alternative consideration set sizes and price elasticities. Table **??** examines the robustness of the counterfactual analyses under the different parameter values. For comparison, Panel A simply reproduces the baseline ($N = 10,000$) and counterfactual ($N = 5,000$) 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 with $N = 10,000$ are very similar to those in the baseline case (where the price elasticity is 6.7). The key comparative statics of reducing the number of sellers also remain robust.

7 Policy Analysis

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.[38] This shows that while demand reinforcement, through the online ranking and review mechanisms, can allow high-quality sellers to accumulate sales, this force is more salient in the initial stage of the market, which disproportionally benefits the earlier cohorts than the latecomers.

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).[39] 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.

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

[39] 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 market congestion and information friction affect the dynamics and efficiency of global e-commerce. While policy makers have increasingly emphasized the opportunities provided by global e-commerce, 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.

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

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

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.

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.

_ , Ali Hortaçsu, and Maria Ana Vitorino, "Advertising, consumer awareness, and choice: Evidence from the US banking industry," *The RAND Journal of Economics*, 2017, *48* (3), 611–646.

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.

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.

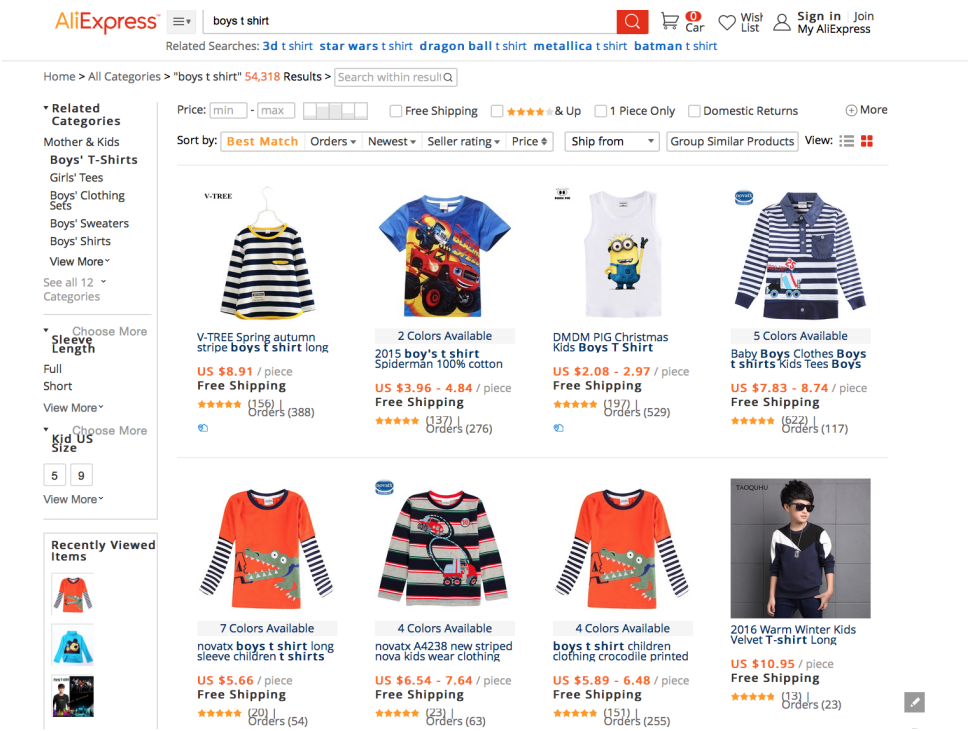
Ursu, Raluca M, "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 2018, *37* (4), 530–552.

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

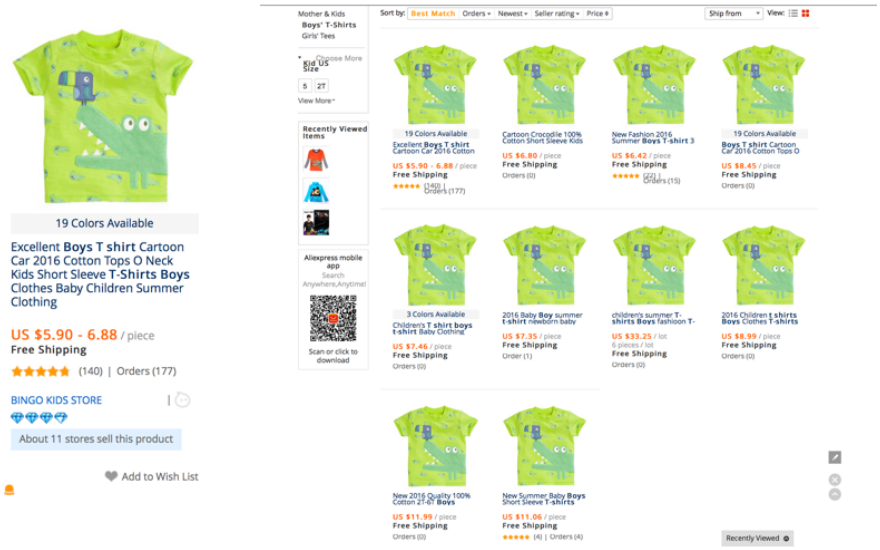
Wildenbeest, Matthijs R, "An empirical model of search with vertically differentiated products," *The RAND Journal of Economics*, 2011, *42* (4), 729–757.

Figure 1: AliExpress: Search Results with and without Grouping

Panel A. Search Results without the Grouping Function



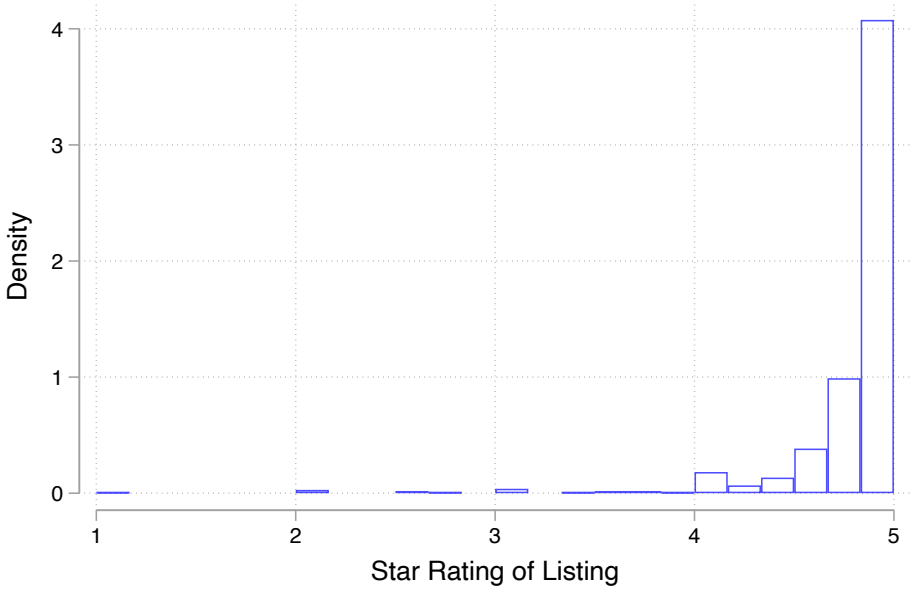
Panel B. Search Results with the Grouping Function



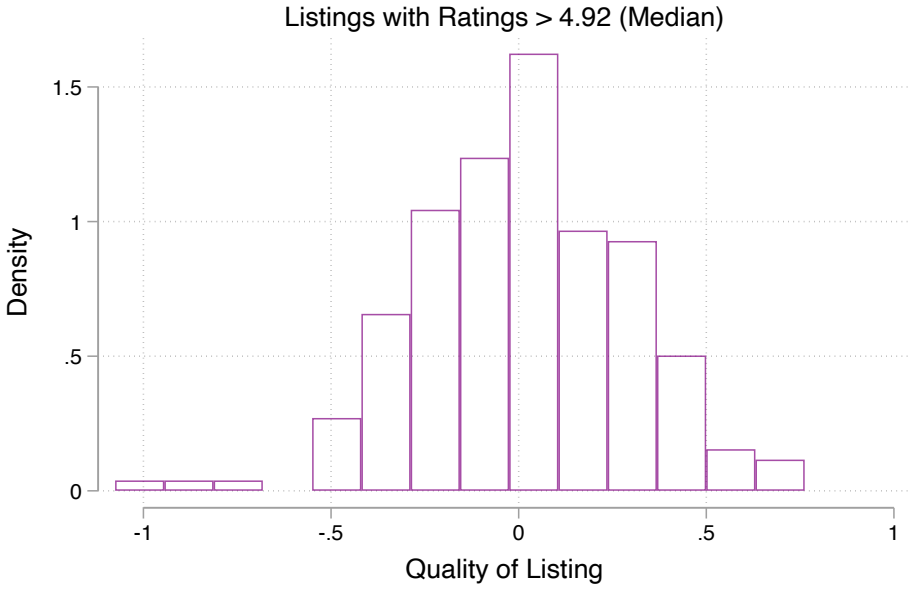
Note: This figure presents examples of search results on AliExpress. Panel A displays the search results of using "children's T-shirts" as keywords without applying the grouping function. Panel B displays the same search results with the grouping function applied.

Figure 2: Information Friction in the Online Marketplace

Panel A. Distribution of Online Star Rating



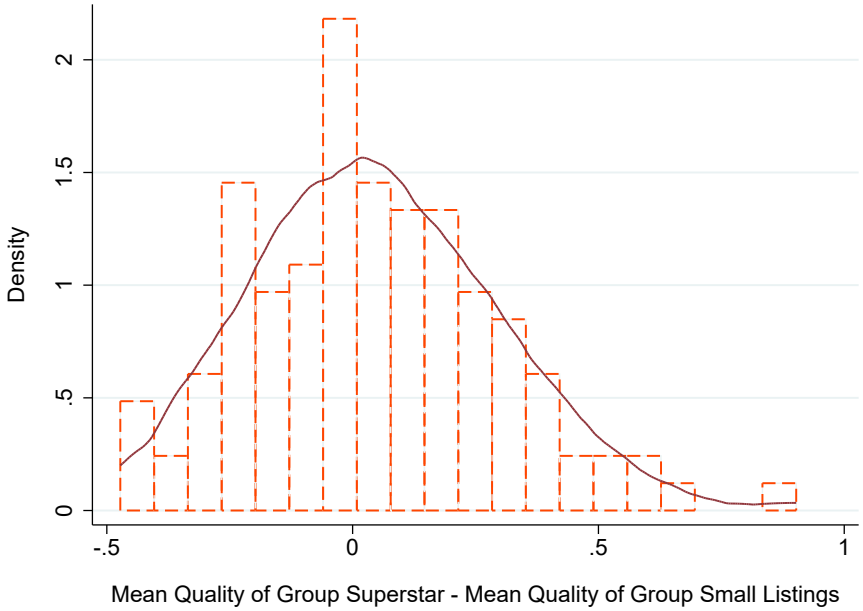
Panel B. Quality Distribution among Top-Rated Listings



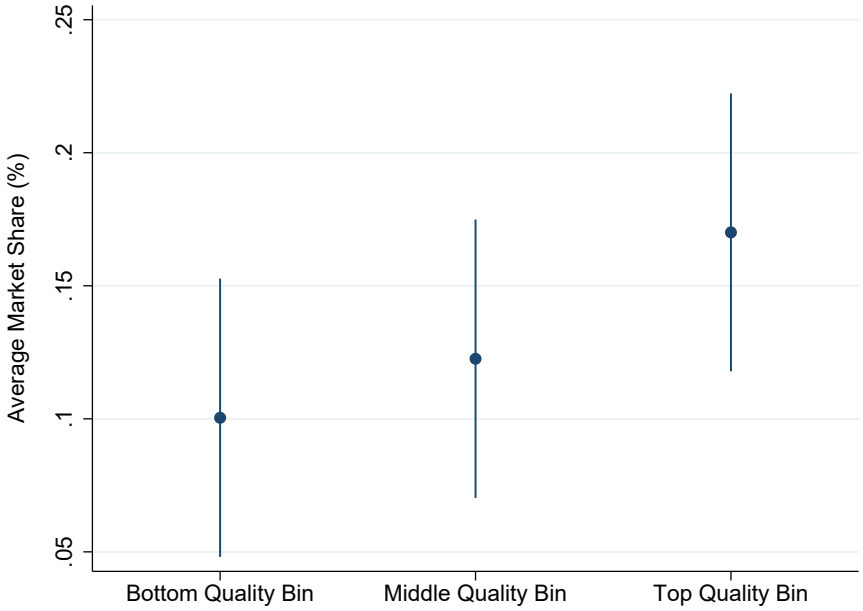
Note: This figure describes the existence of information friction in the online marketplace. Panel A plots the distribution of the online star rating, based on consumer-generated reviews, across all listings. Panel B plots the distribution of quality among listings with above-median star rating. Quality 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

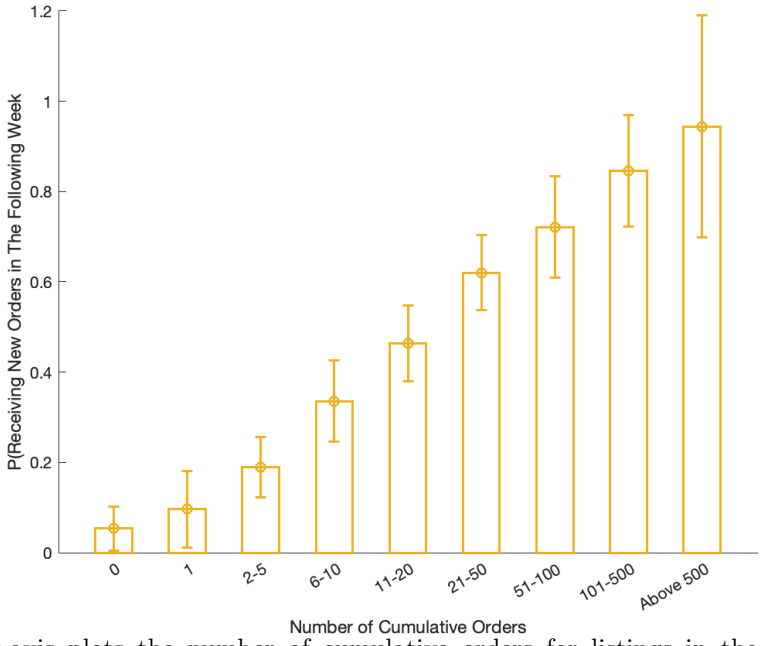


Panel B. Average Market Share over Quality Bins



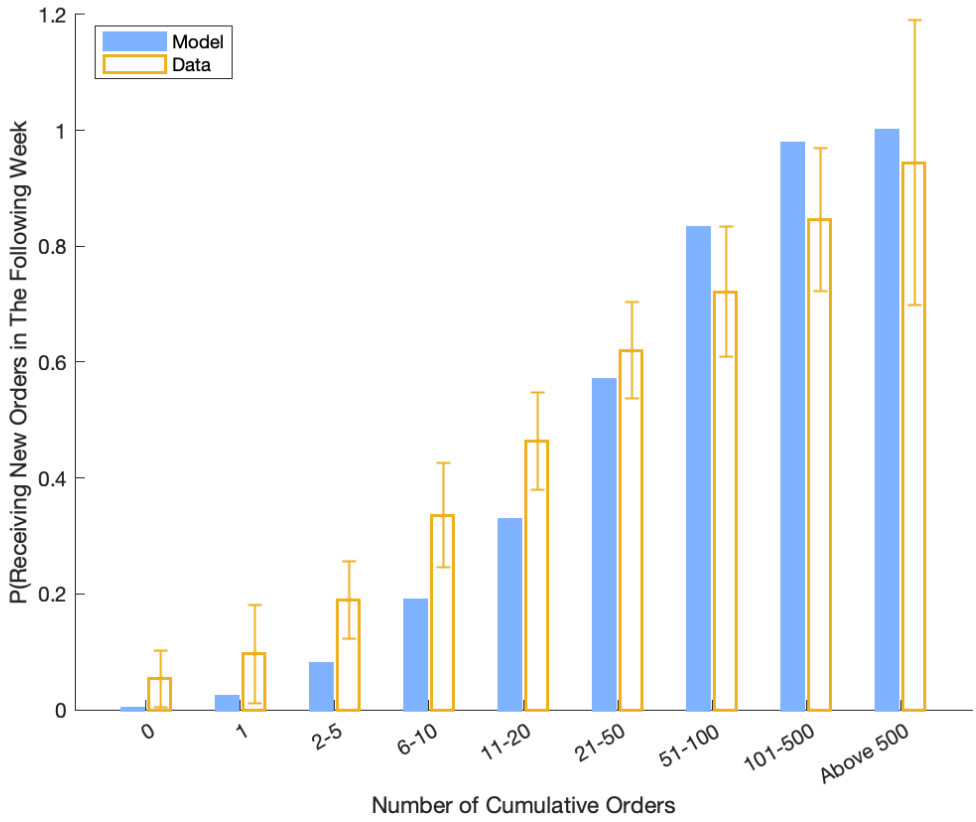
Note: This figure describes the relationship between listings' quality and sales performance. Panel A plots the distribution of the quality gap between group superstars and small listings in the group. Quality is measured in terms of the overall quality index (see Section 2.2 for details on construction 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. The t-test statistic for the group differences is -0.726 between the bottom and middle bins, -1.136 between the middle and top bins, and -1.758 between the bottom and top bins.

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.

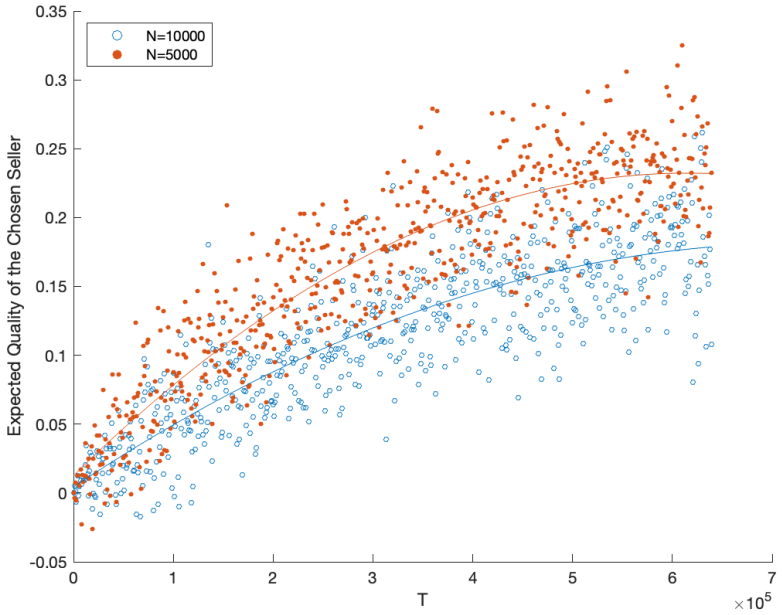
Figure 5: Model Validation - Order Arrival Pattern



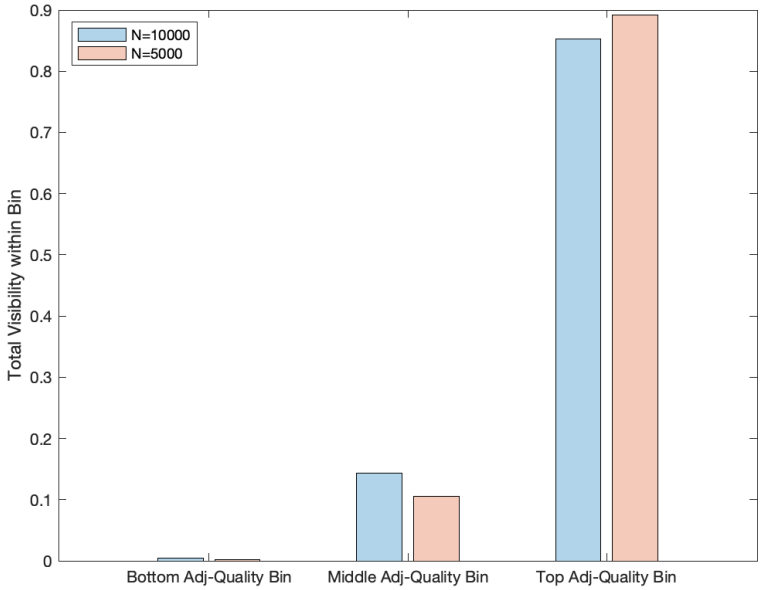
Note: This figure compares the simulated order arrival pattern in the model against its empirical counterpart. In the simulation, we first let the market reach the baseline size and then forward simulate for another week. We record the baseline number of cumulative orders for each listing, as well as 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
			Panel C. Group Level			
Group Size	133	9.46	4.62	9	3	18
Total Orders	133	421.87	480.73	262	103	1353
Total Revenue	133	1922.13	2044.29	1159.94	375.52	5530.58
Ave Price	133	5.44	2.3	5.2	3	8.53
Herfendahl Index	133	6345.52	2380.98	6313.9	2468.67	9930.79
Coeff of Variation of Orders	133	2.17	.62	2.03	1.35	3.31
Coeff of Variation of Revenue	133	2.16	.63	2.02	1.32	3.37
Coeff of Variation of Price	133	.26	.22	.21	.04	.63

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

	Weekly Number of Orders					
	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
Control Mean	0.066	0.066	0.015	0.015	0.004	0.004
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

	Endline Number of Cumulative Orders		
	All Destinations	English-speaking Countries	United States
	(1)	(2)	(3)
Order	0.096	0.186**	0.245***
	(0.231)	(0.076)	(0.058)
Review	0.444	0.078	-0.006
	(0.381)	(0.117)	(0.083)
Observations	784	784	784
Control Mean	.857	.197	.057
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	0.263	5.461	0.478	0.968
S.E.	(0.028)	(0.022)	(0.006)	(0.008)

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	0.324
Top 5% cumulative revenue share	0.608	0.597
Top 10% cumulative revenue share	0.745	0.730
Top 25% cumulative revenue share	0.898	0.891
Top 50% cumulative revenue share	0.974	0.973
2. Dependence of new order on cumulative orders (v_0)	0.103	0.136
3. Quality and sales relationship (σ)		
Cumulative orders share: Top 1/3 quality bin	0.434	0.424
Cumulative orders share: Middle 1/3 quality bin	0.311	0.313
4. Reg. coef. of log price and quality (ρ)	0.125	0.133
5. Experimental treatment effect	0.186	0.136

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

Table 7: Model Simulated Demand Shocks

Percent of Sellers Treated	Number of Initial Orders	Δ Ave Orders, when $\Delta M = 41.9\%$ (counterfactual - benchmark)	
P	O	Treated Sellers	Untreated Sellers
4	1	0.13	-0.01
4	2	0.25	-0.03
4	5	0.64	-0.08
4	10	1.32	-0.16

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 two columns report the change in the average cumulative orders among different groups of 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)	Variety Loss (%) (4)	Ave Seller Profit (5)	Platform Commission ($1K) (6)	Average Consumer Surplus (7)
			Panel A: With 10000 sellers				
$\sigma = 5.46$	0.51	0.78	0.87	0.0	26.7	56.45	0.672
$\sigma = 0$	0.90	0.97	0.99	0.0	33.3	74.97	1.064
			Panel B: With 5000 sellers and $\sigma = 5.46$				
by removing random listings	0.59	0.83	0.90	40.3	57.1	58.26	0.714
by removing listings from niche groups	0.69	0.83	0.90	93.4	57.0	58.04	0.708
			Panel C: With 10000 sellers, $\sigma = 5.46$				
$\lambda = 0.90$	0.45	0.73	0.84	0.0	25.7	54.30	0.621
$\lambda = 0.97$	0.72	0.90	0.94	0.0	29.5	63.05	0.813
$\lambda = 1.50$	0.69	0.98	0.99	0.0	21.3	43.74	0.512
$\lambda = 2.00$	0.50	0.84	0.92	0.0	16.9	33.48	0.378
$\lambda = 5.00$	0.42	0.73	0.85	0.0	13.7	28.00	0.302

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 5.46 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

	New Sellers							
	Ave Number of Orders	Ave Profits	Share for Top 10% Adj-Quality	Share for Top 25% Adj-Quality	Share for Top 33% Adj-Quality	Incumbent Sellers' Ave Profits	Platform Commission ($1K)	Consumer Surplus
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Onboarding to a Large Existing Marketplace								
1000 Seller-Listings								
Baseline Onboarding	0.138	0.14	0.20	0.46	0.59	34.2	60.18	0.797
Pre-Screening for High Quality	0.186	0.18	0.19	0.36	0.47	34.2	60.20	0.798
Pre-Screening for New Variety	0.140	0.14	0.20	0.47	0.60	34.2	60.19	0.798
Initial Quality Information Disclosure	0.146	0.14	0.20	0.48	0.62	34.2	60.21	0.798
Panel B: Onboarding to a New Marketplace								
1000 Seller-Listings								
0.1% Traffic	0.220	0.21	0.19	0.47	0.61		1.25	0.285
0.5% Traffic	1.325	1.26	0.23	0.52	0.67		5.88	0.289
1.0% Traffic	2.807	2.71	0.27	0.56	0.70		11.84	0.296

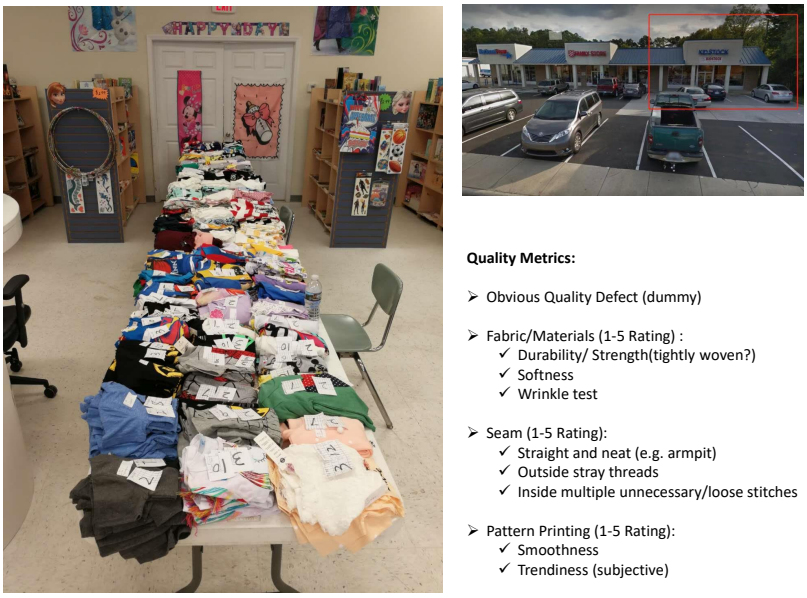
Note: This table simulates different scenarios of onboarding policies. Panel A reports the case in which 1,000 new sellers are onboarded to a large existing marketplace. We consider the following potential enhancements to a blind onboarding policy: (1) the platform conducts screening to ensure that the new sellers have the top 75% cost-adjusted quality among existing sellers, (2) the platform pre-screen to ensure that half of the new sellers carry new product varieties, and (3) the platform can partially reveal product quality by injecting the amount of information equivalent to the first five consumers reviews. The policy is simulated until market size doubles, marked by the arrival and purchasing decisions of 620,000 new consumers. Panel B reports the performance of the same 1,000 sellers when they are onboarded to a new marketplace, but under different flows of consumer traffic relative to the existing marketplace.

Appendices. For Online Publication Only

A Figures and Tables

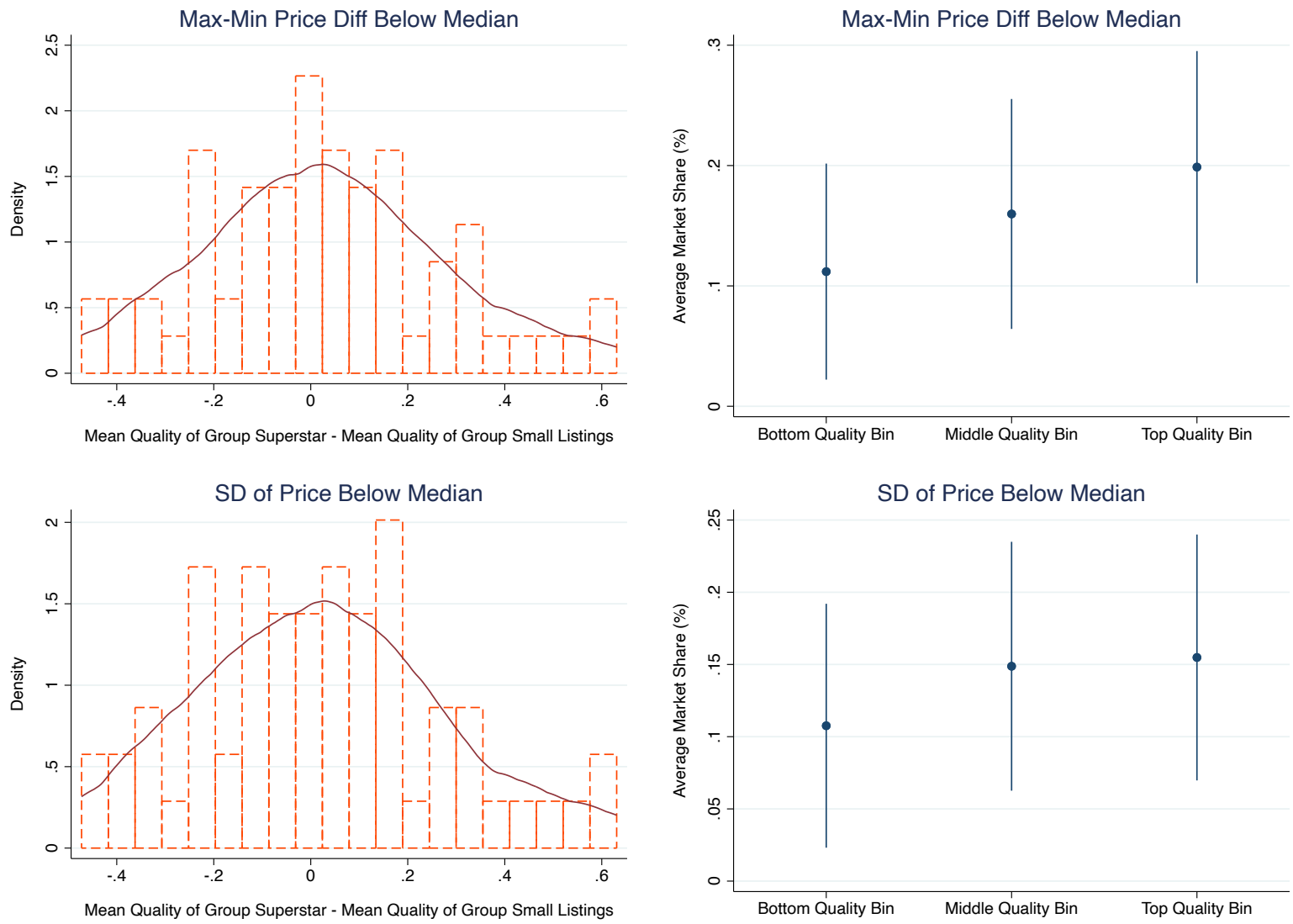
Figure A.1: Product Quality Measurement

Panel A. Inspection and Grading



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

	(1)	(2)
	Dummy=1 if having an order in the following week	
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 past ratings (some small listings in the sample had not received any consumer rating).

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

	Weekly Number of Orders					
	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	-0.004	-0.005	0.054	0.044	-0.008	-0.004	0.001	0.001
	(0.035)	(0.045)	(0.038)	(0.046)	(0.036)	(0.045)	(0.008)	(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

	Endline Number of Cumulative Orders			
	(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: 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.14: Across-Group Sales Distribution

Total share of cumulative orders by top groups	**Data**	**Model**
Top 1% groups	27.47	22.57
Top 5% groups	56.14	49.01
Top 10% groups	70.59	64.47
Top 25% groups	87.53	88.33
Top 50% groups	96.56	99.21

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.15: 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)	Variety Loss (%) (4)	Ave Seller Profit (5)	Platform Commission ($1K) (6)	Average Consumer Surplus (7)
Panel A: $K = 50$, Random Removal							
10000 Seller-Listings	0.67	0.87	0.93	0.0	47.6	109.27	3.427
5000 Seller-Listings	0.72	0.90	0.95	40.3	95.6	109.87	3.456
Panel B: $K = 500$, Random Removal							
10000 Seller-Listings	0.61	0.85	0.92	0.0	48.5	111.69	5.401
5000 Seller-Listings	0.64	0.87	0.93	40.3	96.7	112.48	5.334
Panel C: $K = 500$, Removing Niche Variety Groups							
10000 Seller-Listings	0.61	0.85	0.92	0.0	48.5	111.69	5.401
5000 Seller-Listings	0.83	0.94	0.97	93.4	96.6	115.03	4.968
Panel D: $K \sim$ Positive Poisson(5), Random Removal							
10000 Seller-Listings	0.53	0.80	0.88	0.0	36.2	78.53	1.169
5000 Seller-Listings	0.61	0.85	0.91	40.3	75.8	80.59	1.232

Note: This table reports the results of several counterfactual exercises on cutting the number of sellers. 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) equal to 50. Panel B makes the same comparison but with K at 500. Listings are randomly removed in these two panels. In addition to setting K to 500, Panel C further considers a counterfactual scenario where the removal of the 5,000 seller-listings are from the niche variety groups with the fewest product listings. Finally, Panel D test for robustness when K is drawn from a positive Poisson distribution with parameter 5, using the re-estimated parameter values reported in Column (3) of Table A.16. Section 6.5 describes the counterfactual exercises in more detail.

Table A.16: Robustness: Structural Estimation

	(1) Data	(2) Baseline	(3) Alternative Sample Size	(4) Smaller Price Elasticity
A. Parameter Estimates				
Initial Visibility v_0		0.26	0.27	0.26
Review Noisiness σ		5.46	5.55	5.69
Quality-Cost Correlation ρ		0.48	0.49	0.46
Power of the visibility function λ		0.97	0.94	0.97
B. Simulated vs. Data Moments				
Market share distribution				
Top 1% cumulative revenue share	0.304	0.324	0.322	0.329
Top 5% cumulative revenue share	0.608	0.597	0.590	0.600
Top 10% cumulative revenue share	0.745	0.730	0.723	0.732
Top 25% cumulative revenue share	0.898	0.891	0.887	0.892
Top 50% cumulative revenue share	0.974	0.973	0.971	0.973
Dependence of new order on cumulative orders	0.103	0.136	0.136	0.136
Quality and sales relationship				
Cumulative orders share: Top 1/3 quality bin	0.434	0.424	0.412	0.434
Cumulative orders share: Middle 1/3 quality bin	0.311	0.313	0.319	0.310
Reg. coef. of log price and quality	0.125	0.133	0.140	0.142
Experimental treatment effect	0.186	0.136	0.140	0.141

Note: This table reports the parameter estimates and model fitness in the baseline and alternative model specifications. Column (2) reproduces the baseline results in Tables 5 and 6. Column (3) assumes that the consumer sample size K is drawn from a positive Poisson distribution with mean 5. Column (4) assumes that the price elasticity is low at 2.06 or, equivalently, that γ is 0.39. This low demand elasticity is the one estimated with our offline resale data.

Table A.17: Quality Inference and Visibility Function Based on Cumulative Orders and Reviews

	(1) Quality Inference	(2) Visibility: $\zeta = 0.1$	(3) Visibility: $\zeta = 0.2$
Re-calibrated Parameters			
v_0	0.265	0.260	0.260
λ	0.999	0.900	0.850
σ	5.428	8.000	15.00
ρ	0.462	0.480	0.450
Moments			
Top 1% cumulative revenue share	0.336	0.332	0.387
Top 5% cumulative revenue share	0.604	0.603	0.677
Top 10% cumulative revenue share	0.735	0.737	0.810
Top 25% cumulative revenue share	0.894	0.898	0.943
Top 50% cumulative revenue share	0.974	0.975	0.988
Dependence of new order on cumulative orders	0.134	0.139	0.137
Cumulative orders share: Top 1/3 quality bin	0.451	0.452	0.474
Cumulative orders share: Middle 1/3 quality bin	0.307	0.290	0.284
Reg. coef. of log price and quality	0.127	0.134	0.123
Experimental treatment effect	0.121	0.175	0.315
Baseline: $N = 10,000$			
Cumulative orders share: Top 10% adj-quality	0.55	0.52	0.52
Cumulative orders share: Top 25% adj-quality	0.78	0.77	0.75
Cumulative orders Share: Top 33% adj-quality	0.88	0.86	0.86
Average Seller Profit	26.1	25.9	26.6
Platform Commission ($1K)	56.92	55.25	57.28
Average consumer surplus	0.714	0.598	0.594
Counterfactual: $N = 5,000$			
Cumulative orders share: Top 10% adj-quality	0.65	0.59	0.58
Cumulative orders share: Top 25% adj-quality	0.85	0.82	0.80
Cumulative orders Share: Top 33% adj-quality	0.92	0.89	0.89
Average Seller Profit	56.3	54.8	56.4
Platform Commission ($1K)	60.44	56.71	58.61
Average consumer surplus	0.785	0.639	0.632

Note: This table considers alternative model specifications on how consumers form quality expectation and on platform assigns visibility. In Column (1), consumers use both cumulative orders and reviews to infer listing qualities, as described in Appendix D.5. In Columns (2) and (3), 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.6. ζ determines the relative weight of these two components.

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.[40]

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

[40] 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.[41]

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).

[41] 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 Robustness on Demand Elasticity

In addition to calibrating the demand elasticity to the standard value used in the literature, we also exploit the offline data to estimate an alternative version of the demand elasticity. First or all, we confirm that offline sales price contains useful information: controlling for product quality, it significantly predicts whether a product is eventually sold or not.

Next, we build an estimation strategy that leverages the order of sales. When consumers perfectly observe the quality of every product, the utility of purchasing product i is given by:

$$U_i = \beta + q_i - \gamma p_i + \varepsilon_g.$$

The offline sales features "sampling without replacement," which deviates from the standard logit demand model. As a result, the Law of Large Numbers cannot be applied to derive market shares.

However, we have recorded the order in which products were sold $(i_1, i_2, \ldots, i_n)$. This information allows us to use the sequence of sold products to identify γ. In particular, the probability that product i_1 is chosen can be written as

$$P(i_1) = \frac{\exp(q_{i_1} - \gamma p_{i_1})}{\sum_i \exp(q_i - \gamma p_i)}.$$

After i_1 gets sold, the probability that i_2 is chosen will be

$$P(i_2) = \frac{\exp(q_{i_2} - \gamma p_{i_2})}{\sum_{i \neq i_1} \exp(q_i - \gamma p_i)}.$$

Therefore, the entire probability of our resale outcome will be

$$P = \frac{\exp(q_{i_1} - \gamma p_{i_1})}{\sum_i \exp(q_i - \gamma p_i)} \times \frac{\exp(q_{i_2} - \gamma p_{i_2})}{\sum_{i \neq i_1} \exp(q_i - \gamma p_i)} \times \cdots \times \frac{\exp(q_{i_n} - \gamma p_{i_n})}{\sum_{i \neq i_1, i_2, \ldots, i_{n-1}} \exp(q_i - \gamma p_i)}.$$

We can then use MLE to estimate γ. The identification comes from the fact that γ alters the pecking order of products.

The resulting MLE estimate of γ is 0.2729, with std error equal to 0.0765. This implies a price elasticity of 2.06, which is smaller than the baseline calibrated value 6.7. Therefore, we have re-estimated the model using the lower price elasticity (see Table A.16 Column (4) for the parameter estimates and matched moments), and re-run the counterfactual exercises to examine the roles of information frictions and congestion, as well as the policy experiments (see Table D.1 and Table D.2 below). Overall, the results and takeaways are robust and similar to the main results in the paper.

Table D.1: Robust Counterfactuals with Low Demand Elasticity

	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)	Platform Commission ($1M) (6)
			Panel A: with 10000 sellers			
$\sigma = 5.69$	0.53	0.77	0.87	0.680	0	0.184
$\sigma = 0$	0.90	0.97	0.99	1.081	0	0.244
			Panel B: With 5000 sellers and $\sigma = 5.69$			
by removing random listings	0.61	0.83	0.91	0.725	40.3	0.190
by removing listings from niche groups	0.70	0.82	0.90	0.718	93.4	0.190

Note: This table reports the results of several counterfactual exercises based on the estimates under low demand elasticity. 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 D.2: Robust Policy Simulations with Low Demand Elasticity

	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)	Platform Commission ($1M)
			Panel A: Onboarding to a Large Existing Marketplace			
$\sigma = 5.69$, 1000 Seller-Listings	58.0	409.9	0.20	0.45	0.60	34.6
			Panel B: Onboarding to a New Marketplace			
$\sigma = 5.69$, 1000 Seller-Listings						
0.1% Traffic	217.3	665.9	0.20	0.46	0.62	126.6
0.5% Traffic	1316.0	4060.6	0.23	0.51	0.67	745.1
1.0% Traffic	2792.1	8715.9	0.26	0.55	0.71	1572.4
			Panel C: Onboarding to a New Marketplace			
$\sigma = 5.69$, 500 Seller-Listings						
0.1% Traffic	237.2	727.5	0.21	0.48	0.65	136.5
0.5% Traffic	1394.4	4348.6	0.26	0.56	0.72	784.7
1.0% Traffic	2891.0	9202.5	0.31	0.61	0.76	1629.9

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 under low demand elasticity. 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.

D.5 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.6 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.17 Columns (2) and (3), 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.17 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.