HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Search and Information Frictions on Global E-Commerce Platforms: Evidence from AliExpress

Faculty Research Working Paper Series

Jie Bai
Harvard Kennedy School

Maggie Chen
George Washington University

Jin Liu
New York University

Daniel Yi Xu
Duke University

November 2020; Updated September 2021

RWP20-037

Visit the **HKS Faculty Research Working Paper Series** at:
https://www.hks.harvard.edu/research-insights/publications?f%5B0%5D=publication_types%3A121

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

www.hks.harvard.edu

Search and Information Frictions on Global E-Commerce Platforms: Evidence from AliExpress
Jie Bai, Maggie Chen, Jin Liu, Xiaosheng Mu, and Daniel Yi Xu
NBER Working Paper No. 28100
November 2020, Revised September 2021
JEL No. F14,L15,O12

ABSTRACT

We study how search and information frictions shape market dynamics in global e-commerce. Observational data and self-collected quality measures from AliExpress establish the existence of search and information frictions. A randomized experiment that offers new exporters exogenous demand and information shocks demonstrates the potential role of sales accumulation in enhancing seller visibility and overcoming these demand frictions. However, we show theoretically and quantitatively that this demand-reinforcement mechanism is undermined by the large number of online exporters. Our structural model rationalizes the experimental findings and quantifies efficiency gains from reducing the number of inactive sellers.

Jie Bai
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138
and NBER
jie_bai@hks.harvard.edu

Maggie Chen
Dept. of Economics
George Washington University
2115 G ST, NW, #367
Washington, DC 20052
xchen@gwu.edu

Jin Liu
Department of Economics
New York University
jl7767@nyu.edu

Xiaosheng Mu
Princeton University
xmu@princeton.edu

Daniel Yi Xu
Department of Economics
Duke University
213 Social Science Bldg
419 Chapel Drive
Box 90097
Durham, NC 27708-0097
and NBER
daniel.xu@duke.edu

For acknowledgements and financial disclosure information, see:
https://www.nber.org/papers/w28100

1 Introduction

E-commerce sales have grown tremendously in recent years, reaching $2.9 trillion in 2018 and 12% of total global retail sales (Lipsman, 2019). Within e-commerce, cross-border sales have grown twice as fast as domestic sales, and nearly 40% of online buyers completed a cross-border transaction in 2016 (Pitney Bowes, Inc, 2016). 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.[1] Given these promises and the large market potential, numerous policy initiatives have been adopted worldwide to foster e-commerce growth (e.g, UNCTAD, 2016), with a specific policy target to onboard developing-country SMEs onto e-commerce platforms and allow them to tap into the global market.

Despite the rapid growth of global e-commerce, there is a lack of empirical evidence on the impact of the increased export opportunities on firm growth and market dynamics. While e-commerce potentially exposes prospective exporters to buyers around the world, the sheer number of firms operating on these platforms can create substantial congestion in consumer search.[2] When firms' intrinsic quality is not perfectly observed, search frictions can further slow down resolution of the information problem and hinder market allocation toward better firms.

In this study, we experimentally, theoretically, and quantitatively investigate how these demand-side frictions shape market dynamics in global e-commerce. We first document descriptive evidence consistent with sizable search and information frictions in global online marketplaces. Next, we experimentally identify and demonstrate the role of demand accumulation in helping firms overcome these frictions by improving firms' visibility and generating future demand. Motivated by the reduced-form evidence, we develop a theoretical model that formalizes this demand-driven reinforcement channel under search and information frictions and characterize its efficiency implications. Building on the theoretical framework, we estimate a rich empirical model of the online export market and use the model estimates to quantify the impacts of search and information frictions on firm growth, market allocation, and consumer welfare. Finally, we apply our model to shed light on policies that could facilitate the growth of promising export

[1] 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!"

[2] The logic of this effect can be traced to Stigler (1961), in which consumers are not perfectly informed about all products available for purchase in a given market and can only consider a limited subset.

businesses beyond the initial onboarding stage and improve overall market efficiency.

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 industry of children's T-shirts and collect comprehensive data about sellers operating in this industry, 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 based on actual online purchases and direct interactions with the sellers as well as third-party assessments.

We begin by documenting a set of new stylized facts about the online exporters. First, we compare the sales distribution within identical-looking product varieties. Interestingly, even after we control for horizontal taste differences, there remains meaningful dispersion in sales within identical-looking variety groups, as opposed to "winner-takes-all." This finding is indicative of search frictions: buyers, upon arriving on the platform, face thousands of product offerings but can consider only a finite subset of all seller listings.[3] This raises the question of who gets to grow in the presence of these search frictions. Next, we dive further into the potential determinants of growth and find that quality only weakly predicts sales. The "superstars," which we define as the largest seller in each product variety, do not necessarily have the highest quality (or the lowest price). Intuitively, search frictions introduce a random component in firm growth due to the consumer sampling process. When firms' intrinsic qualities are not perfectly observed (even after they enter a consumer's search set), such frictions can further slow down the revelation of true quality, leading to potential market misallocation. Finally, we find robust evidence that current sales predict the speed of arrival for future sales. This implies that firms with larger past sales—and hence higher visibility—have an advantage in overcoming consumer search frictions and generating future orders. However, if information frictions prevent a firm's visibility from being aligned with its fundamentals, it could take much longer for better firms to distinguish themselves. Overall, market allocation and consumer welfare depend crucially on the interactions of these demand-side forces.

Our interpretation of the stylized facts highlights a *demand-reinforcement mechanism* where accumulating sales boosts a firm's visibility and helps attracting future sales. Having said that, unobserved supply-side actions (such as advertising and display) could also exist and lead to

[3]This is consistent with prior studies that find substantial price dispersion in online marketplaces even for identical products, indicating the presence of significant search frictions (e.g., Clay et al., 2001; Clemons et al., 2002; Hortaçsu and Syverson, 2004; Hong and Shum, 2006).

similar reduced-form relationships between current sales and future sales. To further establish the empirical validity of the demand mechanism, we conduct an experiment in which we generate exogenous demand shocks to a set of small exporters via randomly placed online purchase orders. Since how effectively the additional demand conveys the firm's true fundamentals depends critically on the severity of information frictions, we further interact the order treatment with a review treatment about firms' product and shipping quality to examine the role of information provision. We find that the order treatment leads to a small but significantly positive impact on firms' subsequent sales, establishing the demand mechanism. We document supporting evidence that this effect is mediated by a short-term boost in a listing's ranking and is unlikely to be driven by endogenous supply-side responses. In the meantime, we do not find any significant treatment effect from the reviews, suggesting that the online reputation mechanism may not function very effectively in the presence of large search frictions. This outcome echoes the stylized fact that quality does not strongly predict growth in this market due to the search and information problems, which, when combined together, make it particularly difficult for high-quality sellers to stand out. Using cumulative sales measured at the endline, we estimate an average treatment effect that is much smaller than the size of the initial treatment (0.1-0.25 versus 1), suggesting that the frictions cannot be easily overcome by an individual seller's private effort.

All together, the descriptive and experimental findings are consistent with the presence of sizable search and information frictions and establish an important demand-reinforcement mechanism of online firm growth. Motivated by the reduced-form evidence, we develop a theoretical model that formalizes the process of demand accumulation under consumer search and learning and use the model to examine the efficiency properties of the market. The model extends the classical Polya urn model by incorporating consumer choice and seller heterogeneity in quality. In every period, consumers conduct a fixed sample search where the probability of a seller being sampled is proportional to a power function of its cumulative sales. Consumers observe noisy signals of seller quality based on past reviews and make purchase decisions based on expected quality within the search sample. To examine the implications of search and information frictions, we focus our theoretical analysis on a key comparative static with respect to the number of sellers: while e-commerce lowers the entry barriers to exporting and brings many firms online, the presence of a large number of sellers can exacerbate congestion in consumer search and slow down the rise of high-quality sellers, given consumers' limited search sample size. Tying this analysis back to our policy motivation, we prove theoretically that indeed when there are already many sellers in the market, further increasing the number of sellers strictly worsens allocation

and welfare along the path of market evolution.

Building on the theoretical framework, we estimate a rich empirical model of the online market. We follow the setup of the theoretical model closely and extend the setup to incorporate seller-side heterogeneity in both quality and cost and model sellers' pricing decisions. Our estimate implies that compared with sellers who have made zero sales, sellers who strike a first order are 4.6 times more likely to end up in a subsequent consumer's search sample. Nonetheless, due to the information problem, uncertainty regarding the seller's quality still remains even after the seller successfully strikes the first sale. Our estimate of the review signal noise implies that the posterior uncertainty is reduced by only 3.3% after the first order. This suggests that online reviews are very noisy signals about sellers' quality and that the informational uncertainty is resolved very slowly, i.e., only after a seller accumulates a substantial number of orders. Combined, these findings highlight that search frictions, interacted with information frictions, can constitute an important hurdle for the growth of small 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 model.

We end with several counterfactual exercises to examine the distinctive roles of search and information frictions in firm growth and market allocation and use the estimated model to evaluate potential policy interventions. First, we investigate the impact of lowering search frictions by reducing the number of sellers operating on the platform. The results show that doing so helps mitigate congestion in consumer search, thereby improving allocative efficiency and consumer welfare (as in our theoretical analysis). The improvement is especially strong when there exist large information frictions. This result serves as a cautionary tale against blanket onboarding initiatives that try to bring many firms online: simply giving firms easy access to foreign markets may not be sufficient for generating sustained growth and can in fact exacerbate search and information problems, resulting in market misallocation.

Next, to shed light on the role of information frictions, we remove the noise from the review signals. We find that doing so significantly shifts market share to high-quality sellers and raises consumer surplus. We further decompose this welfare gain into a static gain (better decision-making conditioning on a search sample) and a dynamic gain (better sample quality over time) and find that the latter plays an important role. This result further highlights the important interaction between search and information frictions.

Last but not least, we evaluate a potential policy, common among existing successful platforms, to reduce search congestion by establishing a premium market segment that eliminates

a large number of inactive product listings and sellers. The key question for emerging new e-commerce export platforms in developing economies is when to introduce such segmentation. The key tradeoffs are again highlighted by our theoretical channels based on search and information frictions: on the one hand, segmenting the market in the early stage, such as during the initial onboarding stage when the market is just created, effectively reduces search congestion and facilitates faster allocation of market share toward higher-quality sellers. On the other hand, segmenting the market too early results in loss of information that could be revealed through sellers' performance over time. We show that welfare is indeed nonmonotonic with respect to the timing of implementation, reflecting the balance between the underlying economic forces.

While our setting is specific to e-commerce, the economic insights generalize to broader market settings. It is well understood that there may be excessive entry when firms do not internalize their "business-stealing" from competitors (Mankiw and Whinston, 1986). Our paper illustrates that with the presence of search frictions, such business-stealing can happen when sellers compete for customer attention, beyond simple price competition. We further show that the business-stealing effect can be particularly costly when there exist information frictions, which prevent the best firms from being discovered and reduce the allocative efficiency of the market. Several features of the cross-border e-commerce setting make it a good exhibit of these economic forces: specifically, the lack of market selection, which results in excessive entry, and differences in cultural and business practices across borders, which aggravate the information problem.

Related Literature. Our work contributes to several strands of the existing literature. A large literature in trade and development has documented the important role of quality in determining exporter performance (see Verhoogen, 2020 for an excellent review). We build on a growing body of research that collects detailed information on quality for specific industries (e.g., Atkin et al., 2017; Macchiavello and Miquel-Florensa, 2019; Bai et al., 2019; Hansman et al., 2020). Similar to these earlier works focused on offline settings, we document 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 long-run market shares. Our paper highlights the role of search and information frictions in explaining the disintegration of the customer accumulation process and firms' fundamental quality. Leveraging unique quality data, 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 in trade and development on search and information frictions (Allen, 2014; Macchiavello and Morjaria, 2015; Steinwender, 2018; Startz, 2018). Theoretically, we formalize the process of consumer search 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 mechanism that can potentially help firms overcome these frictions. We then formally model these realistic market frictions and quantify their impacts on market dynamics and consumer welfare. 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. Rather than focusing on firms' own actions, we explore how foreign demand shocks improve firm visibility and help them overcome market search and information frictions.

Third, our work relates to the extensive literature on new exporter dynamics, primarily in the offline setting. A common empirical pattern that emerges from micro data is that young exporting firms start small and have high turnover rates, and those that survive experience rapid growth. Various theoretical mechanisms have been proposed to explain these facts. They include firm learning (Arkolakis et al., 2018; Ruhl and Willis, 2017), demand accumulation (Foster et al., 2016; Piveteau, forthcoming; Fitzgerald et al., 2020), and seller search (Eaton et al., 2016). In contrast to these studies, our paper focuses on demand-side frictions as the key driving force of firm and market dynamics in the online setting. In particular, unlike in the offline export market, the fixed costs of operating are substantially lower in online marketplaces, significantly weakening the role of market selection. Our work shows how the lack of selection reduces consumer search efficiency and endogenously slows down the growth of high-quality exporters.

Lastly, our study also relates to the existing literature on consumer search and consumer consideration sets (for example, Goeree, 2008; Kim et al., 2017; Honka et al., 2017; Dinerstein et al., 2018).[4] We extend existing search models, which have assumed that consumers can perfectly learn a product's utility after a single search, by incorporating information frictions and a process of consumer learning enabled by the online review mechanism (Tadelis, 2016). This approach allows us to examine the interaction between search and information frictions.[5]

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 and motivates the

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

[5]In a different setting, Pallais (2014) and Stanton and Thomas (2016) examine information frictions 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 firms. We further introduce search frictions and show that doing so exacerbates the initial hurdle for high-quality new businesses to distinguish themselves.

experiment. 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 study setting, the market for children's T-shirts on AliExpress, and describe the data collection.

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.[6] 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.[7] 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 firms' sourcing ability (i.e., ability to source high-quality products from manufacturers) as well as the quality of their marketing and shipping services.[8]

For this study, we focus on the children's T-shirt industry. 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 (WTO, 2020). 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.[9] The growth and efficiency of the online retail market therefore matters for upstream manufacturing: in particular, growth of retailers that sell high-quality products in turn benefits their producers.

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

[7] 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.

[8] 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.

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

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.[10] While this helps bring many SMEs onto the platform, the lack of market selection can create important congestion in consumer search, resulting in an excessive number of firms and product offerings competing for consumers' attention in the online marketplace. The resulting welfare implications of the increasing number of market participants for firms and consumers are far less clear in the presence of search and information frictions. 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*. A single *variety group* (hereafter referred to as a *group*) may contain multiple listings that are sold by different sellers but share an identical product design. This is illustrated in Figure 1. This unique feature allows us to compare listings with the same observable product attributes, thereby controlling for consumers' horizontal taste differences. We leverage this feature in our empirical analysis as described below.

2.2 Data

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

(1) Store-Listing-Level Data. We scraped nearly the full universe of product listings in the children's T-shirt industry in May 2018.[11] 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

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

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

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. To collect the quality data, we focus on variety groups with at least 100 cumulative sales (aggregated across all listings in the group) to focus on products that are more relevant for consumer choice. This leaves us with 1,258 product listings sold by 636 stores in 133 variety groups, with varying performance heterogeneity (measured in terms of cumulative sales) within each group. 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.[12] After receiving and cataloging the orders, we worked with a large local consignment store of

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

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 grading 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 and the criteria used. 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.[13] 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 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.1 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

[13]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.

quality—statistically significant, as shown in Table A.2. 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.3 shows that a seller's reply speed is highly consistent over time.[14]

3 New Stylized Facts about Online Exporters

Using the newly assembled micro dataset, we begin by documenting a set of new stylized facts about online exporters. These facts provide suggestive evidence of the presence of search and information frictions and, at the same time, highlight a demand-reinforcement mechanism 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 2, 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%).

Nonetheless, the distribution of superstar sales across groups makes clear that this is not a case of a winner-take-all market; some amount of dispersion still remains. Given that we are comparing products with essentially identical designs, we are controlling for unobserved consumer horizontal tastes. In a frictionless world, we may expect that the listing with the highest quality, relative to price, would win the market.[15] The fact that some dispersion remains indicates that

[14]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.3, 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.3.

[15]Part of the dispersion could be due to heterogeneous preferences for quality and price among consumers. To examine this possibility, we leverage the six-month transaction data, in which we observe buyers' country of origin, to examine sales performance within identical-looking variety groups by country (where we restrict sales to a given country and define the top sellers for each country separately). Figure A.2 shows the patterns for the

frictions exist in this marketplace. This raises the question of who gets to grow in the presence of these frictions. To delve deeper into this issue, we next ask who gets to become superstars.

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

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 due to search and information frictions. Intuitively, search frictions introduce a random component in firm growth due to the consumer sampling process. When firms' intrinsic qualities are not perfectly observed, such frictions can further slow down resolution of the information problem and hinder market allocation toward better firms. That said, this evidence is only suggestive because we have to take into account price differences.[16] To isolate the role of search and information frictions and quantify the degree of misallocation, we rely on a structural model in Section 6.

Fact 3. *On average, it takes 79 days for the first order to arrive; after this period, subsequent orders arrive much faster.*

Finally, we delve further into the growth dynamics and examine how superstars emerge. Using the transaction-level data from the six-month period from March to August 2018, we explore the dynamics of order arrivals. Figure 4 plots the number of days that it takes to receive the n-th order. Panel A shows the order arrival dynamics for the full unbalanced sample of all listings;

US and Russia. To the extent that consumers' tastes are more similar within a country than across countries, the fact that we still observe a sizable amount of dispersion at the top suggests that not all of it is explained by heterogeneous preferences.

[16] Interestingly, we find that superstars do not always charge the lowest price, either: within an identical-looking 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.

Panel B is restricted to listings that accumulated more than 10 orders during the six-month period. A striking pattern emerges: on average, it takes 70-79 days for the first order to arrive; however, conditional on receipt of one order, subsequent orders arrive much faster. For example, on average, the second order arrives 7-15 days after the first order, and the third order arrives 4-6 days after the second. Table A.4 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 results highlight a potential demand-reinforcement mechanism: that is, sellers with larger past sales—and hence higher visibility—have an advantage in overcoming search and information frictions and generating future sales.

However, a key empirical challenge in identifying the role of this demand-side effect is to control for unobserved supply-side actions. For example, it could be that after some initial period of preparation, sellers start to invest in some costly actions, such as paying for advertising or participating in sales and promotion events organized by the platform, which then lead to the first order and subsequent orders. From the observational data, it is difficult to tease apart the demand- and supply-side channels. This motivates us to conduct an experiment to identify the role of demand.

4 Experiment and Findings

To demonstrate the role of demand in helping firms overcome search and information frictions in e-commerce, we conduct an experiment in which we generate exogenous demand and information shocks to a set of small sellers via randomly placed online orders and reviews. We describe the experiment design and present the main findings below.

4.1 Experiment Design

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,[17] 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 buyers. To generate the contents of the shipping and 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 demand. The difference between T1 and T2 identifies any additional impact of alleviating information frictions. 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.5 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 demand and information shocks on firms' subsequent growth, we track all listings for 13 weeks after the initial order placement and estimate the following regression:

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

where the dependent variable is the total number of orders (excluding our own order) for listing i in week t.[18] 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 group fixed effects. In addition, all

[17] 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.

[18] We focus on the impact 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.

regressions control for baseline sales at the store and the listing level. Results without these baseline controls are shown in Table A.6. 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 United States separately. Overall, we see that the order treatment has a small but significantly positive impact on subsequent orders. This demonstrates that positive demand shocks, exogenously generated in our experiment, increase the speed of arrival of future sales. To shed light on the mechanism, Table A.7 shows that the order treatment leads to a small short-term improvement in a listing's ranking. Consistent with this, Table A.8 investigates the dynamic effects of the order treatment and shows that the effect is salient in the short run (i.e., the first month) but decays afterwards. All together, these findings indicate that a key channel for firms to grow in the online marketplace is by accumulating demand to boost their visibility.

To further establish the demand-reinforcement mechanism, we show that the order treatment effect is unlikely to be driven by endogenous supply-side reactions. Table A.9 examines the order treatment effect on seller effort and business strategy and shows that receiving a small order does not lead to any noticeable adjustment in pricing, the shipping service, the listing description (an indication of advertising effort), or the introduction of new listings.

In contrast to the order treatment effect, we do not find any significant treatment effect of the reviews. There are two possible explanations for this: 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 search frictions. Similarly, we also do not find any heterogeneous treatment effects based on quality, as shown in Table A.10.[19] This result echoes the stylized fact that quality does not strongly predict sales performance in this market. In this market environment, search and information frictions combined can make it difficult for high-quality sellers to stand out.

Finally, using cumulative sales measured at the endline, netting out our own order, we estimate an average treatment effect ranging from 0.1 to 0.25, as shown in Table 4. 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-side frictions cannot be easily overcome by individual sellers' private efforts.[20]

[19]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.

[20]In addition, the cost of manipulating orders on AliExpress (an exclusively cross-border platform) is fairly

5 Theory

The above evidence, consistent with the presence of search and information frictions, highlights an important reinforcement mechanism of online firm growth through consumer search and learning: accumulating sales boosts a seller's visibility and helps attract future sales. In this section, we develop a theoretical model that formalizes the process of demand accumulation under search and information frictions in the online marketplace. Using the model, we characterize the market evolution process and examine the efficiency properties of the online market.

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. Consumers' purchase decisions are based on a search procedure as well as a choice procedure. The probability that a seller appears in the search sample of a consumer is governed by the seller's visibility, which is the sum of some initial visibility parameter and total past sales. This reflects the demand-reinforcement mechanism that we want to capture. On the other hand, conditional on being in the search sample, a seller is chosen—i.e., makes the sale—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.

More formally, 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.[21] 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

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.

[21] 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.11 that the alternative procedure of sampling without replacement generates nearly identical quantitative predictions.

common initial visibility level. Specifically, this 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 the demand-reinforcement mechanism.[22]

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 frictions 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 $N, K, v_0, \lambda, \sigma$.

[22] 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' search samples. 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 will see, 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 search sample). 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 mechanism. This departure from the classical model leads to fundamentally different market outcomes.[23]

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.[24] Empirically, we also report the robustness check results of this sampling procedure in Appendix D.4.

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 firms online, the presence of a large number of sellers can exacerbate congestion in consumer search, given consumers' limited search sample size. In what follows, we study the effect of increasing

[23] 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.

[24] 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 and welfare.

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.*[25]

Thus, when there are already many sellers in the market, allocation and welfare worsen 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 search 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 search and information frictions. In Section 7, 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 lack of winner-taking-all in identical-looking variety groups and a weak relationship between quality and seller performance measured in cumulative sales. The question is, then, the following: Under search and information frictions, 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 the demand-reinforcement mechanism, as captured by the parameter λ:

[25] 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* 0*, so that seller* 1 *cannot occupy the entire market.*

When $\lambda > 1$, the demand-reinforcement mechanism 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 search and information frictions and examine to what extent the demand-reinforcement mechanism works to overcome these frictions. The demand side closely follows the setup of the theoretical model and models search frictions due to limited sample search and information frictions due to imperfect observability of underlying quality. 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 perform a counterfactual analysis of the impact of search and information frictions on firm growth, market allocation, and consumer welfare.

6.1 Demand

Search. Following the theoretical setup in Section 5.1, consumers randomly sample K sellers with replacement upon their arrival.[26] We allow for heterogeneity in consumers' search sample size by assuming that K is randomly drawn from 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 search sample 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.

[26] Our model abstracts away from multiple listings within a store and treats each listing as an independent selling entity. This simplification does not capture across-product spillovers within a store, which are likely to matter for large sellers but be relatively less relevant for small sellers. Table A.4 shows that the demand accumulation mechanism is salient even with store fixed effects, i.e., at the listing level within a store.

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 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 search 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.[27] 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 search sample. 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:

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

[27] 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).

where the first-order condition reads

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

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

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

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 from the search sample, 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 search sample size 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.[28] Another key parameter of the

[28] 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.

model is consumers' search sample size 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).[29] Therefore, in our baseline estimate, we assume that K follows a positive Poisson distribution with mean 2. Section 7.1.3 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 search sample. 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.

[29] Studies consistently find that in online marketplaces, while some consumers search intensely, the vast majority search very little (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 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. This result is consistent with the important role of the first order that we observed in Figure 4. 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 $\lambda = 1$.[30] 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 reputation mechanism takes time to play a role even if a seller emerges in a consumer's choice set and successfully makes a sale. As a result, search frictions, interacted with information frictions, constitutes the major hurdle to sellers' initial growth. 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 Validation Using the Experiment

In this section, we conduct simulation exercises using our estimated model to evaluate its ability to rationalize the experimental findings in Section 4.

[30] 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.

Table 7 presents the model-predicted treatment effects for various one-time demand shocks, with different group sizes of treated sellers and sizes of purchase orders. Recall that in our experiment, 4% of the sellers received our orders. 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 (31 T-shirts). As in the experiment, the size of the purchase is 1. It takes time for the new purchase to generate future orders for treated sellers. 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. In our baseline experiment simulation ($P = 2\%$, $O = 1$), we find that our model estimates result in a treatment effect of 0.129—an average seller receiving a random purchase would grow its orders by around 0.13 pieces. This effect size is quantitatively comparable to the range of the average treatment effect shown in Table 4 of between 0.1 and 0.25. We also show that when the size of orders increases from 1 to 2 or 5, the average treatment effect goes up proportionately. However, notice that the effect size is always lower than the size of our treated purchase, which indicates that market frictions are not easily overcome by private effort of sellers.

7 Counterfactual Analyses

The fit of the data moments and the replication of treatment effects validate our estimated empirical model. We now use the model to examine the role of information and search frictions in firm growth and consumer welfare, and evaluate potential policy interventions.

7.1 Dissecting the Roles of Frictions

7.1.1 The Role of Search Frictions: Reducing the Number of Sellers

First, we examine the impact of reducing search frictions by reducing the number of sellers to alleviate the congestion in consumer search. This change is analogous to raising entry costs or the costs of maintaining active listings on the platform. Panel A.1 of Table 8 shows substantial welfare gains from doing this in the presence of the large information frictions estimated at baseline (i.e., $\sigma = 5.39$). To illustrate this result, Panel A of Figure 5 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 5, and gain larger market share: the cumulative market share of sellers in the top 1/3 in terms of quality increases by 18.1% ($= 0.523/0.443 - 1$) when the number of sellers is reduced from $10,000$ to $5,000$; sales-weighted quality improves by 0.17 standard deviations. As a result, the expected consumer surplus increases by 6.6% ($0.738/0.692 - 1$), taking into account the fact that higher-quality sellers charge higher prices.[31]

Next, we examine the potential interaction effect between search and information frictions. As discussed in Section 5.3, the efficiency and welfare implications of reducing search frictions may depend on the degree of information frictions. To test this, Panel A.2 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 search frictions significantly diminishes. In other words, the congestion effect 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 the information frictions), the greater is the negative effect of increasing N (increasing search frictions).

Finally, as we discussed in our theory Section 5.2, the intuition that having fewer sellers could alleviate search congestion and information friction applies broadly to alternative sampling procedures. In Appendix D.4, 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 $(v_0 + s_i)^\lambda \cdot \exp(\zeta \bar{z}_i)$.

7.1.2 The Role of Information Frictions: Reducing the Review Signal Noise

Next, we examine the role of information frictions. Since a seller's true quality is imperfectly observed by consumers, reviews from past purchases serve as a crucial source of information for subsequent consumers. Our estimate implies quite a noisy review signal, with a standard deviation of 5.39. In Panel B.1 of Table 8, we compare our baseline estimate with a case where we reduce σ to zero. In other words, we investigate the case in which a seller-listing's true quality is immediately revealed when the seller accumulates its first order. We find that in this case, the cumulative market share significantly shifts toward higher-quality sellers. We also find that

[31] Without information frictions, consumer surplus can be computed with the standard log sum formula. However, with information frictions, 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.

as a consequence, sales-weighted quality improves by almost one standard deviation and average consumer welfare by 58.5%.

We can further decompose this welfare gain into a static gain and a dynamic gain. At each given point in time, with the search set held constant, a smaller σ allows consumers to make more informed purchase decisions; that is, they are more likely to purchase from high-quality sellers when reviews offer more precise signals of quality. In addition to this static gain, the sampling probabilities of high-quality sellers are boosted over time, allowing them to attract more future sales. To shed light on these two margins, we perform a counterfactual exercise in Panel B.2 where we shut down the dynamic gain by assuming random sampling in each period (instead of sales-based sampling). We can see that the gain from reducing information frictions is much more muted in this case. Sales-weighted quality improves by 0.36 standard deviations, and consumer surplus increases by only 17.9%. The difference in Panels B.1 and B.2 further highlights the interactions between information and search frictions.

7.1.3 Robustness Checks

Table A.11 reports estimated parameter values under alternative search sample sizes and price elasticities. Table A.12 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 reestimated with a larger search sample 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 welfare 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.

7.2 Policy Analysis

We end with an evaluation of a potential platform policy intervention aimed at facilitating the growth of high-quality e-commerce businesses. As shown in Section 7.1, a reduction in the number of sellers could alleviate search congestion, strengthen the review system, and improve

the allocative efficiency of the e-commerce platform. This intuition is consistent with many existing efforts by successful platforms to establish a premium market segment that essentially eliminates a large number of inactive product listings and sellers.[32] However, for emerging new e-commerce export platforms in developing economies, it remains an open question *when* the best time is to initiate such a segmentation effort to screen out small inactive sellers. The key tradeoffs are again highlighted by our theoretical channels based on search and information frictions. The benefit of trimming inactive sellers earlier is illustrated in Figure 5, where we reduce the number of sellers from $10,000$ to $5,000$ effectively in the initial time period. The smaller number of sellers reduces search congestion from the very beginning and facilitates much faster allocation of market share toward higher-quality sellers. On the other hand, the cost of starting this segmentation process too early is the loss of information that would be revealed from sellers' performance over time, including cumulative sales and consumer reviews (which serve as noisy signals of underlying quality).

To shed light on these tradeoffs, in Table 9, we report the results from a series of policy counterfactuals where we start with $10,000$ sellers and later create a premium segment by including only $5,000$ sellers out of the initial 10,000. Specifically, we focus on sellers with positive reviews (posterior quality larger than zero) and take the top $5,000$ based on their cumulative sales.[33] Different columns examine different timings (in terms of the total number of cumulative transactions completed) of implementing the segmentation. The first three rows report the key statistics that summarize the market performance for each implementation timing T after $620,000$ periods of simulations to make the results comparable with those of our baseline case with no segmentation. We find that later implementation improves the average quality of sellers in the new premium segment. Contrasting $T = 10,000$ with $T = 300,000$, we find that the average quality of the top $5,000$ sellers selected into the new segment improves from 0.036 (barely above average) to 0.184. This is driven by the better information that the platform observes in later periods for the inference of seller quality. That said, interestingly, the average purchase quality turns out to be hump-shaped with respect to the implementation time T. This non-monotonicity is primarily explained by fact that search congestion becomes more severe in the pre-implementation periods when we push T to a later date. Comparing across all the columns in Table 9, we see that the average purchase quality along the path of market evolution turns out to be best at $T = 100,000$, at which point the tradeoff between reducing search congestion and accumulating

[32] One prominent example, for instance, is the introduction of Tmall by Alibaba, targeting premium sellers.

[33] In the case when we have fewer than $5,000$ sellers with positive sales (at the beginning of the market), we randomly sample from sellers with zero sales to make up the rest.

quality information is optimally balanced.

8 Conclusion

In this paper, we study exporter dynamics on global e-commerce platforms. Leveraging comprehensive data about online businesses from AliExpress and combining them with unique objective measures of quality covering multiple product and service dimensions, we document sizable variations in firm-product quality in the online marketplace. However, we find that quality only weakly predicts firm performance and growth. Our paper highlights the role of search and information frictions in explaining the disintegration of the demand accumulation process and firm fundamentals and underscores a potential source of market misallocation in e-commerce.

Our findings speak to the need for effective policies facilitating small business growth via e-commerce. While global e-commerce platforms present a promising avenue for small and medium-sized enterprises in developing countries to tap into the global market, simply bringing firms onto these platforms may not be sufficient for generating sustained growth due to the large demand-side frictions. In fact, doing so can exacerbate search and information problems, resulting in market misallocation. Policies should be designed to help firms, especially new businesses, overcome the additional demand-side frictions. In the context of e-commerce, regulating entry, creating a premium market segment, and directing demand toward promising newcomers could help facilitate growth and improve overall market efficiency.

References

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

Arkolakis, Costas, Theodore Papageorgiou, and Olga A Timoshenko, "Firm learning and growth," *Review of Economic Dynamics*, 2018, *27*, 146–168.

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.

Clay, Karen, Ramayya Krishnan, and Eric Wolff, "Prices and price dispersion on the web: Evidence from the online book industry," *The Journal of Industrial Economics*, 2001, *49* (4), 521–539.

Clemons, Eric K, Il-Horn Hann, and Lorin M Hitt, "Price dispersion and differentiation in online travel: An empirical investigation," *Management science*, 2002, *48* (4), 534–549.

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.

Eaton, Jonathan, David Jinkins, James Tybout, and Daniel Yi Xu, "Two-sided search in international markets," *2016 Annual Meeting of the Society for Economic Dynamics*, 2016.

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

Foster, Lucia, John Haltiwanger, and Chad Syverson, "The slow growth of new plants: Learning about demand?," *Economica*, 2016, *83* (329), 91–129.

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

Hortaçsu, Ali and Chad Syverson, "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly Journal of Economics*, 2004, *119* (2), 403–456.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg, "The probit choice model under sequential search with an application to online retailing," *Management Science*, 2017, *63* (11), 3911–3929.

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

Lipsman, Andrew, "Global ecommerce 2019: ecommerce continues strong gains amid global economic uncertainty," 2019.

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.

Pitney Bowes, Inc, "2016 Pitney Bowes global online shopping study," 2016.

Piveteau, Paul, "An Empirical Dynamic Model of Trade with Consumer Accumulation," *American Economic Journal: Microeconomics*, forthcoming.

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.

Ruhl, Kim J and Jonathan L Willis, "New exporter dynamics," *International Economic Review*, 2017, *58* (3), 703–726.

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.

Stigler, George J, "The economics of information," *Journal of Political Economy*, 1961, *69* (3), 213–225.

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.

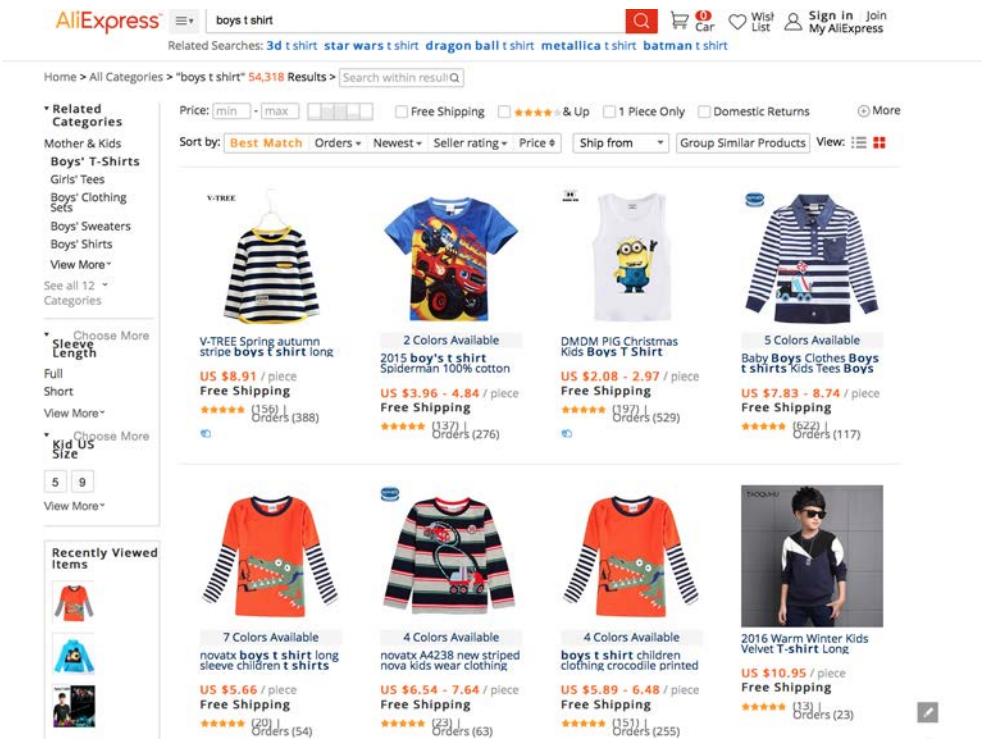
Verhoogen, Eric, "Firm-level upgrading in developing countries," *Working paper*, 2020.

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

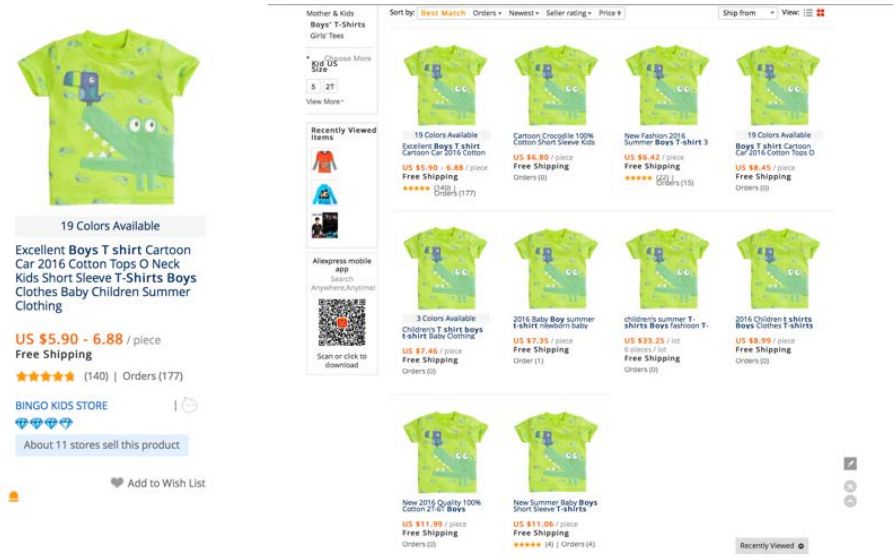
WTO, World Trade Organization, "World trade statistical review 2020," 2020.

Figure 1: AliExpress: Search Results with and without Grouping

Panel A. Search Results without the Grouping Function

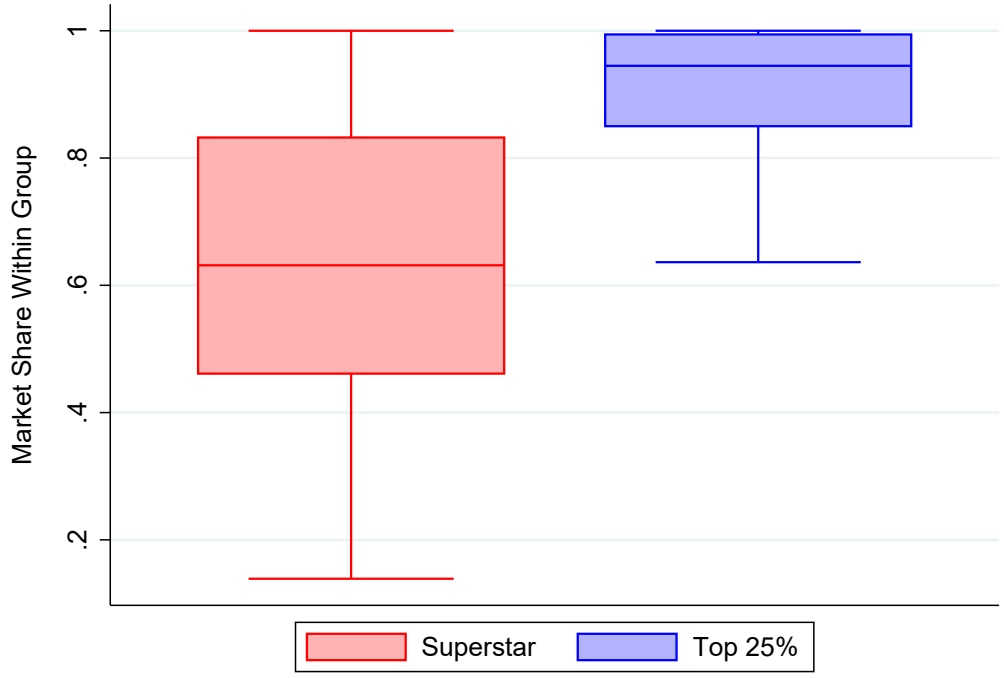


Panel B. Search Results with the Grouping Function



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.

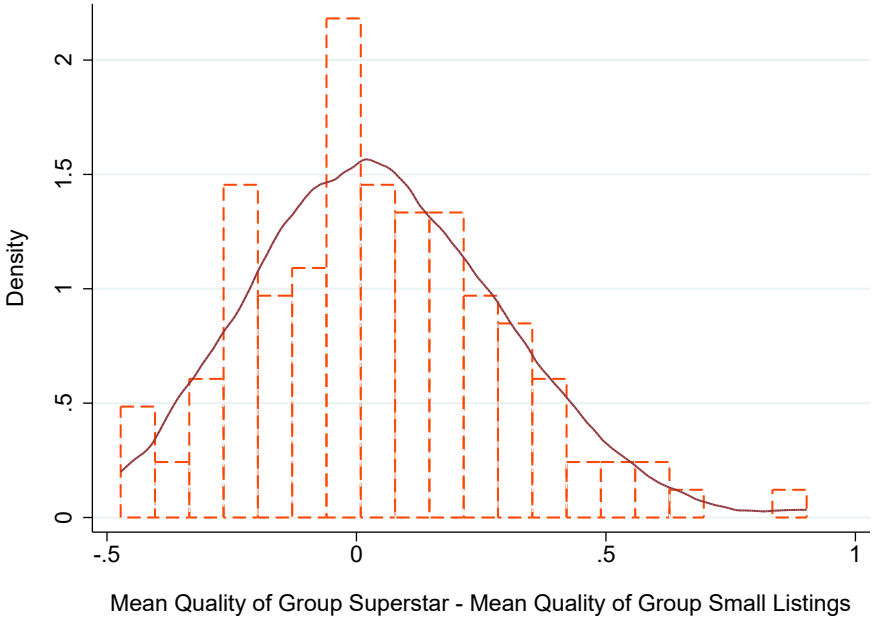
Figure 2: Sales Performance within Identical Varieties



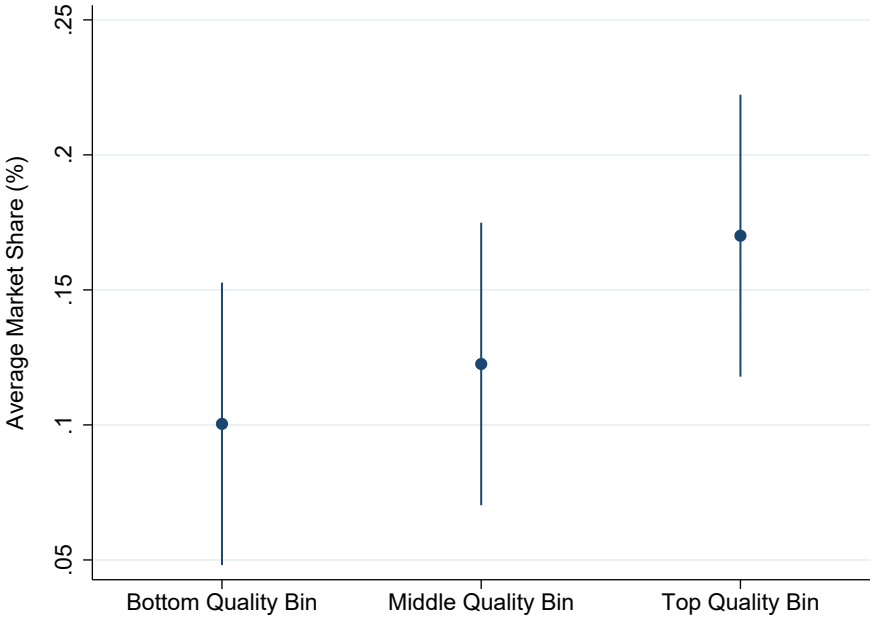
Note: This figure plots the distribution of within-group market shares, measured in cumulative orders, of the top listings in a group using the store-listing-level data described in Section 2.2. Superstar indicates the listing that has the highest number of cumulative orders within a group. Top 25% indicates listings in the top 25th percentile in terms of cumulative orders within a group. We limit the sample to groups with at least 4 listings so that "Top 25%" is well defined.

Figure 3: Quality and Sales Performance

Panel A. Quality Comparison between Group Superstar and Small Listings



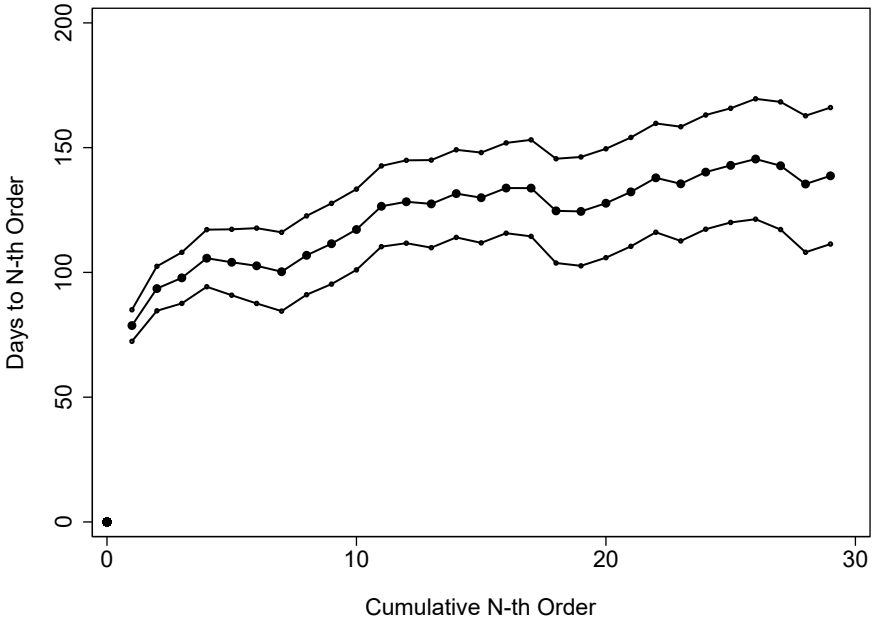
Panel B. Average Market Share over Quality Bins



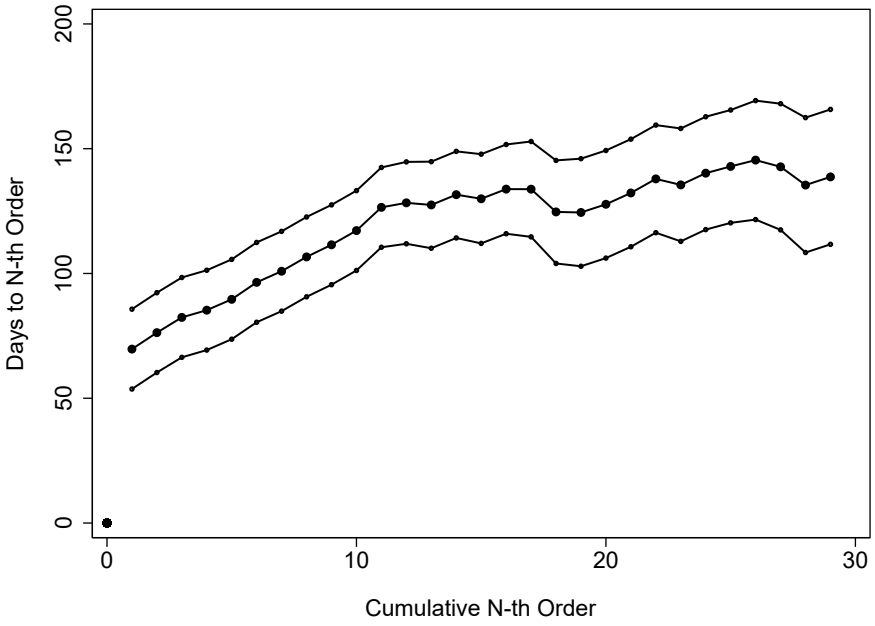
Note: This figure describes the relationship between listings' quality and sales performance. Panel A plots the distribution of the quality gap between group superstars and small listings in the group. Quality is measured in terms of the overall quality index (see Section 2.2 for details on construction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

Figure 4: Dynamics of Order Arrival

Panel A. Unbalanced Panel of All Listings



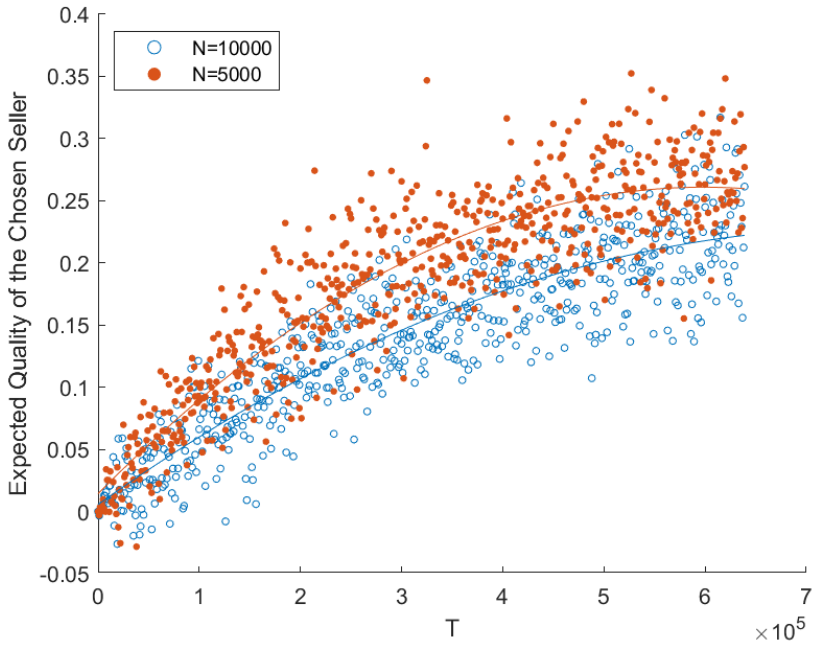
Panel B. Listings with More than 10 Cumulative Orders



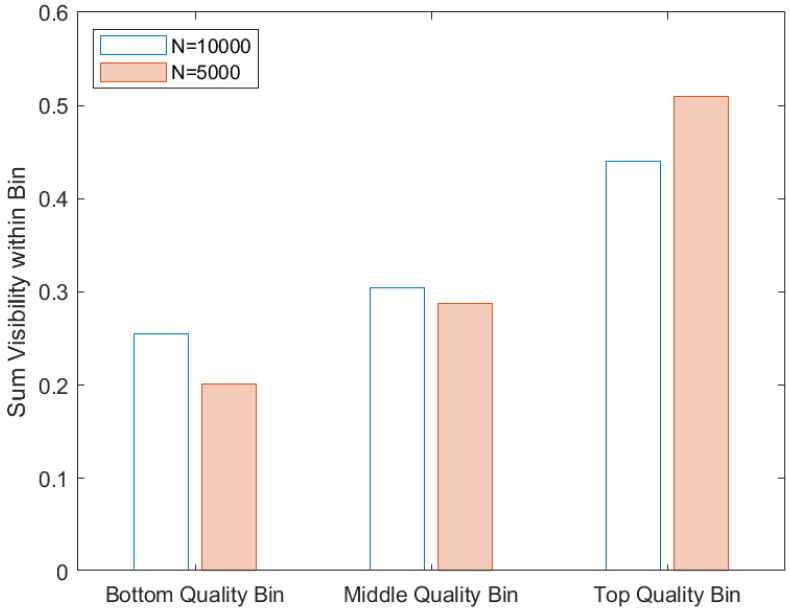
Note: This figure describes the order arrival dynamics using the six-month transaction data described in Section 2.2. The x-axis indicates the n-th cumulative order, and the y-axis shows the number of days until receipt of the n-th order. The bold line in the middle plots the average, and the other two dotted lines plot the 95% confidence interval. Panel A includes the full unbalanced panel of all listings that ever appear in the transaction data. Panel B restricts the sample to listings that accumulate more than 10 orders during the six-month period.

Figure 5: Model Simulation

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



Panel B. Simulated Visibility Distribution over Quality



Note: This figure presents the simulation results from our counterfactual exercise of reducing the number of listings. 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 our baseline parameter estimates in Table 5.

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

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

Note: This table reports the summary statistics for the children's T-shirt market on AliExpress based on the store-listing-level data described in Section 2.2. Panel A reports the summary statistics at the listing level. Panel B reports the summary statistics at the store level for stores carrying these listings. Listing and store-level ratings are only available for those with reviews.

Table 2: Summary Statistics of the Quality Measures

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

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

Table 3: Treatment Effects of Order and Review

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

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

Table 4: Average Treatment Effects Measured at the Endline

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

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

Table 5: Estimated Parameters of the Empirical Model

Parameters	v_0	σ	ρ	λ
Value	0.26	5.39	0.48	0.97
S.E.	(0.029)	(0.025)	(0.006)	(0.011)

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	0.344
Top 5% cumulative revenue share	0.608	0.612
Top 10% cumulative revenue share	0.745	0.739
Top 25% cumulative revenue share	0.898	0.894
Top 50% cumulative revenue share	0.974	0.974
2. Dependence of new order on cumulative orders (v_0)	0.102	0.135
3. Quality and sales relationship (σ)		
Cumulative orders share: Top 1/3 quality bin	0.434	0.443
Cumulative orders share: Middle 1/3 quality bin	0.311	0.309
4. Reg. coef. of log price and quality (ρ)	0.125	0.132

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

Table 7: Model Validation Using the Experiment

Percent of Sellers Purchased	Size of Purchase	Average Effect on Sales: Treated - Control
P	O	$\Delta M = 41.9\%$
4	**1**	**0.129**
4	2	0.255
4	5	0.632

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

Table 8: Counterfactual Analyses

	Total Share for Top 1/3 Quality	Sales-Weighted Quality	Average Consumer Surplus
	Panel A: Search Friction: the Number of Sellers		
A.1 With $\sigma = 5.39$			
10000 Sellers	0.443	0.084	0.692
5000 Sellers	0.523	0.258	0.738
A.2 With $\sigma = 0$			
10000 Sellers	0.885	1.111	1.097
5000 Sellers	0.910	1.181	1.108
	Panel B: Information Friction: the Review Noise		
B.1 Sales-based sampling, 10000 Sellers			
$\sigma = 5.39$	0.443	0.084	0.692
$\sigma = 0$	0.885	1.111	1.097
B.2 Random sampling, 10000 Sellers			
$\sigma = 5.39$	0.302	-0.220	0.301
$\sigma = 0$	0.458	0.146	0.355

Note: This table reports the results of several counterfactual exercises based on the estimated model. Panel A.1 compares market outcomes when the number of product listings is reduced from 10,000 (default) to 5,000, with the baseline estimated level of review noise held fixed. Panel A.2 makes the same comparison but under perfect signals, where the first review fully reveals the listing's true quality. Panel B.1 compares market outcomes with noisy and perfect signals under the default sales-based sampling protocol, where visibility depends on the sum of v_0 and cumulative sales. Panel B.2 makes the same comparison but with a random sampling protocol, where consumers randomly sample listings to form their search sets. Section 7.1 describes the counterfactual exercises in more detail.

Table 9: Policy Counterfactual – Market Segmentation

Time of Segmentation	$T = 10,000$	$T = 50,000$	$T = 100,000$	$T = 200,000$	$T = 300,000$	**No Segmentation**
Statistics						
Average Quality of Purchase	0.386	0.492	0.517	0.504	0.456	0.084
Average Consumer Surplus	0.786	0.827	0.835	0.818	0.792	0.692
Average Quality of Segment	0.036	0.096	0.131	0.165	0.184	-
Moments						
Top 1% cumulative revenue share	0.422	0.403	0.394	0.391	0.388	0.344
Top 5% cumulative revenue share	0.709	0.707	0.710	0.705	0.700	0.612
Top 10% cumulative revenue share	0.823	0.833	0.838	0.837	0.832	0.739
Top 25% cumulative revenue share	0.938	0.948	0.952	0.953	0.951	0.894
Top 50% cumulative revenue share	0.986	0.989	0.990	0.990	0.990	0.974
Cumulative orders share: Top 1/3 quality bin	0.592	0.645	0.681	0.716	0.726	0.443
Cumulative orders share: Middle 1/3 quality bin	0.190	0.202	0.196	0.121	0.131	0.309

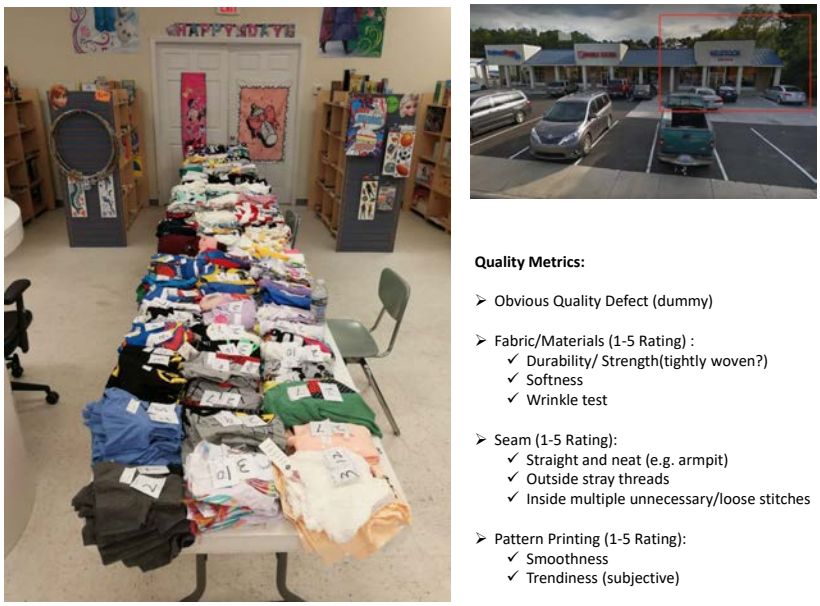
Note: This table reports the simulated market outcomes for a potential platform policy intervention that selects a subset of sellers to establish a premium market segment. Specifically, the policy selects the 5,000 sellers with the highest number of cumulative orders and non-negative expected qualities to form the new market segment. Each of the first five columns considers a different point in time at which the policy intervention takes place, while for comparison, the last column provides the baseline market outcomes without the policy intervention. Section 7.2 describes the policy intervention in more detail.

Appendices. For Online Publication Only

A Figures and Tables

Figure A.1: Product Quality Measurement

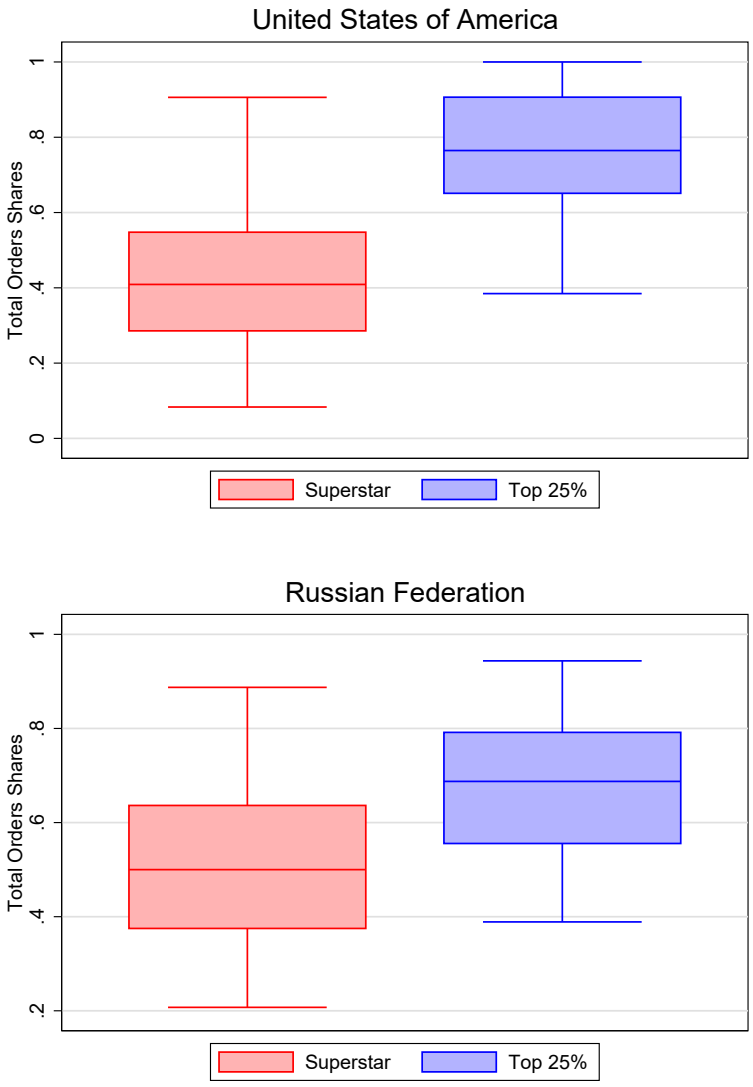
Panel A. Inspection and Grading



Panel B. Variation in Scores



Figure A.2: Sales Performance within Identical Varieties By Country



Note: This figure plots the distribution of within-group market shares, measured in cumulative orders, of the top listings in a group by country. The number of cumulative orders is constructed from the transaction data from March to August 2018, in which we observe the buyer's country of origin. Superstar indicates the listing that has the highest number of cumulative orders within its group for a given country. Top 25% indicates listings in the top 25th percentile in terms of cumulative orders. We limit the sample to groups with at least 4 listings in total and at least 1 listing with positive sales to a given country.

Table A.1: 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.2: 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.3: 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.4: Dependence of New Order Arrival on Cumulative Orders

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

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

Table A.5: 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.6: Treatment Effects of Order and Review: Without Baseline Controls

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

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

Table A.7: 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.8: Dynamic Treatment Effects

	All Destinations		English-speaking Countries		United States	
OrderXMonth1	0.063***	0.065***	0.037***	0.038***	0.038***	0.038***
	(0.024)	(0.021)	(0.008)	(0.008)	(0.007)	(0.007)
OrderXMonth2	0.017	0.020	0.006	0.008	0.007**	0.008**
	(0.032)	(0.027)	(0.010)	(0.009)	(0.003)	(0.004)
OrderXMonth3	0.008	0.011	0.015*	0.016**	0.012*	0.013**
	(0.020)	(0.017)	(0.008)	(0.008)	(0.007)	(0.006)
OrderXMonth4	-0.047	-0.044	-0.009	-0.007	0.011	0.011
	(0.031)	(0.029)	(0.019)	(0.018)	(0.008)	(0.008)
Observations	10192	10192	10192	10192	10192	10192
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the dynamic treatment effects of the experimentally generated orders and reviews. The dependent variable is the weekly number of orders. The baseline controls include the baseline total number of cumulative orders of the store and of the particular product listing. MonthX is a dummy variable that equals one for the X-th month after initial order placement. Standard errors clustered at the listing level are in parentheses. *** implies significance at the 0.01 level, ** at 0.5, and * at 0.1.

Table A.9: Seller Responses after Treatments

Panel A: Pricing Behavior

	AdjustPrice		CutPrice		RaisePrice		ΔLogPrice	
Order	0.022	0.003	0.033	0.008	0.022	0.007	-0.003	-0.005
	(0.035)	(0.044)	(0.038)	(0.046)	(0.036)	(0.045)	(0.009)	(0.011)
Observations	716	456	716	456	716	456	716	456
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Panel B: Shipping Costs

	AdjustShippingCost		CutShippingCost		RaiseShippingCost		ΔLogShippingCost	
Order	-0.010	-0.040	0.010	-0.029	-0.027	-0.040	-0.022	-0.008
	(0.034)	(0.037)	(0.030)	(0.034)	(0.029)	(0.032)	(0.039)	(0.044)
Observations	715	456	715	456	715	456	715	456
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Panel C: Product Description and Introduction of New Listings

	ChangeTitle		ChangeDescription		ChangeNumPictures		LogNewListings	
Order	0.001	0.000	-0.008	-0.008	-0.004	-0.008	-0.096	-0.089
	(0.011)	(0.011)	(0.019)	(0.019)	(0.014)	(0.014)	(0.080)	(0.067)
Observations	764	764	784	784	763	763	758	758
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

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

Table A.10: Heterogeneous Treatment Effects Based on Quality

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

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

Table A.11: Robustness: Structural Estimation

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

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

Table A.12: Counterfactual Robustness

	Total Share for Top 1/3 Quality	Sales-Weighted Quality	Average Consumer Surplus
		Panel A: Baseline	
10000 Sellers	0.443	0.084	0.692
5000 Sellers	0.523	0.258	0.738
		Panel B: $K \sim$ Positive Poisson(5)	
10000 Sellers	0.416	0.029	0.667
5000 Sellers	0.505	0.222	0.736
		Panel C: Price Elasticity=4 ($\gamma = 0.75$)	
10000 Sellers	0.439	0.080	0.707
5000 Sellers	0.525	0.274	0.762
		Panel D: Price Elasticity=10 ($\gamma = 1.89$)	
10000 Sellers	0.454	0.112	0.718
5000 Sellers	0.539	0.291	0.767

Note: This table reports the counterfactual outcomes from reducing the number of sellers under the baseline and alternative model specifications. Panel A reproduces our baseline estimates reported in Panel B.1 of Table 8. In Panel B, the parameters used to simulate the counterfactual are reestimated with K drawn from a positive Poisson distribution with mean 5. Similarly, Panels C and D use the reestimated parameters under different values of γ, corresponding to different price elasticities. The new parameter estimates and their corresponding model fitness are reported in Table A.11.

Table A.13: Robustness: Visibility Function Based on Cumulative Orders and Reviews

	$\zeta = 0.1$	$\zeta = 0.2$
Re-calibrated Parameters		
v_0	0.260	0.260
λ	0.900	0.800
σ	8.000	15.000
ρ	0.481	0.450
Moments		
Top 1% cumulative revenue share	0.340	0.314
Top 5% cumulative revenue share	0.608	0.611
Top 10% cumulative revenue share	0.740	0.761
Top 25% cumulative revenue share	0.899	0.927
Top 50% cumulative revenue share	0.976	0.985
Dependence of new order on cumulative orders	0.138	0.144
Cumulative orders share: Top 1/3 quality bin	0.459	0.444
Cumulative orders share: Middle 1/3 quality bin	0.292	0.296
Reg. coef. of log price and quality	0.131	0.120
Counterfactual:		
Baseline: $N = 10,000$		
Cumulative orders share: Top 1/3 quality bin	0.459	0.444
Average Quality of Purchase	0.131	0.098
Average Consumer Surplus	0.650	0.598
Smaller Market: $N = 5,000$		
Cumulative orders share: Top 1/3 quality bin	0.554	0.509
Average Quality of Purchase	0.323	0.238
Average Consumer Surplus	0.705	0.637

Note: This table considers alternative sampling weights, where the probability that seller i enters the consumer search sample depends on both its cumulative orders and its past reviews, as described in Appendix D.4.

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

C.1 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 sample 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 resampling 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 search 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:

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

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

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

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.[34]

[34] 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.

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 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.[35]

[35] 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

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.

C.2 Proof of Proposition 2: The Case of $\lambda = 1$

Throughout this proof, the true qualities are fixed such that seller 1 has the unique highest quality, i.e., $q^1 > q^i$ for every $i > 1$. We first make some preliminary observations. Because each review is independent from the entire past history, the joint stochastic processes of purchases and reviews can be disentangled such that reviews are drawn *ex ante* and purchase decisions then made based on these fixed reviews and induced beliefs. To be more specific, for every seller i, let $Y_1^i, Y_2^i, \ldots$ be i.i.d. random variables with the distribution $q^i + \mathcal{N}(0, \sigma^2)$. We think of Y_m^i as the m-th review for seller i. These reviews generate beliefs

$$\widehat{q_t^i} = \frac{(Y_1^i + \cdots + Y_m^i)/\sigma^2}{1 + m/\sigma^2},$$

where t is the period in which seller i is chosen for the m-th time. Using these beliefs, the sampling and choice procedures then iteratively determine which seller is chosen in each period. Note that it is possible for a seller i to be chosen only finitely often along certain histories. In these situations, some reviews Y_m^i and the resulting beliefs are never realized, but ex ante, we could still think of them as the reviews and beliefs that would have been realized.

Recall that we want to show that the highest quality seller 1 *almost surely* obtains full market share in the long run. Thus, we can without loss focus on the probability-one event that the time-averaged review $\frac{Y_1^i + \cdots + Y_m^i}{m}$ converges to the true quality q^i for every seller i (by the strong law of large numbers). It suffices to *fix any such collection of reviews* and show that with conditional probability one, seller 1 obtains full market share in the long run.

Henceforth, the reviews are fixed, with the property that the time-averaged reviews converge to the true qualities. Since the absolute value of belief $\widehat{q_t^i} = \frac{(Y_1^i + \cdots + Y_m^i)/\sigma^2}{1 + m/\sigma^2}$ is bounded above by

$\widehat{q_{T-1}^{i_T^j}}$ all have a normal distribution. Thus, the dominated function has a finite expectation as well.

the absolute value of $\frac{Y_1^i+\cdots+Y_m^i}{m}$, and since the latter is bounded as a convergent sequence, we deduce that all beliefs about all sellers are bounded between $-M$ and M for some large positive constant M.

We use this bounded-belief property to show that almost surely every seller is chosen infinitely often. Indeed, in any period $t+1$, the sampling probability of seller i is $\frac{v_t^i}{\sum_j v_t^j} = \frac{v_t^i}{Nv_0+t} \geq \frac{v_0}{Nv_0+t}$. Thus, the expected number of copies of seller i in the sample of size K is at least $\frac{Kv_0}{Nv_0+t}$. On the other hand, under logit choice, each copy of seller i is chosen out of the sample with probability at least $\frac{e^{-M}}{Ke^M}$ (since beliefs are bounded between $-M$ and M). Putting it together, we know that the choice probability of seller i in period $t+1$ is at least

$$\frac{Kv_0}{Nv_0+t} \cdot \frac{e^{-M}}{Ke^M} = \frac{e^{-2M}v_0}{Nv_0+t} = \frac{c}{Nv_0+t}$$

for some positive constant c. Because of this, the probability that seller i is *not chosen after some period* T is at most $\prod_{t\geq T}(1-\frac{c}{Nv_0+t}) = 0$, where the infinite product is zero because the harmonic sum $\sum_{t\geq T}\frac{c}{Nv_0+t}$ diverges. Since this holds for every T, a union bound implies that the probability that seller i is chosen only finitely many times is zero. This is what we desired to show.

Next, as the time-averaged reviews $\frac{Y_1^i+\cdots+Y_m^i}{m}$ converge to the true quality q^i as $m\to\infty$, so do the induced beliefs $\frac{(Y_1^i+\cdots+Y_m^i)/\sigma^2}{1+m/\sigma^2}$. Let δ be a positive number smaller than q^1-q^i for every $i>1$. Then, there exist positive integers $m_1,\ldots,m_N$ such that at any period t when each seller i has been chosen at least m_i times, the belief difference $\widehat{q_t^1}-\widehat{q_t^i}$ between seller 1 and every other seller is at least δ. For every positive integer T, we let $\mathcal{H}_T$ be the set of all histories such that each seller i is chosen at least m_i times by the end of period T but not by the end of period $T-1$. Then, these events $\mathcal{H}_T$ are disjoint and together constitute a probability-one event because each seller is chosen infinitely often. It thus suffices to show that within each $\mathcal{H}_T$, seller 1's market share almost surely converges to one.

We are now in a position to present the main part of the proof. Based on the above analysis, we now fix T and also fix the purchase decisions in the first T periods, with the property that each seller i has been chosen at least m_i times. Given this property, the belief difference between seller 1 and every other seller is at least δ *in all future periods*, regardless of the random future purchase decisions. We use this property to show that seller 1's fraction of total sales $\frac{s_t^1}{t}$ converges to one almost surely. It turns out to be more convenient to work with seller 1's fraction of total visibility, i.e.,

$$r_t^1 = \frac{v_t^1}{\sum_j v_t^j} = \frac{v_0+s_t^1}{Nv_0+t} \quad \in [0,1].$$

This coincides with the sampling probability $R_t^1 = (v_t^1)^\lambda / \sum_j (v_t^j)^\lambda$ precisely when $\lambda = 1$.

It is easy to see $\frac{s_t^1}{t} \to 1$ if and only if $r_t^1 \to 1$ as $t \to \infty$. To show $r_t^1 \to 1$, we note that the key idea is when $\lambda = 1$, r_t^1 is the probability of sampling seller 1 in period $t+1$. However, since logit choice favors higher expected quality, and since seller 1 actually has the highest belief, the probability of *choosing* seller 1 in period $t+1$ exceeds r_t^1. This then implies that in expectation r_{t+1}^1 exceeds r_t^1, making the process $\{r_t^1\}$ a bounded *submartingale*. Such a process almost surely converges, and with some additional analysis, we can show that it converges to the maximum value of 1.

To make this argument more precise, we begin by calculating the probability of choosing seller 1 in any period $t+1$ (with $t \geq T$). Since seller 1's expected quality exceeds that of other sellers by at least δ, it is easy to see that in any sample of K sellers where seller 1 appears $k < K$ times, the logit probability of seller 1 being chosen from the sample is at least $\frac{k}{k+(K-k)e^{-\delta}}$. This is at least $\frac{(1+\eta)k}{K}$ for some positive constant η that depends on δ. Hence, by considering different k and incorporating the corresponding sampling probabilities, we know that the overall choice probability of seller 1 in period $t+1$ satisfies

$$\begin{aligned}
p_{t+1}^1 &\geq (r_t^1)^K + \sum_{1 \leq k \leq K-1} \binom{K}{k} (r_t^1)^k (1-r_t^1)^{K-k} \cdot \frac{(1+\eta)k}{K} \\
&= -\eta (r_t^1)^K + \sum_{1 \leq k \leq K} \binom{K}{k} (r_t^1)^k (1-r_t^1)^{K-k} \cdot \frac{(1+\eta)k}{K} \\
&= -\eta (r_t^1)^K + (1+\eta) r_t^1 \sum_{1 \leq k \leq K} \binom{K-1}{k-1} (r_t^1)^{k-1} (1-r_t^1)^{K-k} \qquad \text{(C.3)} \\
&= -\eta (r_t^1)^K + (1+\eta) r_t^1 \\
&\geq -\eta (r_t^1)^2 + (1+\eta) r_t^1 \\
&= r_t^1 (1 + \eta(1 - r_t^1)).
\end{aligned}$$

With probability p_{t+1}^1, seller 1's visibility increases by 1, and thus $r_{t+1}^1 = \frac{v_t^1+1}{V_t+1} = \frac{r_t V_t + 1}{V_t+1}$, where we define

$$V_t = \sum_j v_t^j = N v_0 + t.$$

Otherwise, $r_{t+1}^1 = \frac{v_t^1}{V_t+1} = \frac{r_t V_t}{V_t+1}$. Since $p_{t+1}^1 \geq r_t^1$ as given by (C.3), it is straightforward to verify $\mathbb{E}[r_{t+1}^1 \mid r_t^1] \geq r_t^1$, so that $\{r_t^1\}$ is indeed a submartingale process that converges almost surely.

However, to show that the limit is almost surely one, we in fact consider a different submartingale process given by $\log(r_t^1) - \frac{1}{v_t^1}$. Its evolution is such that with probability p_{t+1}^1, we have $\log(r_{t+1}^1) - \frac{1}{v_{t+1}^1} = \log(\frac{v_t^1+1}{V_t+1}) - \frac{1}{v_t^1+1}$ and otherwise $\log(r_{t+1}^1) - \frac{1}{v_{t+1}^1} = \log(\frac{v_t^1}{V_t+1}) - \frac{1}{v_t^1}$. From this,

we calculate

$$\begin{aligned}\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] &= \log(\frac{v^1_t}{V_t+1}) - \frac{1}{v^1_t} + p^1_{t+1} \cdot \left(\log(\frac{v^1_t+1}{v^1_t}) + \frac{1}{v^1_t(v^1_t+1)}\right) \\ &= \log(r^1_t) - \frac{1}{v^1_t} - \log(\frac{V_t+1}{V_t}) + p^1_{t+1} \cdot \left(\log(\frac{v^1_t+1}{v^1_t}) + \frac{1}{v^1_t(v^1_t+1)}\right) \\ &> \log(r^1_t) - \frac{1}{v^1_t} - \frac{1}{V_t} + p^1_{t+1} \cdot \frac{1}{v^1_t} \\ &= \log(r^1_t) - \frac{1}{v^1_t} + \frac{1}{V_t}\left(\frac{p^1_{t+1}}{r^1_t} - 1\right),\end{aligned} \tag{C.4}$$

where the penultimate inequality uses $\log(\frac{V_t+1}{V_t}) < \frac{1}{V_t}$ and $\log(\frac{v^1_t+1}{v^1_t}) = -\log(1 - \frac{1}{v^1_t+1}) > \frac{1}{v^1_t+1}$.

Since (C.3) implies $p^1_{t+1} \geq r^1_t$, the above inequality (C.4) tells us that $\{\log(r^1_t) - \frac{1}{v^1_t}\}_{t \geq T}$ is a submartingale that is bounded above by zero. Thus, by the Martingale Convergence Theorem, $\log(r^1_t) - \frac{1}{v^1_t}$ almost surely converges. As $\frac{1}{v^1_t}$ converges as a monotonically decreasing sequence, we deduce that $\log(r^1_t)$ and thus r^1_t almost surely converges to a limit random variable r^1_∞.

Now, suppose r^1_∞ has a positive probability of being less than 1. Then, there exists $\epsilon > 0$ such that $\mathbb{P}\{r^1_\infty < 1 - \epsilon\} > \epsilon$. Because r^1_t converges almost surely to r^1_∞, the weaker notion of convergence in distribution also holds. We thus have

$$\liminf_{t \to \infty} \mathbb{P}\{r^1_t < 1 - \epsilon\} \geq \mathbb{P}\{r^1_\infty < 1 - \epsilon\} > \epsilon.$$

Hence, for this fixed ϵ, we can find τ sufficiently large such that $\mathbb{P}\{r^1_t < 1 - \epsilon\} > \epsilon$ for every $t \geq \tau$. Since $p^1_{t+1} \geq r^1_t(1 + \eta(1 - r^1_t))$ as given by (C.3), there is a probability of at least ϵ that $p^1_{t+1} \geq (1 + \eta\epsilon)r^1_t$. Thus, in (C.4), there is a probability of at least ϵ that

$$\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t}) \geq \frac{\eta\epsilon}{V_t}.$$

With the remaining probability, the difference $\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t})$ is nonnegative, and so taking an ex ante expectation yields

$$\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}}] - \mathbb{E}[(\log(r^1_t) - \frac{1}{v^1_t})] \geq \frac{\eta\epsilon^2}{V_t}.$$

This has to hold for every $t \geq \tau$, which implies by telescoping that

$$\mathbb{E}[\log(r^1_s) - \frac{1}{v^1_s}] - \mathbb{E}[(\log(r^1_\tau) - \frac{1}{v^1_\tau})] \geq \sum_{t=\tau}^{s-1} \frac{\eta\epsilon^2}{V_t} = \sum_{t=\tau}^{s-1} \frac{\eta\epsilon^2}{Nv_0 + t}.$$

However, this is a contradiction because the LHS remains bounded as $s \to \infty$ whereas the RHS is a harmonic sum that diverges. Therefore $r^1_\infty = 1$ almost surely, completing the proof.

C.3 Proof of Proposition 2: The Case of $\lambda > 1$

We show that the market is inefficient when $\lambda > 1$ by demonstrating that there is a positive probability that any seller $i > 1$ is chosen in each period. Indeed, the probability that seller i is sampled K times in each period is the following product (independent of reviews):

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

Since the infinite sum $\sum_t \frac{(N-1)(v_0)^\lambda}{(v_0+t)^\lambda+(N-1)(v_0)^\lambda}$ converges for any $\lambda > 1$ (just as $\sum_t \frac{1}{t^\lambda}$ converges), the infinite product $\prod_{t\geq 0}\left(1-\frac{(N-1)(v_0)^\lambda}{(v_0+t)^\lambda+(N-1)(v_0)^\lambda}\right)$ is strictly positive. Hence, there is a positive probability that all consumers sample (and thus choose) only seller i, completing the proof.

C.4 Proof of Proposition 2: The Case of $\lambda < 1$

We show that a different kind of inefficiency arises when $\lambda < 1$ that involves the market shares of all sellers remaining positive in the long run. Formally, we follow the earlier proof for the case of $\lambda = 1$ and fix the reviews such that the time-averaged reviews converge to the true qualities. We show that conditional on any such collection of reviews, seller i's fraction of total visibility $r^i_t = \frac{v^i_t}{V_t} = \frac{v_0+s^i_t}{NV_0+t}$ satisfies $\liminf_{t\to\infty}\mathbb{E}[r^i_t] > 0$. This would imply that seller i's fraction of total sales also satisfies $\liminf_{t\to\infty}\mathbb{E}[\frac{s^i_t}{t}] > 0$. In particular, $\frac{s^i_t}{t}$ cannot converge to 0 in probability, and thus $\frac{s^1_t}{t}$ cannot converge to 1 in probability, as efficiency would require.

Since time-averaged reviews converge, all beliefs for all sellers are bounded between $-M$ and M for some M. Thus, the probability that seller i is chosen in period $t+1$ satisfies

$$p^i_{t+1} \geq K \cdot \frac{(r^i_t)^\lambda}{\sum_j (r^j_t)^\lambda} \cdot \frac{e^{-M}}{Ke^M},$$

where $\frac{(r^i_t)^\lambda}{\sum_j (r^j_t)^\lambda} = \frac{(v^i_t)^\lambda}{\sum_j (v^j_t)^\lambda}$ is the probability of sampling seller i in each of the K slots and $\frac{e^{-M}}{Ke^M}$ is a lower bound on the logit probability of choosing seller i in that slot. Since each r^j_t belongs to $(0,1)$, we have $\sum_j (r^j_t)^\lambda \leq N$. Thus, there exists a positive constant $\eta = \frac{1}{Ne^{2M}}$ such that $p^i_{t+1} \geq \eta \cdot (r^i_t)^\lambda$. It follows that

$$\frac{p^i_{t+1}}{r^i_t} \geq \eta \cdot (r^i_t)^{\lambda-1} \geq \eta(\lambda-1)\cdot \log(r^i_t),$$

where the second inequality follows from the fact that $x^{\lambda-1} \geq (\lambda-1)\log(x)$ for every $x \in (0,1)$.[36] Since $\log(r_t^i) \leq \log(r_t^i) - \frac{1}{v_t^i} + \frac{1}{v_0}$, we further deduce

$$\mathbb{E}\left[\frac{p_{t+1}^i}{r_t^i}\right] \geq \eta(\lambda-1) \cdot \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] + \frac{\eta(\lambda-1)}{v_0}. \tag{C.5}$$

Now, by the analogue of (C.4) for seller i and taking an ex ante expectation, we have

$$\mathbb{E}\left[\log(r_{t+1}^i) - \frac{1}{v_{t+1}^i}\right] > \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] + \frac{1}{V_t} \cdot \mathbb{E}\left[\frac{p_{t+1}^i}{r_t^i} - 1\right]. \tag{C.6}$$

Thus, (C.5) and (C.6) together imply that $\mathbb{E}\left[\log(r_{t+1}^i) - \frac{1}{v_{t+1}^i}\right] > \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right]$ whenever the RHS of (C.5) exceeds 1. More specifically, we have shown that

$$\mathbb{E}\left[\log(r_{t+1}^i) - \frac{1}{v_{t+1}^i}\right] \text{ is larger than } \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] \text{ whenever the latter is smaller than } -c, \tag{C.7}$$

with $c = \frac{1}{v_0} + \frac{1}{\eta(1-\lambda)}$ being a constant.

Let us then choose as C a larger constant such that $C > c + \frac{1}{v_0}$ and $C > -\mathbb{E}\left[\log(r_0^i) - \frac{1}{v_0}\right]$. We show by induction that $\mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] > -C$ holds in every period t. The base case $t = 0$ holds by the choice of C, so we focus on the induction step and suppose $\mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] > -C$. There are two possibilities: either $\mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] < -c$, or $\mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] \geq -c$. In the first possibility, we know by (C.7) that $\mathbb{E}\left[\log(r_{t+1}^i) - \frac{1}{v_{t+1}^i}\right] > \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right]$, which is in turn larger than $-C$ by the induction hypothesis. In the latter possibility, we can instead deduce from (C.6) that

$$\mathbb{E}\left[\log(r_{t+1}^i) - \frac{1}{v_{t+1}^i}\right] > \mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] - \frac{1}{V_t} \geq -c - \frac{1}{v_0} > -C.$$

Thus, the induction step is valid, and we always have $\mathbb{E}\left[\log(r_t^i) - \frac{1}{v_t^i}\right] > -C$.

It follows that $\mathbb{E}\left[\log(r_t^i)\right] > -C$ in every period t, and thus by Markov's inequality, there is a probability of at most $1/2$ that $r_t^i \leq e^{-2C}$. Thus, there is a probability of at least $1/2$ that $r_t^i > e^{-2C}$. This implies the desired conclusion that $\mathbb{E}[r_t^i] > e^{-2C}/2$ is bounded away from zero, completing the proof.

[36] This inequality clearly holds at $x = 1$. Moreover, for $x \in (0,1)$, the derivative of $x^{\lambda-1}$ is $(\lambda-1)x^{\lambda-2}$ and is more negative than $(\lambda-1)x^{-1}$, the derivative of $(\lambda-1)\log(x)$. Thus, the inequality holds for every $x \in (0,1)$.

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).

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.

Simulate Order and Review. In each period, we use the weighted sampling without replacement to generate the consumer's search sample of size K. Based on its average reviews, we calculate each sampled seller's expected quality and the expected utility of purchasing. 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 Welfare 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 search sample.

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 search sample $(i_1, i_2, \ldots, i_K)$ is

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

where

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

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

D.4 Alternative Specification of 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.13, 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.13 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 7.2 remain robust.