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

Jie Bai, Maggie Chen, Jin Liu, Xiaosheng Mu, Daniel Yi Xu *

July 3, 2021

Abstract

We investigate how search and information frictions shape firm dynamics and market evolution in global e-commerce. First, using detailed data from AliEpxress as well as a rich set of self-collected objective quality measures, we provide stylized facts that are consistent with the presence of search and information frictions. Next, through a randomized experiment that offers exogenous demand and information shocks to small prospective exporters, we establish that accumulating sales helps new exporters' overcome these demand frictions by enhancing their visibility and allow them to attract more future demand and grow. We prove theoretically that this reinforcement mechanism, combined with optimal consumer choice, are important for achieving long-run allocative efficiency. In the short run, however, having a large number of online sellers can congest the consumer search process, slow down the resolution of information problem, and hinder market allocation towards better sellers. Guided by the theoretical findings, we construct and estimate an empirical model of the online market to quantify the extent of these demand-side frictions. Counterfactual analyses show that reducing the number of sellers and alleviating information friction could help to improve market efficiency and raise consumer welfare.

Keywords: e-commerce, exporter dynamics, product quality, information frictions, search frictions

JEL classification: F14, L11, O12

*Contact: Bai: Harvard Kennedy School and NBER (email: jie_bai@hks.harvard.edu); Chen: Department of Economics, George Washington University (email: xchen@gwu.edu); Liu: New York University (email: jl7767@nyu.edu); Mu: Department of Economics, Princeton University (email: xmu@princeton.edu); Xu: Department of Economics, Duke University and NBER (email: daniel.xu@duke.edu). We thank Costas Akolakis, David Atkin, Lauren Bergquist, Ben Faber, Gordon Hanson, Chang-tai Hsieh, Panle Jia, Asim Khwaja, Pete Klenow, Meredith Startz, Tianshu Sun, Christopher Synder, and seminar and conference participants at the Berkeley Economics, BREAD/CEPR/STICERD/TCD Conference, University of Chicago, Harvard Kennedy School, Hong Kong Trade Workshop, IPA SME Working Group Meeting, Michigan Economics, NBER Conferences (China Working Group, DEV), and USC for helpful comments. We thank Chengdai Huang, Haoran Zhang, and Qiang Zheng for excellent research assistance. All errors are our own.

1 Introduction

E-commerce sales have grown tremendously in recent years, reaching $2.9 trillion in 2018 and 12 percent of the total global retail sales (Lipsman, 2019). Within e-commerce, cross-border sales have grown two times faster than domestic sales, and nearly 40 percent of online buyers completed a cross-border transaction in 2016 (Pitney Bowes, 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 into export markets. Furthermore, online exporting lowers many of the traditional barriers of offline exporting, including the needs of building export relationships and setting 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 SMEs in developing countries to 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 such increased export opportunity on firm growth and market dynamics. While e-commerce platforms potentially expose 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, these search frictions can further slow down the resolution of the information problem and hinder market allocation towards better firms.

In this study, we experimentally, theoretically, and quantitatively investigate these mechanisms. We first document descriptive evidence that is 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 to 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. Finally, building on the theoretical framework, we estimate a rich empirical model of the online export market. We 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 businesses beyond the initial onboarding stage and improve the 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

[1] For example, AliExpress, one of the leading cross-border e-commerce platform that we study in this project, states on its website (`https://sell.aliexpress.com/__pc/4DYTFsSkV0.htm`): "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 idea can be traced to Stigler (1961) that consumers not perfectly informed about all products available for purchase in a given market and can only consider a limited subset.

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, multi-dimensional 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 sales distribution within "identical-looking" product varieties. Interestingly, even after controlling for horizontal taste differences, meaningful dispersion in sales remains within identical variety groups, as opposed to "winner-takes-all". This finding is indicative of search frictions: buyers, upon arriving at the platform, face thousands of product offerings but can only consider a limited finite subset of all seller-listings.[3] This raises the question of who gets to grow in the presence of the search problem. 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 (nor the lowest price). Intuitively, search friction introduces a random component in firm growth due to the consumer sampling process. When firms' intrinsic quality is not perfectly observed (even upon entering into a consumer's consideration set), such friction 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, hence higher visibility, have an advantage in overcoming the consumer's search friction and generating future orders. However, if information friction prevents a firm's visibility from being aligned with its fundamentals, it could take much longer for better firms to stand out. Overall, market allocation and consumer welfare depend crucially on the interactions of these demand-side forces.

Our interpretation of the stylized facts centers around a demand driven reinforcement mechanism where each additional consumer order makes the selling firm more visible and hence helps the firm to overcome the search frictions faced by subsequent buyers. However, unobserved supply-side actions (such as advertising and display) could also exist and lead to 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. The treatment allows us to isolate the impact of demand from unobserved supply-side confounding factors. Since how effectively the additional demand conveys the firm's true fundamentals depends critically on the severity of information friction, we further interact the order treatment with a review treatment about firms' product and shipping quality to examine the

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

role of information provision. We find that the order treatment leads to a small but significantly positive impact on firms' subsequent sales. This demonstrates that indeed a key channel for firms to improve their visibility and grow in the online marketplace is by accumulating sales. Quantile analysis reveals, however, that the effect is mainly concentrated at the top: only a small fraction of sellers are able to take advantage of the initial demand shock and grow while the vast majority stay small. The size of the average treatment effect suggests that these demand-side frictions cannot be easily overcome by individual sellers' private efforts. 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 friction. Intuitively, reviews only matter when a seller's listing is discovered by consumers, which is a rare event for small businesses due to their low visibility. Finally, we do not find significant heterogeneous treatment effect based on quality. This echoes the stylized fact that quality does not strongly predict growth in this market due to the search and information problems, which, combined, make it difficult for high-quality sellers to stand out.

All together, the descriptive and experimental findings are consistent with the presence of sizable search and information frictions and highlight an important reinforcement mechanism of online firm growth through accumulating sales. Motivated by the reduced form evidence, we develop a theoretical model that formalizes the process of demand accumulation under consumer search and learning. 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 its cumulative sales. Consumers make purchase decisions based on expected quality within the search sample. We show that this "cumulative-sales-based" sampling is a powerful tool to overcome the demand-side frictions and achieve efficient allocation in the long run, that is, the highest quality firm will have dominant market position. This is true *even* when consumers do not perfectly observe a product's quality and have to rely on noisy review signals by previous buyers. Although efficiency is achieved in the long run, search and information frictions remain crucial determinants of the short-run market outcomes and the speed of convergence towards efficient allocation. Importantly, we show that increasing the number of sellers can congest the consumer search process and slow down the rise of high-quality sellers.

Building on the theoretical framework, we estimate a rich structural model of the online market incorporating the realistic frictions of the market. As in the theoretical model, we assume that consumers conduct a fixed sample search and consider a small finite set of listings. We estimate a linear sales-based visibility function to match the empirical market share distribution and how future order arrival depends on current cumulative sales. On the supply side, we extend the theoretical 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, striking a first order makes a seller 1.2 times more likely

to end up in a subsequent consumer's search sample. That said, uncertainty regarding the seller's quality still remains even after the seller successfully strikes the first sale due to the information problem. Our estimate of the review signal noise indicates substantial information frictions. The posterior uncertainty is only reduced by 7% after the first order, suggesting that the reputation mechanism takes time to play its role and the information problem only resolves slowly over time. Combined, these findings highlight that search friction, interacted with information friction, can constitute an important hurdle for the growth of small prospective exporters.

We end with several counterfactual exercises to examine the distinctive roles of search and information frictions in firm growth and market allocation and evaluate potential policy interventions using the estimated structural model. First, we investigate the impact of reducing search frictions by reducing the number of sellers operating on the platform. The results show that doing so helps mitigate the congestion in consumer search, thereby improving allocative efficiency and consumer welfare. This is especially true when there exists large information frictions. Intuitively speaking, since it takes multiple draws to reveal a listing's true quality, reducing the number of listings increases the number of quality signals each seller obtains and thus allows high-quality sellers to be discovered faster. Over time, sellers with higher quality receive higher visibility.[4] This result points out that just giving firms easy access to foreign markets alone may not be sufficient for generating sustained growth and can in fact exacerbate the search problem, resulting in short-run 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 to promising newcomers could help facilitate growth and improve the overall market efficiency.

Next, to shed light on the role of information frictions, we remove the noise of the review signals. We find that doing so significantly shifts market share to high-quality sellers. The resulting consumer surplus is 13.8% higher compared to the baseline. 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.

While our setting is specific to e-commerce, the economic insights generalize to broader market settings. It is well understood that there could 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 friction, the business stealing can happen when sellers compete for customer attention, beyond simple price competition, and such a business stealing effect could be particularly costly when there exists information frictions since it further prevents the best firms from being "discovered" and

[4] Given the small search sample size, relative to all products available, our model features limited gains from variety. We consider variety effects in Section 7.3. Our main empirical model focuses on the implications of excessive entry for search friction.

reduces the short-run allocative efficiency of the market. Several features of the cross-border e-commerce setting make it a good exhibit of these economic forces, including the lack of market selection, resulting in excessive entry, and differences in cultural and business practices across borders, aggravating the information problem.

Our work contributes to several strands of the existing literature. Extensive work in international trade has studied the empirical patterns of new exporter dynamics 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 theories have been proposed to explain these facts. They include firm learning (Arkolakis, Papageorgiou, and Timoshenko, 2018; Ruhl and Willis, 2017), demand accumulation(Foster, Haltiwanger, and Syverson, 2016; Piveteau, 2016; Fitzgerald, Haller, and Yedid-Levi, 2020), and seller search (Eaton et al., 2016). In contrast to these studies, our paper focuses on demand-side frictions, rather than the exporter's own decision, as the key driving force of firm and market dynamics in the online setting. In particular, unlike 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. Methodologically, we bring in new sources of variations to first experimentally identify the mechanisms underlying new exporter's demand accumulation process and then formally model the realistic frictions of the market. Our findings also connect to the existing literature in trade that examines the roles of search and information frictions on market demand and seller reliability in explaining price variations and trade patterns (Allen, 2014; Macchiavello and Morjaria, 2015; Steinwender, 2018; Startz, 2018).[5]

Another complementary literature explaining exporter performance highlights the role of quality (see Verhoogen, 2020, for an excellent review). Since product or service quality is rarely observed in standard firm surveys, most of the earlier literature has focused on indirect measures of quality estimated based on market shares and prices, for instance, (Verhoogen, 2008; Khandelwal, 2010). We build on a growing body of development research that collects detailed information on quality for specific industries (e.g, Bai, 2016; Atkin, Khandelwal, and Osman, 2017; Hansman et al., 2020). Similar to these earlier works in offline settings, we document large variations in firm-product quality online. However, we find quality plays a less pronounced role in explaining exporter growth and long-run market shares. Our paper explains the disintegration of the customer accumulation process and firm fundamentals, such as quality, and underscores the potential sources of market misallocation in the e-commerce context.

[5]Despite the growing importance of e-commerce in international trade, empirical work on the online setting has remained scarce and has so far primarily focused on patterns of online trade and the role of geographic distance (Hortaçsu, Martínez-Jerez, and Douglas, 2009; Lendle et al., 2016). In this paper, we examine exporter growth dynamics in online trade and study the roles of search and information frictions. We establish a set of new stylized facts about e-commerce exporters. These facts point to new trade models that extend the standard heterogeneous firm and trade framework to incorporate important features of the online marketplace.

Our study also relates to the existing literature on consumer search and consideration set (for example, Goeree (2008); Kim, Albuquerque, and Bronnenberg (2017); Honka, Hortaçsu, and Vitorino (2017); Dinerstein et al. (2018).)[6] Theoretically, we extend the 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 reputation mechanism (Ching, Erdem, and Keane, 2013; Tadelis, 2016). This allows us to examine the interaction between search and information frictions.[7] Empirically, we first experimentally identify the impact of accumulating sales on search, as reflected through future purchases. After that, we leverage the unique quality data to quantify the scope of market misallocation through the lens of the empirical model.

Last but not least, findings from our study also speak broadly to the development literature on interventions to help micro, small, and medium enterprises. Echoing the literature on productivity differences across firms, most of the earlier work has emphasized supply-side interventions, including providing credit access, quality inputs, and managerial training (e.g, De Mel, McKenzie, and Woodruff, 2008; Kugler and Verhoogen, 2012; Banerjee, 2013; Bloom et al., 2013). More recently, a growing set of work has begun to look at demand-side interventions. A closely related study to ours is Atkin, Khandelwal, and Osman (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 the impact of foreign demand shocks on search and information about the firm.

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 experiment. Section 4 describes the experiment design and main findings. Sections 6 and 6.3 build and estimate 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 setting of the study, the market of children's t-shirts on AliExpress, and describe the data collection.

[6] We refer interested readers to a recent review article by Honka, Hortacsu, and Wildenbeest (2019) for the broader literature.

[7] In a different setting, Pallais (2014) examines information friction in entry-level labor market and shows 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. The previous paper focuses on the information problem. We further introduce search friction and show that doing so exacerbates the initial hurdle for high-quality new businesses to stand out.

2.1 The Market of 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.[8] 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.[9] Most sellers on the platform are retailers rather than manufacturers, and they 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 marketing and shipping services.[10]

For this study, we focus on the industry of children's t-shirts. 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.[11] The growth and efficiency of the online retail market therefore matters for the manufacturing upstream: in particular, the growth of retailers that sell high quality products will in turn benefit their producers. The vibrant entry and growth dynamics in the online market also provide an ideal setting to study 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 the entry and fixed operation costs of exporting and allowing sellers, large and small, to tap into the export markets.[12] While this does help to 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 in the online marketplace competing for consumers' attention. The resulting welfare implication of having an increasing number of market participants on firms and consumers is far less clear in the presence of search and information frictions. This forms the key trade-off 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 *"group"*) may contain multiple listings, sold by different sellers, that

[8] Sources: https://sell.aliexpress.com/ and https://sell.aliexpress.com/__pc/4DYTFsSkV0.htm.

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

[10] 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, high-quality versus low-quality factories, and how much quality inspection effort the sellers put in. We document this formally using detailed quality measures we collect from the study (see Section 2.2).

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

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

share 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 analyses 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 the 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 universe of product listings in the industry of children's t-shirt in May 2018.[13] We collected all the information that a buyer can view on the listings' pages, including total cumulative orders (quantity sold), current prices, discounts (if any), ratings, buyer protection schemes (if any), and detailed product attributes. We further collected information about the stores that carry these products, including the year of opening and other products the stores carry.

(2) Transaction Records. For each product listing, 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 6-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 report only customer reviews and the total volume of transactions without the full transaction history (e.g., Amazon and eBay). The availability of the real-time transaction records enables us to closely track each product listing's sales activities over time.[14]

(3) Measures of Quality Finally, we complement the platform data with a rich set of objective quality measures we collected for a representative sub-sample of all product listings, covering quality of products measured in 8 dimensions, quality of shipping, and quality of seller service. These quality measures were collected through three channels: (i) actual purchase of the products, (ii) direct communications with the sellers, and (iii) third-party assessment. Appendix C.1 provides a detailed discussion of the quality measurement process. Table 1 presents summary statistics of the various quality measures.[15]

To measure product quality, we placed actual orders of children's t-shirts on AliExpress.[16] After receiving and cataloging the orders, we worked with a large local consignment store of children's clothing

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

[14]The transaction data omits information on price. To complement that, we further conducted a weekly data scraping from May to August 2018 for listings in the experimental sample to track price dynamics.

[15]To construct the quality indices, we first standardize the quality metric in each dimension and then average across all dimensions. Table B.1 decomposes the variation of the overall quality index to that explained by each individual quality metric.

[16]We placed an order on each listing in our experimental sample as well as their medium-size and superstar peers in the same variety group (see Section 4 for details on the sampling procedure).

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, quality was assessed along 8 dimensions: durability, fabric softness, wrinkle test, seams (straightness and neatness), outside stray threads, inside loose stitches, pattern smoothness, and trendiness. Figure 2 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, except trendiness, capture vertical quality differentiation. For example, at equal prices, consumers would prefer t-shirts with more durable fabric, straighter seams and fewer loose stray 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 2, there exists considerable quality difference both across and within variety groups, depending on which factories the retailers choose to source from and/or how much quality inspection effort is put in.[17]

To measure shipping quality, we recorded the date of each purchase, the date of shipment, the date of delivery, carrier name, and the condition of the package upon arrival. The information is used to construct four measures of shipping quality: (i) the time lag between order placement and shipping, (ii) the time lag between shipping and delivery, (iii) whether the package is delivered, and (iv) whether the package is 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. Appendix C.1 describes the messages. We rate service quality based on whether the message received a reply, the time it took to receive a reply, and whether the questions were acknowledged and properly addressed.

In Table B.2, we find all three quality indices — product, shipping and service — to be positively correlated with the online star ratings, although the correlations are relatively weak and only statistically significant for shipping and service qualities.

2.3 Summary Statistics

For most of the empirical analyses, we restrict the sample to variety groups with at least 100 cumulative sales (aggregated across all listings in the group) in order to focus attention on products that are more relevant for consumer choice. This leaves us with 133 variety groups, featuring 1265 product listings sold by 625 stores. This forms our study sample. Table 2 summarizes the product level (Panel A) and store level (Panel B) characteristics for the study sample.[18] There are 1265 product listings in the study

[17] To cross-validate the quality measures, we 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 quality metrics are strongly correlated with the subjective price evaluations.

[18] Table B.4 shows the summary statistics for all the listings and stores that appear in the original 4586 variety groups, including those with 0 group total sales.

sample. The average price is $5.3. About 48 percent of the listings offer free shipping, and the average shipping price to the US is $0.69. At the store level, there are 625 stores carrying these varieties. Most exporters are young with an average age of 1.29 years. The average cumulative sales is 9,090 with a standard deviation of 18,300, indicating large performance heterogeneity. We observe similar patterns of performance heterogeneity at the listing level. At a given point in time, more than 37% of the listings have zero sales and the median has 2, whereas the largest listings have accumulated 2362 orders.

3 New Stylized Facts of Online Exporters

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

First, we exploit the unique feature of AliExpress during our study period that allowed us group product listings into different "identical-looking" varieties. Leveraging this unique feature, we first look at how sales performance varies within a single variety group. We focus on popular variety groups with more than 10 listings. As shown in Figure 3, we see that sales are quite concentrated at the top within each group. The group's superstar, defined as the listing with the highest cumulative orders within the group, accounts for about 72.9% of the total sales of the group; the top 25% captures nearly all (93.2%).

Nonetheless, looking at the distribution of the superstar sales across groups, it is also clear that this is not a case of winner-takes-all; some amount of dispersion still remains. Given that we are comparing products with the almost identical design, we are controlling for unobserved consumer horizontal taste. In a friction-less world, one may expect that the listing with the highest quality, relative to price, would win the market.[19] The fact that some dispersion remains indicates that frictions exist in this marketplace. This raises the question of who gets to grow in the presence of these frictions. To delve more into that, 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. Superstar is defined to be the listing with the highest sales in the group and small listings are those with fewer than 5 cumulative orders. Figure 4 plots the distribution of quality difference 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 we sampled. Consistent with this, Figure 5 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.

[19] 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 where we observe buyers' country of origin to examine sales performance within identical variety groups by country (where we restrict sales to a given country and define top sellers for each country separately). Figure B.1 shows the patterns for the US and Russia. To the extent that consumers' tastes are similar within country, the fact that we still observe sizable amount of dispersion at the top suggests that not all of it is explained by heterogeneous preferences.

These observations indicate the difficulties for high quality sellers to gain market share due to search and information frictions. Intuitively, search friction introduces a random component in firm growth due to the consumer sampling process. When firms' intrinsic quality is not perfectly observed, such friction can further slow down the resolution of the information problem and hinder market allocation towards better firms. It is worth noting that the evidence is only suggestive because we have to take into account price differences. [20] To isolate the role of information friction 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 that, subsequent orders arrive much faster.*

Finally, we delve more into the growth dynamics and examine how superstars emerge. Using the transaction-level data over a period of six months from March to August 2018, we explore the dynamics of order arrivals. Figure 6 plots the number of days it takes to receive the n-th order. Panel A shows the order arrival dynamics for the full unbalanced sample of all listings; Panel B restricts 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, conditioning on having 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 B.3 regresses the dummy of receiving an order in a given week on log of past cumulative orders of a product listing, with and without store fixed effect. The results show a demand-driven reinforcement channel where past sales influence future sales: firms with larger past sales, hence higher visibility, have an advantage in overcoming the search friction and generating future orders.

However, a key empirical challenge of 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 as well as 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 to overcome search and information frictions in e-commerce, we conduct an experiment in which we generate exogenous demand and information shocks

[20] Interestingly, we find that superstars do not always charge the lowest price: within an identical variety group, the listing with the highest sales only charges the lowest price for 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 can be partly driven by price.

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

Randomization: Of the 1,265 product listings in the study sample, 790 are small listings with fewer than 5 orders. We randomize the 790 small listings into three groups of different order and review treatments: a control group C without any order and review treatment, T1 which receives 1 order randomly generated by the research team and a star rating, and T2, which, in addition to receiving an order and a star rating, further receives a detailed review on product and shipping quality.

Given that ratings are highly inflated on AliExpress (out of the 6487 reviews we observe over a 6-month window, 85.9% are five stars), for all the treatment groups we leave a five-star rating to the order unless there is any obvious quality defect or shipping problem. This is to mimic the behavior of actual buyers. To generate the contents of the shipping and product reviews, we use the Latent Dirichlet Allocation topic model in natural language processing to analyze past reviews and construct the review messages based on the identified key words. Appendix C.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 for comparisons across otherwise "identical" listings, we leverage the grouping function and stratify the randomization by variety group. For varieties sold by two small sellers (and other big sellers), we assign 1/2 to control and 1/2 to treatment. The latter is randomly split into T1 and T2 with equal probabilities. For varieties sold by more than two small sellers (and other big sellers), we assign 1/3 to each of C, T1, and T2. This randomization procedure is powered to identify the impact of receiving an order, followed by the impact of reviews. In the end, we have 303 listings in C, 259 in T1, and 228 in T2. Table B.5 presents the balance checks and shows that the randomization was balanced across baseline characteristics.[21]

4.2 Treatment Effects of Demand and Information Shocks

We define a dummy variable "Order", which equals to 1 if a listing received the order treatment regardless of the review treatment (i.e., in T1 or T2). Figure B.2 plots the distribution of cumulative net orders (subtracting our own order) 3 months after the intervention. We see a small shift to the right among the treated listings, especially from the 0 to 1 margin. Overall, most listings remain small

[21] We leverage the experiment design to collect information on quality. Measuring product and shipping quality involves making actual purchases. In addition to the 487 small listings in the two treatment groups, we also purchased from the largest listings in the 133 variety groups as well as all the medium-size listings with cumulative orders between 6 and 50 in the same variety groups. This allows us to examine the relationship between quality and size. For service quality, we reached out to all 635 stores in the 133 variety groups and directly communicated with the sellers via the platform. For those with multiple listings included in the 133 groups, we randomly selected one listing for inquiry.

except for a few outliers that managed to grow.[22]

Next, we estimate the following regression to examine the impact of the order and review treatments on weekly cumulative orders after the initial order treatment:

$$\text{WeeklyOrders}_{it} = \alpha + \beta \text{Order}_i + \gamma_1 \text{Review}_i \times \text{PostReview}_t + \lambda_t + \nu_{g(i)} + \epsilon_{it} \tag{1}$$

where the dependent variable is the total number of orders (excluding our own order) for listing i in week t. 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 after the reviews were provided in week 7. The specification leverages the panel structure of our data since the reviews were only given upon receiving the orders.[23] λ_t and $\nu_{g(i)}$ are week and group fixed effects. In addition, all regressions control for baseline sales, both at the store level and the listing level. Standard errors are clustered at the listing level.

The results are shown in Table 3. We include baseline controls of cumulative orders at the store level and product level in the regressions. Results without these baseline controls are very similar and are shown in Table B.6. We see that the order treatment has a small but significantly positive impact on subsequent orders. This demonstrates that indeed a key channel for firms to grow in the online marketplace is by accumulating demand. On the other hand, the impact of the reviews are insignificant, suggesting that the online reputation mechanism may not function effectively in the presence of large search friction. Intuitively, reviews only matter when consumers click and visit a seller's product listing page, which is a rare event for small sellers due to their low visibility.

Table 4 examines 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 quickly afterwards. This is consistent with the fact that only a small number of sellers are able to take advantage of the short-term boost in visibility generated from the order treatment to overcome the initial hurdle of growth.[24] Quantile regression results in Table 5 further show that the impact of the order treatment concentrates in the very top quantiles while the majority of the listings experience no significant impact, consistent with the endline cumulative net orders distribution in Figure B.2. At the same time, heterogeneous treatment effect analyses in Table B.8 show that these rising new stars are not necessarily those with high quality.[25] This echoes the stylized fact in Section 2 that quality does not strongly predict growth in this market. In this market

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

[23] Most of the orders, 801 out 826, arrived within the first 7 weeks. 2 orders arrived later and 23 orders went missing. We left the online reviews in week 7 after the initial order placement when we had received majority of the orders.

[24] In Table B.7, we examine the treatment effect on listings' relative ranking and find that receiving one order indeed leads to a small, short-term improvement in listing visibility.

[25] 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.

environment, search and information frictions combined can make it difficult for high-quality sellers to stand out.[26]

All together, the experimental findings are consistent with the presence of search and information frictions, and show that in such an environment accumulating initial demand acts as a crucial force in shaping firms' subsequent growth. Having said that, the size of the estimated average treatment effect ranges from 0.11 to 0.25, as shown in Table 5. The magnitude is much smaller than 1, 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.[27]

5 Theory

The reduced form evidence highlights an important reinforcement mechanism of online firm growth through consumer search and learning: accumulating sales boosts a seller's visibility and helps attracting 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. The model extends the classical Poly urn model by incorporating consumer choice and seller heterogeneity in quality. Using the model, we characterize the market evolution process and derive predictions on both the long-run and short-run market outcomes, in terms of market concentration and allocation, of increasing search and information frictions.

5.1 Baseline Setup: Consumer Search without Learning

Consider $N \geq 2$ sellers with qualities ranked as $q^1 > \cdots > q^N$. The chance of each seller ending up in consumer's choice set is governed by each seller's "visibility" v^i. Let $v_0^i > 0$ denote the initial visibility of seller i. Consistent with our empirical and experimental findings, we assume that adding up seller i's initial visibility v_0^i and its cumulative sales up to t gives $\{v_t^i\}$. The evolution of v_t^i is stochastic and governed by a model of consumer search. Suppose that consumers arrive sequentially at the platform and each conducts a fixed sample search of $K \geq 2$ sellers upon arrival, where the probability of a seller i being sampled increases with the seller's visibility. Specifically, suppose at the end of period $t \geq 0$ the "total visibility" of sellers $V_t = \sum_{i=1}^{N} v_t^i$. Let $r_t^i = \frac{v_t^i}{V_t}$ be the relative visibility share of seller i, then in period $t+1$ the consumer samples K sellers with replacement, where each sample draws seller i with i.i.d. probability r_t^i.

[26] We also examine the treatment effect on seller effort and business strategy. We find in Table B.9 that receiving a small order does not lead to any noticeable adjustment in pricing, shipping service, listing description (reflecting advertising effort), and introduction of new listings.

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

After obtaining a sample size of K consisting of sellers $i_1, \ldots, i_K$, the consumer chooses to purchase from a particular seller i_k with probability $g_{i_k, i_{-k}}$ satisfying $g \geq 0$ and $\sum_{k=1}^{K} g_{i_k, i_{-k}} = 1$. For most of our analysis (both theoretical and empirical), we will focus on the specific "choice rule" given by the logit model, where

$$g_{i_k, i_{-k}} = \frac{e^{\alpha \cdot q^{i_k}}}{\sum_{\ell=1}^{K} e^{\alpha \cdot q^{i_\ell}}}$$

for some parameter $\alpha > 0$.[28] Some of our theoretical results hold more generally, which we will discuss in due course.

The baseline model focuses on search frictions and assumes that consumers can perfectly observe the products' true qualities from the very beginning. Section 5.4 extends the the baseline setup by incorporating information frictions and a process of consumer learning.

Given the search and choice rules, the probability that the consumer purchases from seller i in period $t+1$ can be written as

$$p_{t+1}^{i} = K \sum_{i_1, \ldots, i_K:\ i_1 = i} \prod_{k=1}^{K} r_t^{i_k} \cdot g_{i_1, i_{-1}}.$$

This is because $\sum_{i_1, \ldots, i_K:\ i_1 = i} \prod_{k=1}^{K} r_t^{i_k} \cdot g_{i_1, i_{-1}}$ is the probability that seller i is chosen as the first seller in the sample. By symmetry, this is also the probability that seller i is chosen as the k-th seller in the sample, for each $1 \leq k \leq K$. Summing across these events yields the total probability seller i is chosen. Hence, with probability p_{t+1}^{i} (which depends on $v_t^1, \ldots, v_t^N$), the visibility of seller i increases by 1, leading to $v_{t+1}^{i} = v_t^{i} + 1$ and $v_{t+1}^{j} = v_t^{j}$ for every $j \neq i$. This fully describes how $\{v_t^i\}$ evolve over time.

To summarize, the primitives of the baseline model are $N, \{q^i\}_{i=1}^{N}, \{v_0^i\}_{i=1}^{N}, K, \alpha$. We point out that if we had set $K = 1$ (i.e. each consumer only samples one seller) or $\alpha = 0$ (i.e. consumer samples multiple sellers but chooses randomly), then this model would return to the classic Polya urn model. However, we will find that the case of $K \geq 2$ and $\alpha > 0$, which we focus on, has drastically different dynamics compared to the classic model. We comment on this distinction at the end of the section.

5.2 Long Run Market Outcomes

We first show that efficient allocation is achieved in the long run despite the existence of consumer search friction, i.e., the highest quality seller will have dominant market position in the long run.

[28] Since we assume that consumers sample with replacement, the sample $i_1, \ldots, i_K$ may involve the same seller multiple times — in such a situation our formulation above implies that each seller is chosen with *weighted* probability, with weights equal to the frequency it appears in the sample. We made this assumption primarily to simplify the exposition of our theory. In our empirical model, we show that all the results are robust to assuming that consumers sample without replacement, provided $N >> K$.

Proposition 1. *Given the primitives of the model, the stochastic processes of* $\{v_t^i\}$ *and* $V_t = \sum_i v_t^i$ *are such that the ratio* $r_t^1 = \frac{v_t^1}{V_t}$ *converges almost surely to 1.*

This result holds for any *general* choice rule g that strictly favors higher quality sellers relative to random choice, including in particular the case of logit choice described in the setup. Full proofs of this and later results are provided in Appendix A. Here, for the sake of illustration, we will describe the main intuitions using a special case where $K = 2$ and consumers choose among the sampled sellers with fixed probabilities that favor the higher quality seller. Specifically, we set $g_{ii} = 1/2$ and $g_{ij} = 3/4$ for every $j > i$. The probability that seller 1 is chosen in period $t + 1$ is then

$$p_{t+1}^1 = (r_t^1)^2 + 2r_t^1(1 - r_t^1) \cdot 3/4 = r_t^1(1.5 - 0.5r_t^1), \tag{2}$$

With probability p_{t+1}^1, we have $r_{t+1}^1 = \frac{v_t^1+1}{V_t+1}$, otherwise $r_{t+1}^1 = \frac{v_t^1}{V_t+1}$. Hence we have[29]

$$\begin{aligned} \mathbb{E}[r_{t+1}^1 \mid r_t^1] &= \frac{v_t^1}{V_t + 1} + p_{t+1}^1 \cdot \frac{1}{V_t + 1} \\ &= r_t^1 + \frac{0.5r_t^1(1 - r_t^1)}{V_0 + t + 1}. \end{aligned} \tag{3}$$

Since the second term on the RHS is positive, it can be shown that the process of $\{r_t^1\}$ represents a bounded sub-martingale that converges almost surely to 1. That is, in the long run there is probability close to one that the consumer chooses seller 1, the highest quality seller.

5.3 Short Run Market Outcomes

Although allocative efficiency is always achieved in the long run, different primitives of this model lead to different dynamics and different speed of convergence. In this section we derive a comparative statics result regarding how market outcomes in the first T periods depend on the number of sellers, which captures the amount of search frictions: for a fixed search sample size K, increasing the number of sellers increases the congestion in consumer search. Note that under sampling with replacement, increasing the number of sellers (in a way that replicates the distribution of qualities) is equivalent, in terms of welfare, to increasing the initial visibility $\{v_0^i\}_{i=1}^N$ by some constant factor.[30] Therefore, to examine the welfare effects of changing the number of sellers, we consider scaling each v_0^i by the same factor $1 + \epsilon$ and ask how this change affects the expected quality the consumer gets in period $1 \sim T$, which reflects market allocation and consumer welfare.

[29] In this special case it is sufficient to condition on r_t^1 because $V_t = V_0 + t$ is deterministic, and v_t^1, p_{t+1}^1 can be written in terms of r_t^1. For general choice rules we need to condition on all the information at the end of period t.

[30] The key observation is that for all the sellers with the same quality level, their *total* visibility at the end of period t determines the probability that any of them will be sampled in period $t + 1$. Because of this, the evolution of their total visibility can be characterized without reference to their individual shares. Therefore it is without loss to "merge" the sellers with the same quality, corresponding to scaling v_0^i.

Proposition 2. *Let $T \geq 2$ be a given positive integer. Suppose total initial visibility satisfies $V_0 \geq \frac{1}{2}KT^2$, then scaling each v_0^i by a factor of $1+\epsilon$ (for any $\epsilon > 0$) would lead to a strict decrease in the expected quality received by the consumer in each of the periods $2 \sim T$, under any logit choice rule. In other words, increasing the number of sellers is welfare-decreasing in early periods.*

Proposition 2 demonstrates that market allocation and consumer welfare strictly worsen in the earlier periods as the number of sellers on the platform increases (equivalent to scaling up initial visibility). To illustrate the main intuition, we return to the special case with $K = 2$ and fixed probability $3/4$ of choosing the higher quality seller out of the sample. A key component that governs allocation and welfare in the short run is the evolution of relative visibility for different quality types, which determines the sample probabilities of different quality types. Focusing on the highest quality seller, seller 1, Equation (3) characterizes the evolution of its relative visibility. Although r_t^1 converges almost surely to 1 in the limit (Proposition 1), the short run dynamics is shaped by V_0. The equation highlights two countervailing forces:

First, larger V_0, or larger N, increases the denominator of the second term on the RHS and hence slows down the convergence speed. Intuitively, the presence of more sellers dampens the positive impact of one additional order on a seller's future probability of being sampled. While this logic applies to all sellers, the negative effect is most relevant for the demand accumulation process of the high-quality sellers, which are favored by the choice rule (and thus benefit from the "positive impact" more often). As a result, it takes longer time for the high quality sellers to accumulate demand and stand out.

At the same time, there is a potential countervailing force as captured by the numerator of the second term on the RHS of Equation (3). The numerator represents the noise of the stochastic process, which can be interpreted as random luck due to the sampling process. Since the function $x(1-x)$ is concave, the smaller the variation in r_t^1 is, the higher $r_t^1(1-r_t^1)$ is in expectation. Increasing V_0 reduces the variation in r_t^1, and thus may act as a potential countervailing force that speeds up the convergence to the long run efficient outcome. It is easy to see that in the earlier periods when t is small, the first channel (i.e., the impact on the denominator) is more likely to dominate, leading to worse allocation and welfare in the short run as V_0 or N increases.[31]

In the arguments above, we have focused on the evolution of the relative visibility, r_t^1, and examined how this is affected by V_0 or N. Although r_t^1 is related to allocation efficiency and welfare, this relation turns out to be indirect. What really matters for the expected quality the consumer gets in period $t+1$ is the probability p_{t+1}^1 of *choosing* the highest quality seller. This choice probability can be written as a function of r_t^1 as in Equation (2). Thus, in our actual proof, we derive the following equation for the

[31] The theoretical result stated in Proposition 2 guarantees that welfare in the first T periods decreases with V_0 when V_0 is large enough to begin with. However, the bound $V_0 \geq \frac{1}{2}KT^2$ is not tight for this comparative statics result to hold. Numerically, we show that allocation and welfare get worse during the entire periods that we need to run the model to match the empirical data.

dynamics of p^1_{t+1} (illustrated in the special case):

$$\mathbb{E}[p^1_{t+2}] - \mathbb{E}[p^1_{t+1}] = \frac{V_t}{(V_t+1)^2} \cdot \mathbb{E}[r^1_t(1-r^1_t)(0.75-0.5r^1_t)]. \tag{4}$$

This identity tells us that the choice probability of seller 1 increases over time. Moreover, the increment on the RHS involves a factor $\frac{V_t}{(V_t+1)^2}$ which depends on V_0 since $V_t = V_0 + t$, and an expectation of a function of r^1_t. Therefore, the same two forces discussed above carry through: allocation and welfare become worse in the short run as N increases and search frictions intensify.

5.4 Incorporating Learning

We now consider an extension of the baseline model where the true qualities q^i are not known ex-ante, but learned over time through past purchases and reviews. Suppose that consumers hold a common prior that q^i are independently and normally distributed, with prior mean $\widehat{q_0}$ and prior variance σ_0^2 (so that $\tau_0 \equiv \frac{1}{\sigma_0^2} > 0$ is the prior precision about each q^i).[32] Consumers who have purchased from seller i leave noisy reviews which serve as signals about q^i.

Specifically, suppose at the end of period t the cumulative sales of each seller i is s^i_t, i.e. $v^i_t = v^i_0 + s^i_t$, and the consumers' common posterior mean of q^i is $\widehat{q^i_t}$. Then in period $t+1$ the following occurs:

1. A consumer arrives at the platform and samples K sellers, $i_1, \ldots, i_K$, with replacement, according to probabilities $\{r^i_t = \frac{v^i_t}{V_t}\}$;

2. The consumer chooses to purchase from seller i_k with probability $g(\widehat{q^{i_k}_t}, \widehat{q^{i_{-k}}_t}) = \frac{e^{\alpha \cdot \widehat{q^{i_k}_t}}}{\sum_{m=1}^K e^{\alpha \cdot \widehat{q^{i_m}_t}}}$, which is the logit choice probability computed from expected qualities. Then $s^{i_k}_{t+1}$ and $v^{i_k}_{t+1}$ both increase by 1 from their period t values. All other sellers' sales and visibility are unchanged;

3. The consumer who purchases from seller i_k in period $t+1$ produces a publicly observed review about its quality. This review/signal takes the form $z_{t+1} = q^{i_k} + \epsilon_{t+1}$ where ϵ_{t+1} is an independent $\mathcal{N}(0, \sigma^2)$ random variable with precision $\tau \equiv \frac{1}{\sigma^2}$ about q^{i_k};

4. Let $\bar{z}^{i_k}_{t+1}$ be the average of the past $s^{i_k}_{t+1}$ reviews about q^{i_k}, up to and including period $t+1$. Then by familiar results about Bayesian updating of normal signals, the posterior expected value of q^{i_k} at the end of period $t+1$ is given by

$$\widehat{q^{i_k}_{t+1}} = \frac{\tau_0 \cdot \widehat{q_0} + s^{i_k}_{t+1}\tau \cdot \bar{z}^{i_k}_{t+1}}{\tau_0 + s^{i_k}_{t+1}\tau}, \tag{5}$$

[32] For ease of notation we restrict to a symmetric prior such that all sellers have the same prior mean and prior variance. Our theoretical results do generalize.

where we take the weighted average of the prior mean $\widehat{q_0}$ and the past average review $\bar{z}_{t+1}^{i_k}$ with weights given by their respective precision. Written this way, we see that the posterior mean $\widehat{q_{t+1}^{i_k}}$ is a deterministic function of sales amount, which is also the number of reviews about seller i_k, and the average review.

While our empirical analysis implements the belief updating formula (5), it is worthwhile to point out the following equivalent formulation, which relates $\widehat{q_{t+1}^{i_k}}$ to $\widehat{q_t^{i_k}}$ in a recursive manner:

$$\widehat{q_{t+1}^{i_k}} = \frac{\left(s_t^{i_k}\tau + \tau_0\right) \cdot \widehat{q_t^{i_k}} + \tau \cdot z_{t+1}}{\left(s_t^{i_k}\tau + \tau_0\right) + \tau}. \tag{6}$$

This will be convenient for our theoretical analysis. In particular, it shows that the $2N$ random variables $\{v_t^i, \widehat{q_t^i}\}$ follow a Markov process such that in period $t+1$, a particular seller i_k is chosen with probability given by the sampling process (based on $\{v_t^i\}$) and logit choice rule (based on $\{\widehat{q_t^i}\}$). Once this seller is chosen, its visibility becomes $v_{t+1}^{i_k} = v_t^{i_k} + 1$ and its expected quality $\widehat{q_{t+1}^{i_k}}$ evolves according to (6) where $z_{t+1} \sim q^{i_k} + \mathcal{N}(0, \sigma^2)$. For all other sellers $i \neq i_k$, their visibility v_{t+1}^i and expected qualities $\widehat{q_{t+1}^i}$ are unchanged from period t. These summarize the dynamics of the model with learning.

Interestingly, despite the additional information frictions, we find that allocative efficiency is again achieved in the long run under the logit choice rule. Formally,

Proposition 3. *Consider the model with learning, and suppose the true qualities are ranked as $q^1 > \cdots > q^N$. Then under any logit choice rule, the ratio $r_t^1 = \frac{v_t^1}{V_t}$ converges almost surely to 1.*

There is a subtle difference between this result and the previous Proposition 1. Proposition 1 shows that in the baseline setting without learning, long run efficiency is achieved by any general choice rule that favors higher quality sellers in the sample. In contrast, our proof of Proposition 3 demonstrates that the same result does not always hold with learning. Indeed, if the choice rule "overly favors" higher expected quality (such as the rule that selects the highest expected quality seller in the sample with probability 1), then every seller can get lucky in early periods, receive favorable reviews and maintain this initial advantage indefinitely, due to its *expected* quality being persistently higher than the remaining sellers that never get chosen and learned about. This kind of inefficiency can occur with positive probability if consumers do not "experiment" sufficiently with sellers whose current expected qualities are not the highest, much like in a multi-armed bandit problem.

Despite this possibility, Proposition 3 shows that long run efficiency is restored so long as in every sample, each seller is chosen with strictly positive probability regardless of its expected quality. This is in particular satisfied by the logit choice rule. Under such a condition, we show that every seller is chosen infinitely often along almost every history. With infinite signals for each seller, consumers will

be able to identify the highest quality seller in the long run. So eventually the dynamics of the current model with learning will resemble the baseline setting where true qualities are known. This leads to the best seller getting predominant market share.

Proposition 3 establishes the long run efficiency result under both search and information frictions. In Appendix A, we derive an analogous result to Proposition 2, focusing on market outcomes in the short run. We show that market allocation and welfare get worse in the early periods as the number of sellers increases. Furthermore, there is a potential interaction effect between search and information frictions: the noisier the review signal is (larger information frictions), the greater is the negative effect of increasing N (increasing search frictions).

Overall, the theoretical results in this section verify that the reinforcement mechanism that we documented in Section 3 and 4 indeed enables high quality sellers to overcome their initial hurdle of visibility and eventually take over the market in the long run. The possibility of information friction does not fundamentally change this force, as long as there is sufficient experimentation.

We emphasize that *both* sales-based sampling and consumer choice are important for the reinforcement mechanism, and thus for long run efficiency to be guaranteed. If sampling were completely random (instead of being proportional to visibility), then even in the absence of information friction the highest quality seller would not eventually occupy the entire market, due to the fact that lower quality sellers are sampled and chosen too often.[33]

On the other hand, if a consumer choses randomly out of the sample (instead of logit choice based on qualities), then we would return to the classic Polya urn model where quality has no influence on the dynamics. Thus, assuming that the sellers have equal visibility ex-ante, it is equally likely for the lowest quality seller to eventually dominate the market, as it is for the highest quality seller.[34] Inefficiency again persists in the long run.

Apart from predicting convergence to efficiency, Section 5.3 further shows that the speed of convergence is decreasing in the number of sellers N, at least in the short run. This result is particularly important since on-boarding more sellers (i.e. increasing N) has been a central policy focus of many government to facilitate SME growth. Our result formalizes a key countervailing force where more sellers congest the search process and slow down the rise of high-quality sellers. Of course, the quantitative magnitude of this mechanism and its impact on market transition remain an empirical question.

[33]To show this formally, we note that with random sampling and known qualities, the probability that each seller i is chosen in period t is a constant $p_i > 0$ that does not depend on t. Thus by the Law of Large Numbers, seller i's market share r_t^i almost surely converges to p_i.

[34]In the classic Polya urn model, it is known that almost surely each r_t^i converges as $t \to \infty$. However, the limit distribution of these relative visibility follows the symmetric and full-support Dirichlet distribution, rather than the point mass at $r_t^1 = 1$ as shown in Proposition 1.

6 An Empirical Model of the Online Marketplace

We build on the theoretical model in Section 5 to incorporate several additional empirical features of the online market. As in the theoretical model, our setting focuses on demand-side mechanisms to highlight the role of the search friction due to limited sample search and the information friction due to imperfect observability of underlying quality. On the supply side, though, we further incorporate seller-side heterogeneity in both quality and cost and model sellers' pricing decisions. We then structurally estimate the model to fit the key data moments and perform counterfactual analysis examining the impact of search and information frictions on firm growth, market allocation, and consumer welfare.

6.1 Demand

Search: The model of consumer search closely follows the theoretical setup in Section 5.1 where consumers randomly sample K sellers with replacement upon their arrival.[35] We assume that the search sample size K is randomly drawn from the 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 (which we now assume to be equal across sellers) and cumulative sales s_t^i, reflecting the fact that products sold by larger sellers are often positioned more saliently on the platform.[36] Consider an ordered sample $\phi = (i_1, i_2, \ldots, i_K)$ of size K. The probability that this sample is considered by the consumer is given by:

$$P(\phi|\mathbf{v_t}) = \prod_{k=1}^{K} r_t^{i_k}, \tag{7}$$

where we recall $r_t^i = \frac{v_t^i}{\sum_j v_t^j}$ is seller i's relative visibility.

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 follow the description in Section 5.4. In particular, we assume that prior beliefs follow a standard normal distribution $q^i \sim N(\mu_0, \sigma_0^2)$ with $\mu_0 = 0, \sigma_0^2 = 1$. Empirically, we standardize our quality measures to be consistent with this assumption. If the buyer chooses to purchase from seller i, she observes a noisy signal z^i of

[35] 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 is likely to matter for large sellers but relatively less so for small sellers. Table B.3 shows that the demand accumulation mechanism is salient even with store fixed effect, i.e., at the listing level within store.

[36] In practice, the platform allows consumers to rank listings by Best Match, Past Orders, Newest, and Prices. The default (by Best Match) uses AliExpress's recommender system which is trained by fitting a machine learning model using historical data on listings, sellers and buyers. Listings with more cumulative sales get more visibility in the recommender systems and when consumers rank the listings by Past Orders. Without individual consumer level data on search, display and clicks history, we estimate a reduced form function that captures how cumulative sales boosts a listing's likelihood of entering into a consumer's consideration set.

the true quality q^i and leaves a truthful review. z^i follows a normal distribution centered around the true quality with standard deviation σ, which governs the noisiness of the review signals.

The consumers' common posterior expectation of each seller i's quality, denoted by $\widehat{q^i}$, follows the Bayesian updating rule as described in Equation (5). 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 $\bar{z}_t^i \equiv \frac{\sum_{j=1}^{s_t^i} z_t^i}{s_t^i}$ (seller i's rating, or average past review) and s_t^i (cumulative sales).

Purchase and Review: We extend the baseline logit demand framework described in Section 5.1 to incorporate prices and an outside option of non-purchase 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 consumer's idiosyncratic preference shock with I.I.D type I extreme value distribution. β and γ are the constant and the coefficient on price.

6.2 Supply

On the supply side, we extend the baseline setup in Section 5.1 to incorporate seller heterogeneity in cost, in addition to quality, as well as a pricing strategy that approximates the observed data. Specifically, each seller is characterized with exogenous cost c^i that is potentially correlated with fundamental quality q^i. Each seller's pair of (c^i, q^i) are drawn from a random distribution upon the firm's entry into the online platform. We denote the correlation between c_i and q_i as ρ. 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 to use product price as a signal for unobserved quality.

Price Adjustment: Since the consumer's search depends on each seller's previous cumulative orders, one might naturally think that sellers would have incentive to compete for future demand by dynamic pricing. However, in our sample, we observe very infrequent price adjustment.[37] More importantly, we do not observe systematic pattern of price increase 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

[37] In our study sample with 1265 listings, there were only 142 price adjustments during the 13 week post-treatment periods. We also find little empirical evidence of the life-cycle price dynamics for sellers, in particular, for those with higher measured quality. The lack of price movement is consistent with that documented in (Fitzgerald, Haller, and Yedid-Levi, 2020).

i, which is the probability that it appears in the sample and gets chosen by the consumer. Thus D_i depends on the rich set of public information $\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}$, which are the prices, ratings, and cumulative sales of all sellers at the time of price adjustment:

$$D_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = K \sum_{i_2,\ldots,i_K} \prod_{k=2}^{K} r_t^{i_k} \cdot \frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]}, \tag{8}$$

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

Given the demand function D_i, seller i solves the following problem:

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

where the first order condition reads

$$p_i - c_i = -\frac{D_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s})}{\partial D_i / \partial p_i(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s})} \tag{9}$$

Given the additive structure of D_i based on the realized samples ϕ, we can easily define the key piece of demand elasticity with

$$\begin{aligned}\frac{\partial D_i}{\partial p_i}(\mathbf{p}, \bar{\mathbf{z}}, \mathbf{s}) = -K\gamma \sum_{i_2,\ldots,i_K} \prod_{k=2}^{K} r_t^{i_k} &\left(\frac{exp[(\widehat{q^i} - \gamma p^i)]}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]} \right) \\ &\times \left(1 - \frac{exp[(\widehat{q^i} - \gamma p^i)] \cdot (1 + |\{2 \leq k \leq K : \ i_k = i\}|)}{1 + exp[(\widehat{q^i} - \gamma p^i)] + \sum_{k=2}^{K} exp[(\widehat{q^{i_k}} - \gamma p^{i_k})]} \right)\end{aligned}$$

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 sample, conditioning on being in the consumer's search sample ϕ. 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 gets 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 an average entrant.

6.3 Model Estimation

6.3.1 Parametrization and Identification

Our model has six structural parameters: $\{v_0, \sigma, K, \beta, \gamma, \rho\}$. The consumer demand depends on. the initial visibility parameter v_0, the review signal noise σ, the search sample size K, 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, it provides relatively little information of the overtime variation in their cost. So 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.[38] Another key parameter of the 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).[39] Therefore, in our baseline estimate, we assume that K follows a positive Poisson distribution with mean 2. Section 7.3 performs robustness checks with different parameter values of γ and K.

The rest of the structural parameters $\{v_0, \sigma, \rho\}$ 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 order on cumulative orders
3. The regression coefficient of log price and the measured quality
4. The conditional distribution of cumulative orders for each measured quality segment

We simulate our model from the start until the sellers' average cumulative orders reach the level in our data (44 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. Intuitively, a small initial visibility v_0 increases the relative importance of early orders in a seller's life cycle. The amplification effect of cumulative orders is more pronounced in this case, and it increases the skewness of market distribution. The dependence of a seller's new order on cumulative orders plays a similar role in disciplining v_0. Conditioning on v_0, the correlation between a seller's cumulative orders and measured quality identifies the review signal noise σ. If the review was very precise, then higher quality sellers would grow their orders rapidly once they end up in consumer's search sample. In contrast, a larger σ results in a flattened relationship between quality and the cumulative orders. Finally, a competing force that could result in a low correlation between cumulative order and quality is the cost-quality

[38] Aliexpress' market share in the number of cross-border buyers is 16-20% (source: https://techjury.net/blog/aliexpress-market-share/#gref). Given this, we set the share of the outside option to 80%.

[39] In the online marketplaces, studies consistently find that while some consumers search intensely, the vast majority search very little (e.g., Hong and Shum (2006); Moraga-González and Wildenbeest (2008); Wildenbeest (2011)).

dependence ρ. Hence we also require our simulated data to be consistent with the observed correlation between price and quality.

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

6.3.2 Estimation Results

Table 6 presents the parameter estimates with standard errors. The parameter v_0 that governs the initial visibility is estimated to be 5.18. To interpret the magnitude, consider the initial stage of a market where one seller makes his/her first sales while all other sellers have made zero sales; the visibility for the former increases by 19% relative to the latter. The review noise σ is estimated to be 2.52. This implies that the standard deviation of the posterior belief is reduced only by 7% 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 about sellers' quality and that the uncertainty about each seller's quality is resolved very slowly, i.e. only after a substantial amount of orders. This indicates that the reputation mechanism takes time to play a role even if a seller emerged in a consumer's choice set and successfully made a sale. As a result, the search friction, interacting with the information friction, constitutes the major hurdle of seller's initial growth. Finally, the estimate for ρ is 0.53. Given the empirical marginal distribution of cost and the standard normal quality distribution, this translates into a coefficient of correlation between quality and cost of 0.52.

Table 7 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 all very well.

7 Counterfactual Analyses

We conduct counterfactual exercises to examine the role of information and search frictions on firm growth and consumer welfare. We evaluate potential policy interventions through the lens of our estimated structural model. The results are reported in Table 8.

7.1 The Role of Search Friction: 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 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 gain of doing that in the presence of large information frictions estimated at baseline (i.e., $\sigma = 2.52$). To understand this

result, Figure 7 shows that the expected sample quality, weighted by the choice probabilities, improves at a faster rate when N is reduced by a half. The results are consistent with Proposition 2 and the theoretical discussion in Section 5.4: 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 Figure 8 and gain larger market share: the cumulative market share of the top 1/3 quality sellers increases by 11.3 percentage point when the number of sellers is reduced from 1250 to 625; the sales-weighted quality is 60% higher. As a result, the expected consumer surplus increases by 25.4%, taking into account the fact that higher quality sellers charge higher prices.

Next, we examine the potential interaction effect between search and information frictions. As discussed in Section 5.4, the efficiency and welfare implications of reducing search frictions may depend on the amount of information frictions. To test this, Panel A.2 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 friction 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 (larger information frictions), the larger is the negative effect of increasing N (increasing search frictions). [40]

7.2 The Role of Information Friction: Reducing the Review Signal Noise

We then go on to examine the role of information frictions. Since a seller's true quality is imperfectly observed by consumers, review from past purchases serves as a crucial source of information for subsequent consumers. Our estimate implies a quite noisy review signal, with a standard deviation of 2.52. In Panel B.1 of Table 8, we compare our baseline with a case where we reduce σ to zero. In other words, we investigate a case that a seller-listing's true quality is immediately revealed when it accumulates its first order. We find that in this case, the cumulative market share significantly shifts towards the higher quality sellers. As a consequence, we also find that average consumer welfare improved by 13.8%.

We can further decompose this welfare gain into a static gain and a dynamic gain. At each given point in time, holding fixed the search set, smaller σ allows consumers make more informative purchase decisions, that is, they are more likely to purchase from high-quality sellers when reviews are more precise signals of quality. In addition to the static gain, the sampling probabilities of high-quality sellers get 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 friction is much more muted in this case. Consumer surplus actually decreases

[40] The counterfactual exercises in Panel A speak to policies targeting at the creation of new marketplaces. For existing platforms, which already host a large number of sellers and listings, one can imagine an analogous exercise by screening out inactive seller-listings.

because of the countervailing price effect. The difference in Panels B.1 and B.2 further highlights the interactions between information and search frictions.

7.3 Robustness Checks

Table 9 examines the robustness of the counterfactual analyses under different parameter values. Panel A reproduces the baseline and counterfactual outcomes of reducing N using the parameter estimates in Table 6. Panel B shows the outcomes under parameter values re-estimated with the search sample size K drawn from the positive Poisson distribution with mean 5. Holding fixed the number of sellers, we see that larger K leads to better allocation as it reduces search friction by allowing more sellers to enter into the search sample of consumers, whose choice rule favors higher-quality sellers. At the same time, 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.

Next, we perform robustness checks in which we allow multiple sellers to belong to the same variety groups (as motivated by Panel B of Figure 1). Specifically, in Panel C, we consider a case where each variety group contains 10 sellers with identical horizontal attributes.[41] Thus, the realization of taste shocks ε_i would be identical for sellers in the same variety group. This implies that in the event where multiple sellers from the same group get sampled, the consumer always favors the seller with higher expected quality within the group. This force would lead to better allocation as compared to the baseline. At the same time, there exists a countervailing negative force as the number of distinct varieties get reduced (by 10 times). Comparing the outcomes in Panel A and Panel C, holding fixed the number of sellers, we see that the market allocation indeed improves as we reduce the number of varieties; consumer surplus is slightly lower due to the countervailing loss from the variety effect. That said, the loss from the variety effect is modest in our setting, mainly due to the relatively small search sample size ($K << N$) that constrains any variety effect in the first place. Reassuringly, Panel C shows that reducing the number of sellers once again improves allocation and consumer welfare in the case with fewer varieties.

Last but not least, in Panels D and E, we perform two additional robustness checks under different parameter values of γ that correspond to different price elasticities (4 and 10, respectively). The key comparative statics of reducing the number of sellers remain robust.

8 Conclusion

In this paper, we study exporter dynamics on global e-commerce platforms. Leveraging comprehensive data about the online businesses from AliExpress and combining that with unique objective measures

[41] Our baseline estimates correspond to a case where each seller represents a distinct variety.

of quality covering multiple product and service dimensions, we document sizable variation 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 the potential source of market misallocation in e-commerce.

Our findings speak to effective policies on 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 to these platforms may not be sufficient to generate sustained growth due to the large demand-side frictions. In fact, doing so can exacerbate the search and information problems, resulting in market misallocation. Policies should be designed to help firms, especially new businesses, to overcome the additional demand-side frictions. In the context of e-commerce, regulating entry, creating a premium market segment, and directing demand to promising newcomers could help to facilitate growth and improve the overall market efficiency.

References

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

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

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

Bai, Jie. 2016. "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation." Tech. rep.

Banerjee, Abhijit Vinayak. 2013. "Microcredit under the Microscope: What Have We Learned in the Past Two Decades, and What Do We Need to Know?" Annu. Rev. Econ. 5 (1):487–519.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India*." The Quarterly Journal of Economics 128 (1).

Broda, Christian and David E Weinstein. 2006. "Globalization and the Gains from Variety." The Quarterly journal of economics 121 (2):541–585.

Ching, Andrew T, Tülin Erdem, and Michael P Keane. 2013. "Learning models: An assessment of progress, challenges, and new developments." Marketing Science 32 (6):913–938.

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

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

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." The Quarterly Journal of Economics 123 (4):1329–1372.

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

Eaton, Jonathan, David Jinkins, James Tybout, and Daniel Yi Xu. 2016. "Two-sided Search in International Markets." In 2016 Annual Meeting of the Society for Economic Dynamics.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi. 2020. "How Exporters Grow." Tech. rep.

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

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

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

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

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

Honka, Elisabeth, Ali Hortacsu, and Matthijs Wildenbeest. 2019. "Empirical search and consideration sets." In Handbook of the Economics of Marketing, vol. 1. Elsevier, 193–257.

Hortaçsu, Ali, F Martínez-Jerez, and Jason Douglas. 2009. "The geography of trade in online transactions: Evidence from eBay and mercadolibre." American Economic Journal: Microeconomics 1 (1):53–74.

Hortaçsu, Ali and Chad Syverson. 2004. "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 119 (2):403–456.

Khandelwal, Amit. 2010. "The long and short (of) quality ladders." The Review of Economic Studies 77 (4):1450–1476.

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

Kugler, Maurice and Eric Verhoogen. 2012. "Prices, Plant Size, and Product Quality." The Review of Economic Studies 79 (1):307–339.

Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina. 2016. "There goes gravity: eBay and the death of distance." The Economic Journal 126 (591):406–441.

Lipsman, A. 2019. "Global Ecommerce 2019: Ecommerce Continues Strong Gains Amid Global Economic Uncertainty. eMarketer." Saatavissa (viitattu 8.4. 2020): https://www. emarketer. com/content/global-ecommerce-2019 .

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

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

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

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

Pitney Bowes, Inc. 2016. "2016 Pitney Bowes Global Online Shopping Study." http://www.pitneybowes.com/us/global-ecommerce/case-studies/the-growing-importance-of-international-ecommerce.html. Accessed: 2020-07-20.

Piveteau, Paul. 2016. "An empirical dynamic model of trade with consumer accumulation." Tech. rep.

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

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

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

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

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

Stigler, George J. 1961. "The economics of information." Journal of political economy 69 (3):213–225.

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

UNCTAD, Torbjörn, Fredriksson. 2016. "Unlocking the Potential of E-Commerce for Developing Countries." Tech. rep.

Verhoogen, Eric. 2020. "Firm-Level Upgrading in Developing Countries." Tech. rep.

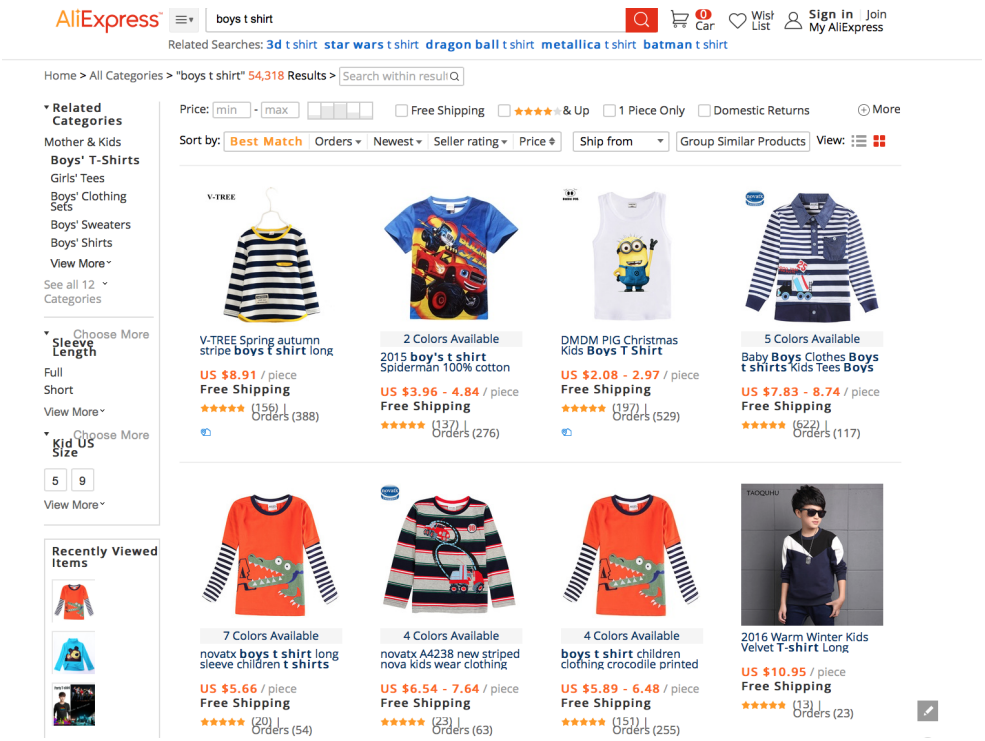
Verhoogen, Eric A. 2008. "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector." The Quarterly Journal of Economics 123 (2):489–530.

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

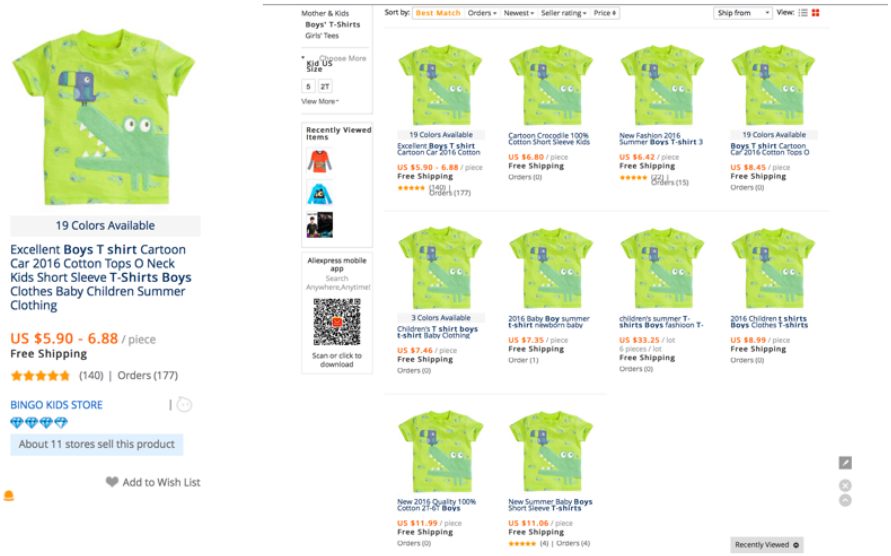
WTO, World Trade Organization. 2020. "World trade statistical review 2020." .

Figure 1: AliExpress: Search Results with and without Grouping

Panel A. Search Results without the Grouping Function



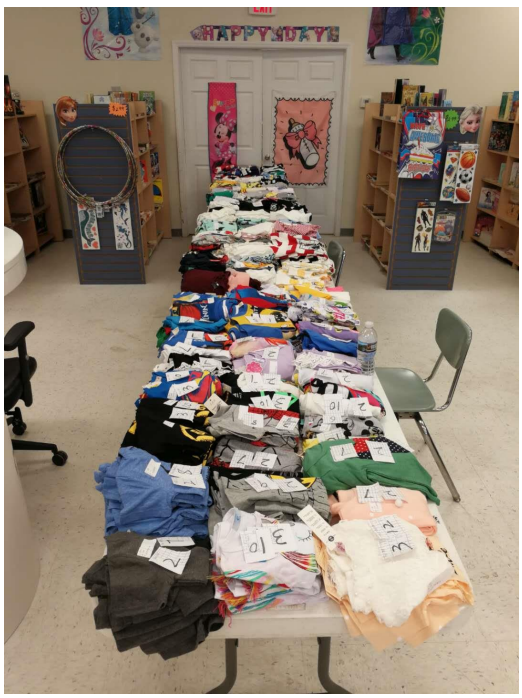
Panel B. Search Results with the Grouping Function



Note: This figure presents examples of search results on AliExpress. Panel A displays the search results using "children's t-shirts" as keywords, without applying the grouping function provided by the website. Panel B displays the same search results while applying the grouping function.

Figure 2: Quality Assessment

Panel A. Quality Assessment



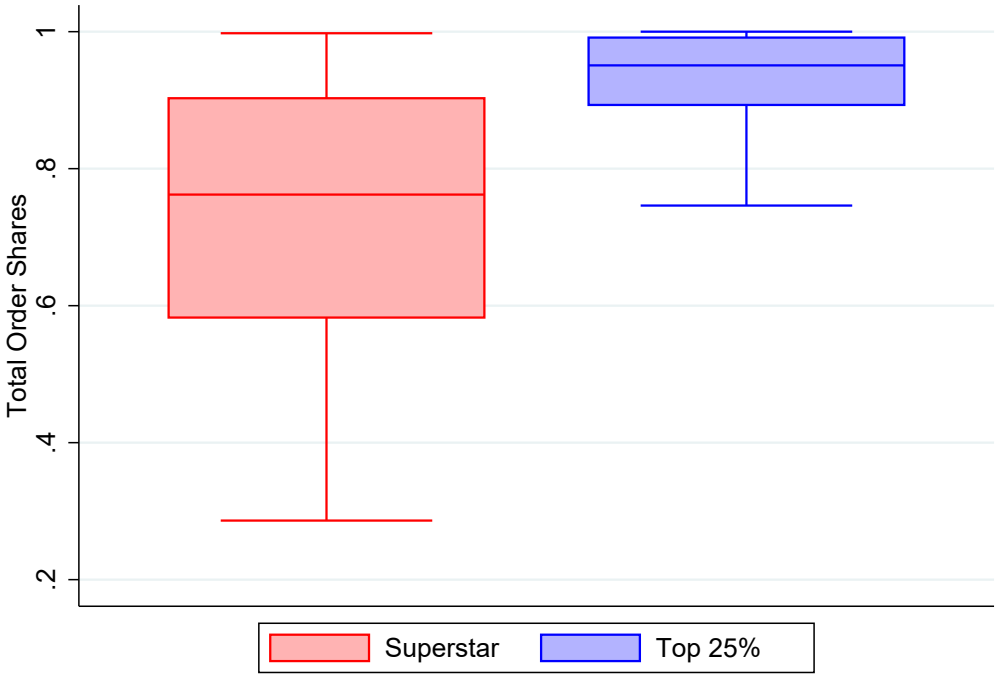
Quality Metrics:

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

Panel B. Variation in Quality

Note: Panel A displays the purchased t-shirts, sorted by groups, from our experiment; the quality assessment agent located in Durham, North Carolina; and the quality metrics used in the assessment process. Panel B shows examples of the t-shirts that receive low scores in specific quality metrics.

Figure 3: Sales Performance Within Identical Variety

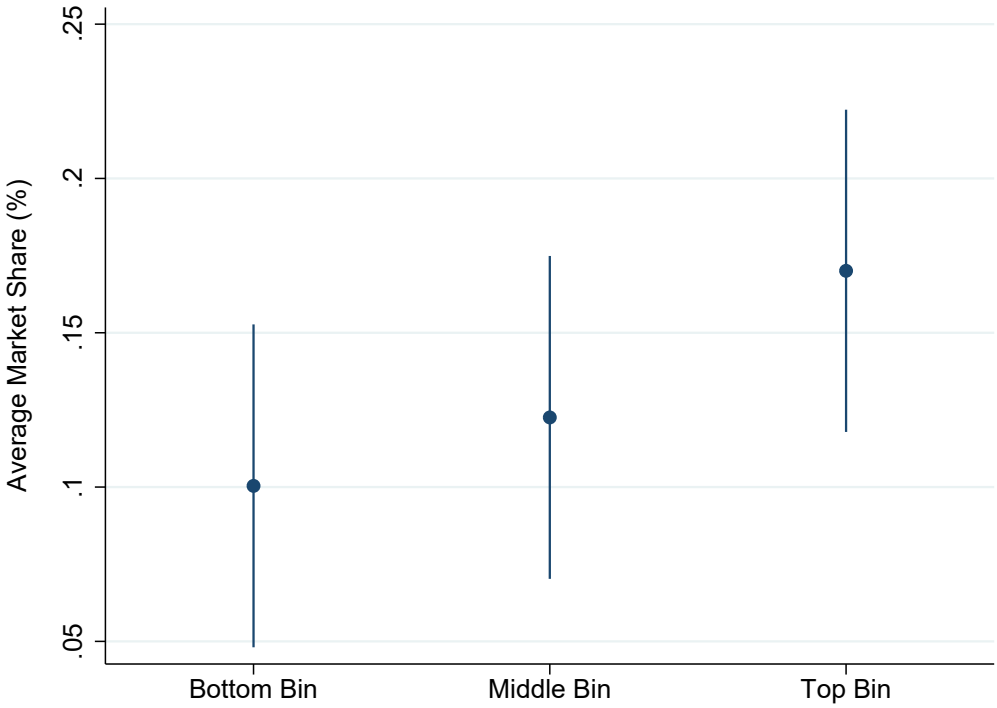


Note: This figure plots the distribution of the total share of cumulative orders for top listings across groups. "Superstar" indicates listings that have the highest cumulative orders within its group variety. "Top 25%" indicates listings that have the top 25% cumulative orders within its group variety. We limit to groups with at least 4 listings so that "Top 25%" is well defined.

Figure 4: Quality Comparison Between Group Superstar and Small Listings

Note: This figure plots the distribution of the quality differences between group superstars and group small listings. Quality is measured by the Overall Quality Index (see Section 2.2 for details on the construction of quality indices). Group superstar is defined to be the listing with the largest number of cumulative orders in each group. Small listings is defined to be the listings with fewer than 5 cumulative orders.

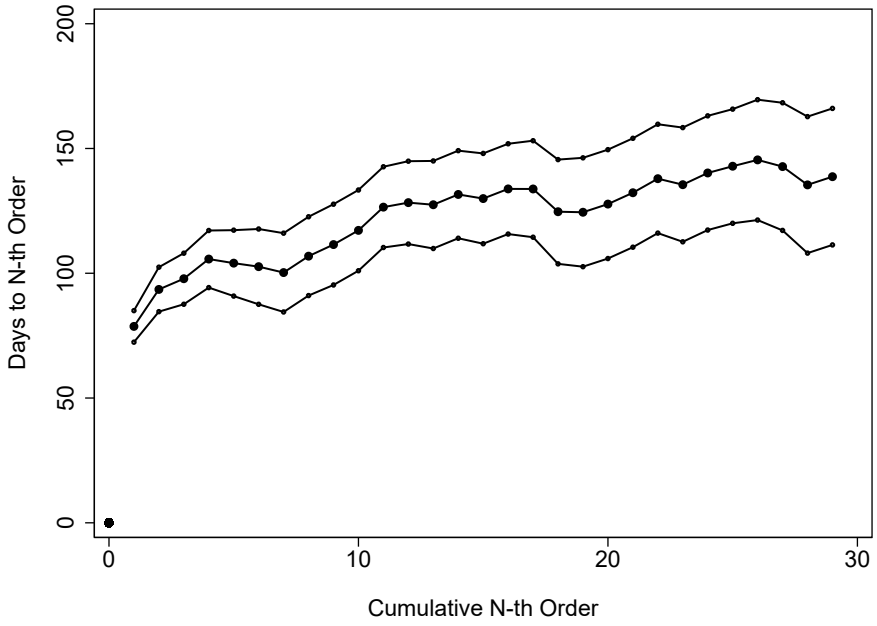
Figure 5: Average Market Share over Quality Bins



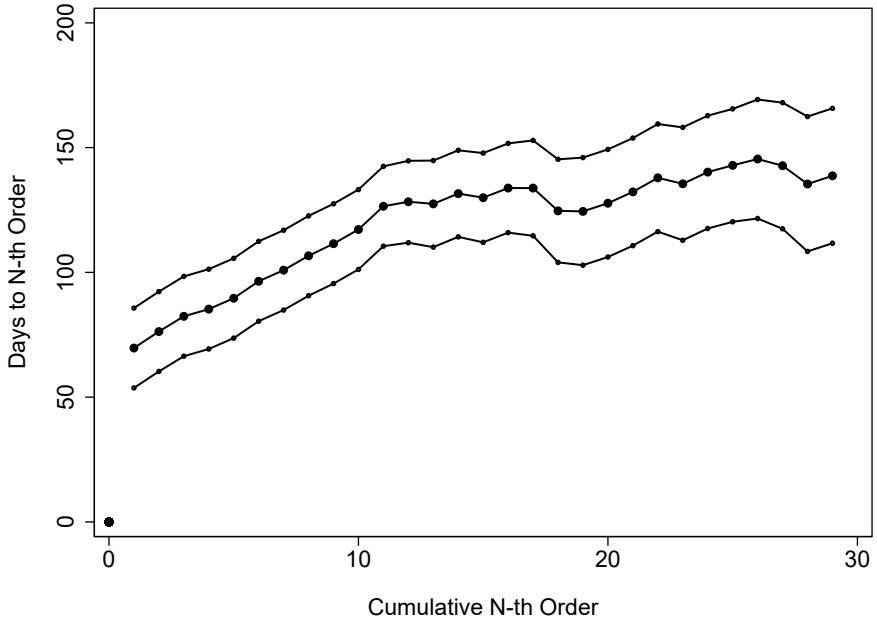
Note: This figure plots the regression coefficients and their 95% confidence intervals from regressing the listings' shares based on cumulative orders on the quality bins they belong to.

Figure 6: Dynamics of Order Arrival

Panel A. Unbalanced Panel of All Listings

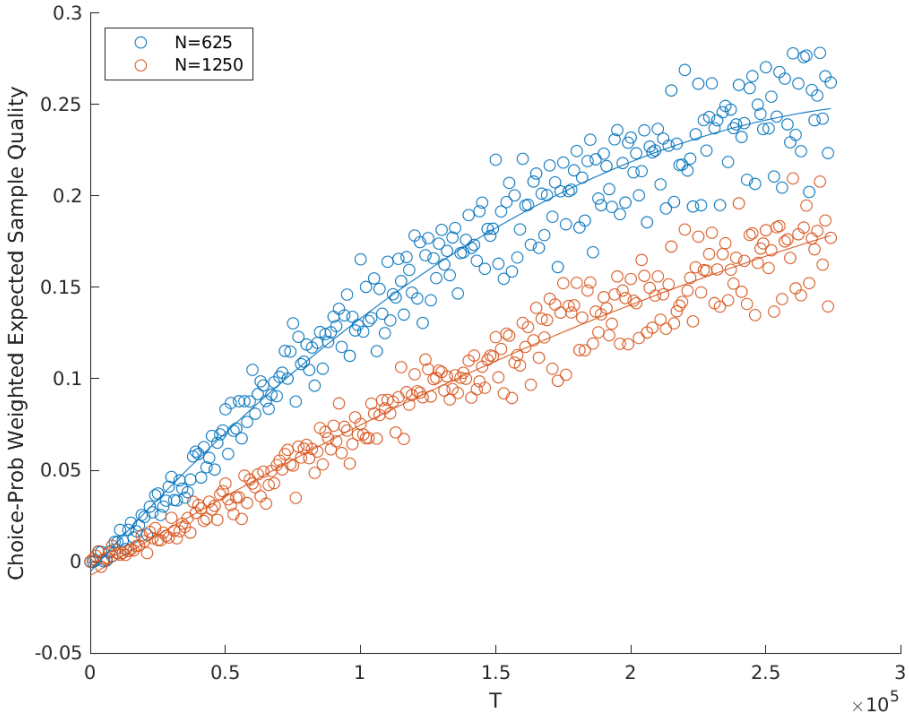


Panel B. Listings with More than 10 Cumulative Sales



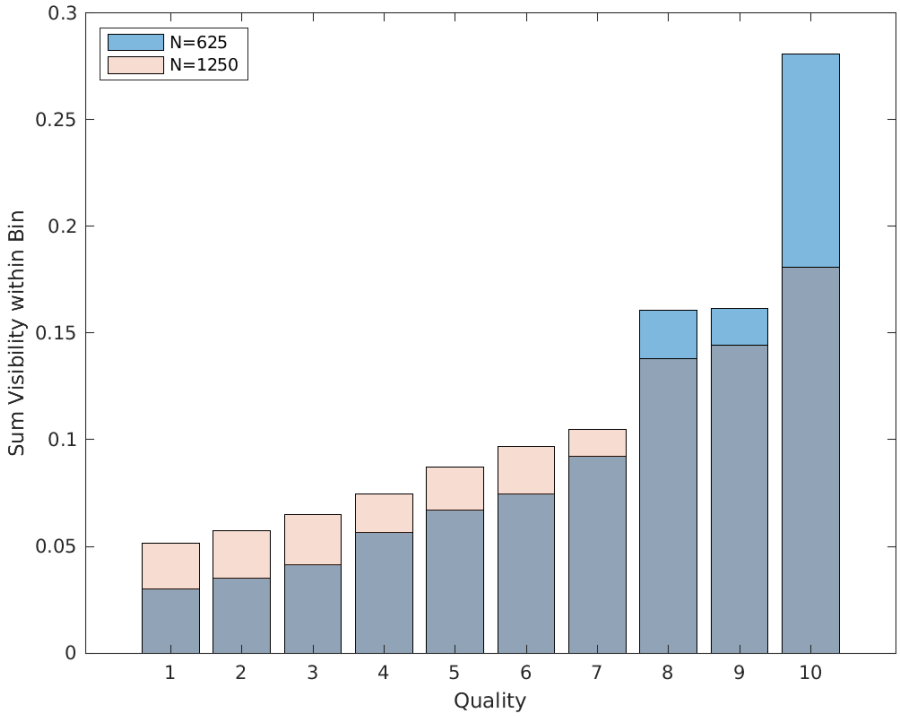
Note: This figure describes the order arrival dynamics. The x-axis indicates the n-th order and the y-axis shows the number of days till receiving the n-th order. The bold line in the middle plots the average and the other two dotted lines plot the 95% confidence interval. We use the six-month transaction history data described in Section 2.2. Panel A include the full unbalanced panel of all listings appeared in the transaction data. Panel B restrict the sample to listings that accumulated more than 10 orders during the six-month period.

Figure 7: Counterfactual Expected Choice Quality Overtime



Note: This figure plots the choice probability weighted expected quality within consumers' search sample overtime. Data are simulated based on our baseline estimates of parameters.

Figure 8: Counterfactual Visibility Distribution Over Quality



Note: This figure plots the distribution of visibility over quality bins in the last model period. Data are simulated based on our baseline estimates of parameters.

Table 1: Summary Statistics of Quality Measures

	Observations	Mean	Std Dev	Median	Intra-Group Corr	Within-Group SD	Between-Group SD
				Panel A: Product Quality			
NoObviousQualityDefect	796	.93	.26	1	.08	.25	.07
Durability	791	2.63	.79	3	.91	.24	.75
MaterialSoftness	791	3.21	.72	3	.94	.17	.71
WrinkleTest	791	3.08	.48	3	.91	.15	.46
SeamStraight	791	4.22	.47	4	.2	.42	.21
OutsideString	791	2.8	1.56	3	.4	1.22	.99
InsideString	791	.78	1.18	0	.46	.87	.8
PatternSmoothness	773	3.43	1.53	4	.87	.55	1.44
Trendiness	791	3.13	1.35	3	.91	.4	1.3
				Panel B: Service and Shipping Quality			
BuyShipTimeLag	823	3.66	3.24	3	.32	2.67	1.85
ShipDeliveryTimeLag	802	12.92	4.15	12	.19	3.73	1.83
LostPackage	820	.02	.14	0	.09	.13	.04
PackageDamage	795	0	.05	0	0	.05	.
ReplyWithinTwoDays	1258	.69	.46	1	.08	.44	.13
				Panel C: Quality Indices			
ProductQualityIndex	769	0	.41	-.02	.7	.22	.34
ShippingQualityIndex	793	.04	.43	.12	.09	.41	.13
ServiceQualityIndex	1258	0	1	.67	.08	.96	.28
OverallQualityIndex	763	.01	.29	.01	.54	.2	.22

Notes: This table reports the summary statistics of the various quality measures. Sections 2.2 and C.2 provide details on the measurement process and each of the quality metrics. Panel C reports the aggregate quality indices constructed by standardizing scores of individual quality metrics and taking their average within each quality category. The number of observations changes slightly across quality measures because some items were lost in the shipping process and in the quality assessment process.

Table 2: Summary Statistics of the Children's T-shirts Market on AliExpress

	Observations	Mean	Std Dev	Median	5th Pctile	95th Pctile
			Panel A. Listing Level			
Orders	1265	44.36	159	2	0	218
Price	1265	5.3	3.08	4.36	2.8	9.86
Revenue	1265	202.12	710.31	7.55	0	1009.6
Star Rating	626	4.8	.41	4.92	4	5
Total Comments	1265	27.27	111.7	1	0	135
Free Shipping Indicator	1265	.48	.5	0	0	1
Shipping Cost to US	1265	.68	.9	.21	0	2.42
			Panel B. Store Level			
Age	629	1.29	1.69	0	0	5
Total Orders	625	9090.42	18300.21	3472	46	30651
T-shirts Orders	638	87.96	246.16	4	0	532
T-shirts Revenue	638	400.75	1124.45	16.1	0	2253.91
Store Rating	598	4.72	.15	4.7	4.5	4.9

Note: This table reports the summary statistics for the children's t-shirts market on AliExpress. See Section 2.2 for details on the data. 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.

Table 3: Treatment Effects of Order and Review

	All Destinations		English-speaking Countries		United States	
Order	0.024	0.027*	0.015***	0.016***	0.017***	0.018***
	(0.020)	(0.014)	(0.005)	(0.005)	(0.003)	(0.003)
ReviewXPostReview	0.001	-0.017	0.020	0.016	0.016	0.014
	(0.023)	(0.027)	(0.019)	(0.017)	(0.017)	(0.016)
Observations	10270	10270	10270	10270	10270	10270
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE at listing level	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, calculated using the transaction data collected in August 2018. 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 review on shipping and product quality. "PostReview" is a dummy that equals one for the weeks after the reviews were given (i.e., from week 7 onward). Standard errors clustered at the listing level are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table 4: Dynamic Treatment Effects

	All Destinations		English-speaking Countries		United States	
OrderXMonth1	0.063***	0.064***	0.037***	0.037***	0.038***	0.038***
	(0.024)	(0.021)	(0.008)	(0.008)	(0.007)	(0.007)
OrderXMonth2	0.019	0.019	0.007	0.008	0.008**	0.008**
	(0.032)	(0.027)	(0.010)	(0.009)	(0.003)	(0.004)
OrderXMonth3	0.009	0.009	0.015*	0.015**	0.012*	0.013**
	(0.020)	(0.017)	(0.008)	(0.008)	(0.007)	(0.006)
OrderXMonth4	-0.044	-0.044	-0.009	-0.008	0.011	0.011
	(0.031)	(0.029)	(0.018)	(0.018)	(0.007)	(0.007)
Observations	10270	10270	10270	10270	10270	10270
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE at listing level	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, calculated using the transaction data collected in August 2018. 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 treatment. Standard errors clustered at the listing level are in the parentheses. *** implies significance at 0.01 level, ** 0.5, * 0.1.

Table 5: Average and Quantile Treatment Effects Measured at the Endline

	(1) Average	(2) 10th %	(3) 50th %	(4) 90th %	(5) 95th %	(6) 99th %
Orders from the US	0.248***	0.010	0.032*	0.673***	0.931***	3.856
	(0.063)	(0.019)	(0.019)	(0.198)	(0.263)	(2.447)
Orders from English-Speaking Countries	0.190**	0.020	0.050**	0.637***	1.080**	3.625*
	(0.093)	(0.025)	(0.022)	(0.239)	(0.464)	(1.962)
Orders from All Countries	0.110	0.096	0.149*	0.593	0.977	3.394
	(0.308)	(0.110)	(0.077)	(0.652)	(1.216)	(4.973)
Observations	790	790	790	790	790	790
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the average and quantile treatment effects of the experimentally generated orders and reviews. Each cell in the table reports a regression coefficient. The dependent variable is the endline number of cumulative orders, calculated using the transaction data collected in August 2018. The independent variable is the order treatment dummy that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. 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, and Columns 2 to 6 report the quantile treatment effects. Standard errors are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table 6: Estimated Parameters of the Empirical Model

Parameters	v_0	σ	ρ
Value	5.18	2.52	0.53
S.E.	(0.018)	(0.029)	(0.005)

Note: v_0 governs the initial visibility; σ is the review noise; ρ is the parameter that maps to the correlation between cost and quality. Standard errors are reported in the parentheses.

Table 7: Matching Moments

Moments	**Data**	**Model**
1. Market share distribution (v_0)		
Top 1% cumulative revenue share	0.179	0.235
Top 5% cumulative revenue share	0.537	0.512
Top 10% cumulative revenue share	0.721	0.657
Top 25% cumulative revenue share	0.910	0.837
Top 50% cumulative revenue share	0.981	0.947
2. Dependence of new order on cumulative orders (v_0)	0.103	0.129
3. Quality and sales relationship (σ)		
Cumulative orders share: Top 1/3 quality bin	0.434	0.479
Cumulative orders share: Middle 1/3 quality bin	0.311	0.294
4. Reg. coef. of log price and quality (ρ)	0.125	0.150

Note: This table reports the data moments and the model moments evaluated at the parameter estimates reported in Table 6. See Section 6 for more discussion on the choice of moments.

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 $\sigma = 2.52$			
1250 Sellers	0.479	0.404	0.254
625 Sellers	0.592	0.648	0.292
A.2 With $\sigma = 0$			
1250 Sellers	0.719	0.912	0.289
625 Sellers	0.792	1.054	0.336
	Panel B. Information Friction: the Review Noise		
B.1 Sales-based sampling, 1250 *sellers*			
$\sigma = 2.52$	0.479	0.404	0.254
$\sigma = 0$	0.719	0.913	0.289
B.2 Random sampling, 1250 *sellers*			
$\sigma = 2.52$	0.273	-0.041	0.125
$\sigma = 0$	0.355	0.176	0.111

Note: This table reports the results of several counterfactual analyses using the estimated model. Panel A compares market outcomes with different numbers of sellers under noisy and perfect signals. Panel B compares market outcomes with noisy and perfect signals under sales-based and random sampling protocols. Specifically, in the former case, visibility is proportional to the sum of v_0 and cumulative sales. In the latter case, consumers randomly choose listings to form their search sample.

Table 9: Model Robustness

	Total Share for Top 1/3 Quality	Sales-Weighted Quality	Average Consumer Surplus
		Panel A: Baseline	
1250 Sellers	0.479	0.404	0.254
625 Sellers	0.592	0.648	0.292
		Panel B: $K \sim$ Positive Poisson(5)	
1250 Sellers	0.528	0.499	0.259
625 Sellers	0.647	0.739	0.300
		Panel C: With Variety Effect: 125 varieties	
1250 Sellers	0.484	0.415	0.252
625 Sellers	0.594	0.640	0.285
		Panel D: Price Elasticity=4 ($\gamma = 1.89$)	
1250 Sellers	0.478	0.410	0.262
625 Sellers	0.602	0.676	0.307
		Panel E: Price Elasticity=10 ($\gamma = 0.75$)	
1250 Sellers	0.486	0.421	0.270
625 Sellers	0.610	0.681	0.315

Note: Panel A reproduces our baseline estimates reported in Panel B.1 of Table 8. In Panel B, the parameters are re-estimated with K drawn from the positive Poisson distribution with mean 5. The new estimates are $\hat{v}_0 = 5.18, \hat{\sigma} = 0.52, \hat{\rho} = 0.53, \hat{\beta} = 1.319$. Panels C and D re-simulated the market dynamics considering variety effects as described in Section 7.3. Panel C uses the baseline parameter estimates with K drawn from the positive Poisson distribution with mean 2, whereas Panel D uses the parameter estimates with K drawn from the positive Poisson distribution with mean 5. Panels E and F re-estimate the model under different values of γ, corresponding to different price elasticities. In Panel E, fixing $\gamma = 1.89$, the new estimates are $\hat{v}_0 = 6.258, \hat{\sigma} = 3.180, \hat{\rho} = 0.484, \hat{\beta} = 2.202$. In Panel F, fixed $\gamma = 0.75$, the new estimates are $\hat{v}_0 = 6.606, \hat{\sigma} = 3.537, \hat{\rho} = 0.444, \hat{\beta} = 2.197$.

Appendices. For Online Publication Only

A Theory Appendix

A.1 Proof of Proposition 1

We first illustrate the main ideas of the proof using the special case described in the main text, where $K = 2$ and the higher quality seller is chosen with fixed (instead of logit) probability 3/4. In this case the probability that seller 1 is chosen in period $t+1$ is

$$p^1_{t+1} = (r^1_t)^2 + 2r^1_t(1 - r^1_t) \cdot 3/4 = r^1_t(1.5 - 0.5r^1_t),$$

because $g_{11} = 1/2$ and $g_{1j} = 3/4$ for every $j > 1$. With probability p^1_{t+1} we have $r^1_{t+1} = \frac{v^1_t+1}{V_t+1}$, otherwise $r^1_{t+1} = \frac{v^1_t}{V_t+1}$. Hence we have

$$\begin{aligned}
\mathbb{E}[r^1_{t+1} \mid r^1_t] &= \frac{v^1_t}{V_t+1} + \frac{p^1_{t+1}}{V_t+1} \\
&= r^1_t - \frac{r^1_t}{V_t+1} + p^1_{t+1} \cdot \frac{1}{V_t+1} \\
&= r^1_t + \frac{0.5r^1_t(1-r^1_t)}{V_t+1} \\
&= r^1_t + \frac{0.5r^1_t(1-r^1_t)}{V_0+t+1}.
\end{aligned} \tag{A.1}$$

Since r^1_t is always bounded in $(0,1)$, the above identity shows that the process $\{r^1_t\}$ is a bounded *sub-martingale* with $\mathbb{E}[r^1_{t+1} \mid r^1_t] > r^1_t$. Hence there exists a limit random variable $r^1_\infty \in [0,1]$ such that $r^1_t \to r^1_\infty$ almost surely.[42]

We now show that r^1_∞ is almost surely equal to 1. For this it turns out to be useful to consider a different sub-martingale process, given by $\log(r^1_t) - \frac{1}{v^1_t}$. Its evolution is such that with probability p^1_{t+1} we have $\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} = \log(\frac{v^1_t+1}{V_t+1}) - \frac{1}{v^1_t+1}$, and otherwise $\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} = \log(\frac{v^1_t}{V_t+1}) - \frac{1}{v^1_t}$. From this we calculate

$$\begin{aligned}
\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] &= \log(\frac{v^1_t}{V_t+1}) - \frac{1}{v^1_t} + p^1_{t+1} \cdot \left(\log(\frac{v^1_t+1}{v^1_t}) + \frac{1}{v^1_t(v^1_t+1)}\right) \\
&= \log(r^1_t) - \frac{1}{v^1_t} - \log(\frac{V_t+1}{V_t}) + p^1_{t+1} \cdot \left(\log(\frac{v^1_t+1}{v^1_t}) + \frac{1}{v^1_t(v^1_t+1)}\right) \\
&> \log(r^1_t) - \frac{1}{v^1_t} - \frac{1}{V_t} + p^1_{t+1} \cdot \frac{1}{v^1_t},
\end{aligned} \tag{A.2}$$

where the last 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 $p^1_{t+1} = r^1_t(1.5 - 0.5r^1_t) \geq r^1_t = \frac{v^1_t}{V_t}$, the above equation tells us that $\{\log(r^1_t) - \frac{1}{v^1_t}\}$ is also a sub-martingale that

[42] In the standard Polya Urn model without choice, $p^1_{t+1} = r^1_t$ and thus r^1_t is exactly a martingale.

is bounded above by zero.[43]

Now suppose r^1_∞ has 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 T sufficiently large such that $\mathbb{P}\{r^1_t < 1-\epsilon\} > \epsilon$ for every $t \geq T$. Since $p^1_{t+1} = r^1_t(1.5 - 0.5r^1_t)$, there is probability at least ϵ that $p^1_{t+1} \geq (1+0.5\epsilon)r^1_t$. So in Equation (A.2) there is probability at least ϵ that

$$\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t}) \geq -\frac{1}{V_t} + \frac{(1+0.5\epsilon)r^1_t}{v^1_t} = \frac{\epsilon}{2V_t}.$$

With 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 non-negative, so taking an ex-ante expectation yields

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

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

$$\mathbb{E}[\log(r^1_\tau) - \frac{1}{v^1_\tau}] - \mathbb{E}[(\log(r^1_T) - \frac{1}{v^1_T})] \geq \sum_{t=T}^{\tau-1} \frac{\epsilon^2}{2V_t} = \sum_{t=T}^{\tau-1} \frac{\epsilon^2}{2(V_0+t)}.$$

But this is a contradiction because the LHS remains bounded as $\tau \to \infty$, whereas the RHS is a harmonic sum that diverges. Therefore $r^1_\infty = 1$ almost surely, as we desire to show.

We now generalize this argument to any $K \geq 2$ and any logit parameter $\alpha > 0$. Logit choice satisfies the following property: In any sample of K sellers where seller 1 appears $k < K$ times, the probability of choosing seller 1 strictly exceeds $\frac{k}{K}$. Below we show that this property alone is sufficient to guarantee long run monopoly. Indeed, since the possible samples of size K are finite, we can choose $\eta > 0$ such that in any sample of K sellers where seller 1 appears $k < K$ times, the probability of choosing seller 1 is at least $\frac{(1+\eta)k}{K}$. Hence, by considering different k, we know that the choice probability of seller 1

[43] Because $\frac{1}{v^1_t}$ converges as a monotonically decreasing sequence, the fact that $\log(r^1_t) - \frac{1}{v^1_t}$ converges almost surely in fact provides a different proof that r^1_t converges almost surely.

satisfies

$$
\begin{aligned}
p^1_{t+1} &\geq (r^1_t)^K + \sum_{k=1}^{K-1} \binom{K}{k} (r^1_t)^k (1-r^1_t)^{K-k} \cdot \frac{(1+\eta)k}{K} \\
&= -\eta (r^1_t)^K + \sum_{k=1}^{K} \binom{K}{k} (r^1_t)^k (1-r^1_t)^{K-k} \cdot \frac{(1+\eta)k}{K} \\
&= -\eta (r^1_t)^K + (1+\eta) r^1_t \sum_{k=1}^{K} \binom{K-1}{k-1} (r^1_t)^{k-1} (1-r^1_t)^{K-k} \\
&= -\eta (r^1_t)^K + (1+\eta) r^1_t \\
&\geq -\eta (r^1_t)^2 + (1+\eta) r^1_t \\
&= r^1_t (1 + \eta(1-r^1_t)).
\end{aligned}
\tag{A.3}
$$

Thus we can modify the previous Equation (A.1) to get

$$
\begin{aligned}
\mathbb{E}[r^1_{t+1} \mid r^1_t] &= \frac{v^1_t}{V_t+1} + \mathbb{E}[p^1_{t+1} \mid r^1_t] \cdot \frac{1}{V_t+1} \\
&\geq r^1_t - \frac{r^1_t}{V_t+1} + \frac{r^1_t(1+\eta(1-r^1_t))}{V_t+1} \\
&= r^1_t + \frac{\eta r^1_t (1-r^1_t)}{V_t+1} \\
&= r^1_t + \frac{\eta r^1_t (1-r^1_t)}{V_0+t+1},
\end{aligned}
$$

with η replacing the constant 0.5 in Equation (A.1). This shows that the process $\{r^1_t\}$ is still a bounded sub-martingale which converges to a limit random variable r^1_∞.

Moreover, we can follow essentially the same argument as before to deduce $r^1_\infty = 1$ almost surely. Specifically, from $p^1_{t+1} \geq r^1_t(1+\eta(1-r^1_t))$ we can deduce that for $t \geq T$ (see the above argument for the choice of T), there is probability at least ϵ that $1 - r^1_t \geq \epsilon$ and thus $p_{t+1} \geq (1+\eta\epsilon) r^1_t$. (A.2) implies that in this event $\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t}) \geq \frac{\eta\epsilon}{V_t}$. So for $t \geq T$ we have

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

This leads to the same contradiction as before, which completes the proof.

A.2 Proof of Proposition 2

As explained in the main text, doubling the number of sellers (in the setting with known qualities) is equivalent to doubling initial visibility levels v^i_0. So we focus on the effect of increasing v^i_0 on short run welfare.

Note that the expected quality the consumer gets in period $t+1$ can be written as

$$Q_{t+1} = \mathbb{E}\left[\sum_{i=1}^{N} p_{t+1}^i \cdot q^i\right] = \sum_{i=1}^{N} \mathbb{E}[p_{t+1}^i] \cdot q^i, \quad \forall t \geq 0.$$

The choice probabilities p_{t+1}^i are functions of the previous relative visibility levels r_t^i. Since at $t=0$ these r_t^i are unaffected by a common scaling of v_0^i, so are the choice probabilities in period 1. Thus the expected quality in period 1 is unchanged under scaling. This explains why the result only states a strict decrease in welfare from period 2 onwards.

In our proof it will be convenient to use $T+2$ in place of T, and correspondingly assume $V_0 \geq \frac{1}{2}K(T+2)^2$. We will demonstrate a strict decrease in the expected quality in periods $2 \sim T+2$. The expected quality in period $t+2$ is $\sum_{i=1}^{N} \mathbb{E}[p_{t+2}^i] \cdot q^i$, which by the Abel summation formula can be rewritten as

$$\mathbb{E}[p_{t+2}^1] \cdot (q^1 - q^2) + \left(\mathbb{E}[p_{t+2}^1] + \mathbb{E}[p_{t+2}^2]\right) \cdot (q^2 - q^3) + \cdots + \left(\mathbb{E}[p_{t+2}^1] + \cdots \mathbb{E}[p_{t+2}^{N-1}]\right) \cdot (q^{N-1} - q^N) + q^N,$$

where the last term is q^N because $\mathbb{E}[p_{t+2}^1] + \cdots \mathbb{E}[p_{t+2}^N] = 1$ as $p_{t+2}^1 + \cdots + p_{t+2}^N \equiv 1$. We will show that for each $1 \leq I \leq N-1$, scaling v_0^i leads to a strict decrease in $\sum_{i \leq I} \mathbb{E}[p_{t+2}^i]$ for $0 \leq t \leq T$. This is sufficient to imply that the expected quality also decreases. Intuitively, if choice probabilities shift from higher quality sellers down to lower quality sellers, then expected quality decreases.

Studying $\sum_{i \leq I} \mathbb{E}[p_{t+2}^i]$ for a general choice rule turns out to be not very easy, so we again illustrate our proof strategy in the special case where $K=2$, and the higher quality seller in the sample is chosen with probability 3/4. For now let us also focus on $I=1$, so we are simply dealing with $\mathbb{E}[p_{t+2}^1]$, the expected choice probability of seller 1 in period $t+2$. Our goal is to show the following lemma:

Lemma 1. *Suppose $V_0 \geq (T+2)^2$, then in this special case $\mathbb{E}[p_{t+2}^1]$ strictly decreases for every $0 \leq t \leq T$, under any proportional increase in $\{v_0^i\}$.*

Proof. We begin with the observation that $r_{t+1}^1 = \frac{r_t^1 V_t + 1}{V_t + 1}$ with probability p_{t+1}^1, and $r_{t+1}^1 = \frac{r_t^1 V_t}{V_t + 1}$ with remaining probability $1 - p_{t+1}^1$. Thus, as calculated before,

$$\mathbb{E}[r_{t+1}^1] = \mathbb{E}[r_t^1] + \frac{\mathbb{E}[p_{t+1}^1 - r_t^1]}{V_t + 1}.$$

We can also calculate that

$$\begin{aligned}
\mathbb{E}[(r_{t+1}^1)^2] &= \mathbb{E}\left[(\frac{r_t^1 V_t}{V_t + 1})^2 + p_{t+1}^1 \cdot \frac{2r_t^1 V_t + 1}{(V_t+1)^2}\right] \\
&= \mathbb{E}[(r_t^1)^2] + \mathbb{E}\left[\frac{-(r_t^1)^2(2V_t+1) + p_{t+1}^1(2r_t^1 V_t + 1)}{(V_t+1)^2}\right] \\
&= \mathbb{E}[(r_t^1)^2] + \frac{\mathbb{E}[p_{t+1}^1 - (r_t^1)^2]}{(V_t+1)^2} + \frac{\mathbb{E}[(p_{t+1}^1 - r_t^1) r_t^1] \cdot 2V_t}{(V_t+1)^2}.
\end{aligned}$$

These identities are true in general, but in our special case, we can plug in $p_{t+1}^1 = r_t^1(1.5 - 0.5r_t^1)$ to

rewrite the RHS in terms of r_t^1 only. With a bit of algebra we obtain

$$\begin{aligned}
&1.5\mathbb{E}[r_{t+1}^1] - 0.5\mathbb{E}[(r_{t+1}^1)^2] \\
=&1.5\mathbb{E}[r_t^1] - 0.5\mathbb{E}[(r_t^1)^2] + 1.5\frac{\mathbb{E}[0.5r_t^1(1-r_t^1)]}{V_t+1} - 0.5\frac{\mathbb{E}[1.5r_t^1(1-r_t^1)]}{(V_t+1)^2} - 0.5\frac{\mathbb{E}[0.5(r_t^1)^2(1-r_t^1)]\cdot 2V_t}{(V_t+1)^2} \\
=&1.5\mathbb{E}[r_t^1] - 0.5\mathbb{E}[(r_t^1)^2] + \frac{\mathbb{E}[(r_t^1)(1-r_t^1)(0.75-0.5r_t^1)]\cdot V_t}{(V_t+1)^2}.
\end{aligned}$$

That is, as described also in the main text,

$$\mathbb{E}[p_{t+2}^1] - \mathbb{E}[p_{t+1}^1] = \frac{V_t}{(V_t+1)^2}\cdot\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]. \tag{A.4}$$

This key identity tells us that the choice probability of seller 1 increases over time. Moreover, the increment on the RHS involves a factor $\frac{V_t}{(V_t+1)^2}$ which depends on V_0 since $V_t = V_0 + t$. Note that the function $\frac{x}{(x+1)^2}$ decreases in x precisely when $x \geq 1$. Thus we need $V_0 \geq 1$ to ensure $\frac{V_t}{(V_t+1)^2}$ decreases under any scaling of initial visibility levels. This is in fact necessary and sufficient for $\mathbb{E}[p_{t+2}^1]$ to decrease under scaling for $t = 0$, because the other terms $\mathbb{E}[p_1^1]$ and $\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]$ in Equation (A.4) depend only on r_0^1, which is a constant unaffected by scaling.

To ensure $\mathbb{E}[p_{t+2}^1]$ continues to decrease under scaling for $t \leq T$, we apply induction on t and seek to show that the increment $\frac{V_t}{(V_t+1)^2}\cdot\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]$ decreases under scaling. Intuitively, our assumption $V_0 \geq (T+2)^2$ ensures that the first factor $\frac{V_t}{(V_t+1)^2}$ decreases "significantly" under scaling. If that is the case, then even though the second factor $\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]$ may increase, this increase will be smaller in magnitude than the decrease in $\frac{V_t}{(V_t+1)^2}$. Below we fill in the details.

Consider increasing each v_0^i to $(1+\epsilon)v_0^i$ for some $\epsilon > 0$. We use $\widetilde{v_0^i} = (1+\epsilon)v_0^i$ and $\widetilde{V_0} = (1+\epsilon)V_0$ to denote the resulting primitive parameters. Using similar notation, we let $\widetilde{r_t^i}$ to denote seller i's (random) relative visibility at the end of period t, under the scaled primitive parameters. For induction to work we want to show that if $V_0 \geq (T+2)^2$ then

$$\frac{V_t}{(V_t+1)^2}\cdot\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)] > \frac{\widetilde{V_t}}{(\widetilde{V_t}+1)^2}\cdot\mathbb{E}[\widetilde{r_t^1}(1-\widetilde{r_t^1})(0.75-0.5\widetilde{r_t^1})] \quad \forall 0 \leq t \leq T, \tag{A.5}$$

where $\widetilde{V_t} = \widetilde{V_0} + t = (1+\epsilon)V_0 + t$. We focus on the case $t = T$, since the result for earlier periods is weaker in the sense that $V_0 \geq (T+2)^2$ automatically implies $V_0 \geq (t+2)^2$.

The key difficulty is to estimate the ratio $\frac{\mathbb{E}[\widetilde{r_t^1}(1-\widetilde{r_t^1})(0.75-0.5\widetilde{r_t^1})]}{\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]}$. To do this, we consider any "purchase history" $b_1, \ldots, b_T \in \{1, \ldots, N\}$, where $b_t = i$ means that in period t the consumer purchases from seller i. Any such history determines the evolution of visibility levels in the original model and in the scaled

version:

$$r_t^i = \frac{v_0^i + |\{1 \le \tau \le t : \ b_\tau = i\}|}{V_0 + t};$$
$$\widetilde{r_t^i} = \frac{(1+\epsilon)v_0^i + |\{1 \le \tau \le t : \ b_\tau = i\}|}{(1+\epsilon)V_0 + t}.$$

Thus, when we look at the same b-sequence, the corresponding relative visibility levels cannot increase very much under scaling: with $C = |\{1 \le \tau \le t : \ b_\tau = i\}$ we have

$$\frac{\widetilde{r_t^i}}{r_t^i} = \frac{((1+\epsilon)v_0^i + C) \cdot (V_0 + t)}{(v_0^i + C) \cdot ((1+\epsilon)V_0 + t)} = 1 + \frac{\epsilon v_0^i t - \epsilon V_0 C}{(v_0^i + C) \cdot ((1+\epsilon)V_0 + t)} \le 1 + \frac{\epsilon t}{V_0} \le e^{\frac{\epsilon t}{V_0}} \quad \forall 0 \le t \le T.$$

Since this holds for every i, we also have $\frac{1-\widetilde{r_t^1}}{1-r_t^1} = \frac{\sum_{i>1} \widetilde{r_t^i}}{\sum_{i>1} r_t^1} \le e^{\frac{\epsilon t}{V_0}}$, and likewise the same upper bound applies to the ratio $\frac{0.75-0.5\widetilde{r_t^1}}{0.75-0.5r_t^1}$.

It follows that, for the same b-sequence,

$$\frac{\widetilde{r_t^1}(1-\widetilde{r_t^1})(0.75-0.5\widetilde{r_t^1})}{r_t^1(1-r_t^1)(0.75-0.5r_t^1)} \le e^{\frac{3\epsilon T}{V_0}}.$$

Now, we cannot directly conclude that the ratio of expectations is also bounded by $e^{\frac{3\epsilon T}{V_0}}$, because the same b-sequence does not occur with the same probability under the original process as under the scaled process. We have to also estimate the ratio of these probabilities for the same b-sequence. Specifically, under the original process the probability of $b_1, \ldots, b_T$ can be written as $\prod_{t=0}^{T-1} P^{b_t}(r_t^1, \ldots, r_t^N)$, where P^i represents the choice probability of seller i as a function of relative visibility levels.

Similarly, the probability of $b_1, \ldots, b_T$ under the scaled process is $\prod_{t=0}^{T-1} P^{b_t}(\widetilde{r_t^1}, \ldots, \widetilde{r_t^N})$. Since each P^i is a quadratic polynomial with non-negative coefficients, and $\frac{\widetilde{r_t^i}}{r_t^i} \le e^{\frac{\epsilon t}{V_0}}$, we have $\frac{P^{b_t}(\widetilde{r_t^1},\ldots,\widetilde{r_t^N})}{P^{b_t}(r_t^1,\ldots,r_t^N)} \le e^{\frac{2\epsilon t}{V_0}}$ for each t. Thus under scaling of initial visibility levels, the probability of any b-sequence increases by a factor of at most $\prod_{t=0}^{T-1} e^{\frac{2\epsilon t}{V_0}} = e^{\frac{\epsilon T(T-1)}{V_0}}$.

Hence we arrive at the following bound:

$$\frac{\mathbb{E}[\widetilde{r_t^1}(1-\widetilde{r_t^1})(0.75-0.5\widetilde{r_t^1})]}{\mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)]} \le e^{\frac{3\epsilon T}{V_0}} \cdot e^{\frac{\epsilon T(T-1)}{V_0}} = e^{\frac{\epsilon T(T+2)}{V_0}}.$$

To prove the desired inequality $\frac{V_T}{(V_T+1)^2} \cdot \mathbb{E}[r_t^1(1-r_t^1)(0.75-0.5r_t^1)] > \frac{\widetilde{V_T}}{(\widetilde{V_T}+1)^2} \cdot \mathbb{E}[\widetilde{r_t^1}(1-\widetilde{r_t^1})(0.75-0.5\widetilde{r_t^1})]$, it suffices to show that

$$\log \frac{V_T}{(V_T+1)^2} - \log \frac{\widetilde{V_T}}{(\widetilde{V_T}+1)^2} > \frac{\epsilon T(T+2)}{V_0},$$

which is equivalent to

$$\log \frac{\widetilde{V_T}}{V_T} > \frac{\epsilon T(T+2)}{V_0} + 2 \log \frac{(V_T+1)\widetilde{V_T}}{(\widetilde{V_T}+1)V_T}$$

On the RHS, $\log \frac{(V_T+1)\widetilde{V_T}}{(\widetilde{V_T}+1)V_T} = \log(1+\frac{\widetilde{V_T}-V_T}{(\widetilde{V_T}+1)V_T}) = \log(1+\frac{\epsilon V_0}{(\widetilde{V_T}+1)V_T}) \leq \frac{\epsilon V_0}{(\widetilde{V_T}+1)V_T} \leq \frac{\epsilon}{V_0}$. So the entire RHS is bounded above by $\epsilon \cdot \frac{T^2+2T+2}{V_0}$. In comparison, the LHS is $\log \frac{\widetilde{V_T}}{V_T} = \log(1+\frac{\epsilon V_0}{V_0+T}) \geq \frac{\epsilon V_0}{V_0+T} - (\frac{\epsilon V_0}{V_0+T})^2 \geq \frac{\epsilon V_0}{V_0+T} - \epsilon^2$, where we use the inequality $\log(1+x) \geq x - x^2$ which is valid for all $x \geq 0$.

Therefore, this argument goes through as long as $\epsilon \cdot \frac{V_0}{V_0+T} - \epsilon^2 > \epsilon \cdot \frac{T^2+2T+2}{V_0}$. Since $V_0 \geq (T+2)^2 > T^2+2T+2$, we have $\frac{T^2+2T+2}{V_0} \leq \frac{T^2+2T+2+T}{V_0+T}$. Thus the preceding inequality holds whenever $\epsilon \cdot \frac{V_0}{V_0+T} - \epsilon^2 > \epsilon \cdot \frac{T^2+3T+2}{V_0+T}$, which can be simplified to $\frac{V_0-T^2-3T-2}{V_0+T} > \epsilon$. Again using $V_0 \geq (T+2)^2$, we have $\frac{V_0-T^2-3T-2}{V_0+T} \geq \frac{(T+2)^2-T^2-3T-2}{(T+2)^2+T} = \frac{T+2}{T^2+5T+4} > \frac{1}{T+3}$. Putting it together, we conclude that whenever $\epsilon \in (0, \frac{1}{T+3}]$, scaling v_0^i to $(1+\epsilon)v_0^i$ would lead to the desired inequality (A.5). But then the same inequality holds for all $\epsilon > 0$, because a large scaling can be decomposed into a sequence of small scalings involving $\epsilon \leq \frac{1}{T+3}$. ■

By now we have shown that increasing initial visibility decreases $\mathbb{E}[p_{t+2}^1]$ for every $t \leq T$. To finish the proof for this special case, we need to also show that for every I, increasing initial visibility decreases $\mathbb{E}[\sum_{i \leq I} p_{t+2}^i]$. This generalization is straightforward thanks to the fixed probability choice rule, because the total choice probability of the top I sellers depends only on their total visibility shares:

$$\sum_{i \leq I} p_{t+2}^i = (\sum_{i \leq I} p_{t+2}^i)^2 + 2(\sum_{i \leq I} p_{t+2}^i) \cdot (\sum_{j > I} p_{t+2}^j) \cdot 3/4 = (\sum_{i \leq I} r_{t+1}^i) \cdot (1.5 - 0.5 \sum_{i \leq I} r_{t+1}^i).$$

Thus we can study the evolution of $\sum_{i \leq I} r_{t+1}^i$ and $\sum_{i \leq I} p_{t+2}^i$ by "merging" the top I sellers, without worrying about their individual visibility levels and choice probabilities.[44] This procedure reduces the problem to the previous case $I = 1$, so the same analysis shows that $\mathbb{E}[\sum_{i \leq I} p_{t+2}^i]$ decreases. Hence expected quality decreases in the first $T+2$ periods, as we desire to show.

To generalize this proof to any sample size $K \geq 2$ and any logit choice rule, the main challenge is to derive a recursive formula similar to Equation (A.4). This is recorded in the next lemma:

Lemma 2. *Fix the primitives $N, \{q^i\}, K, \alpha$ and also fix $I \in \{1, \ldots, N-1\}$. Then there exist polynomials $H_1, \ldots, H_{K-1}$ with the following properties:*

1. *Each H_m is a nonzero N-variable homogeneous polynomial with degree $2K - m$ and non-negative coefficients;*
2. *The coefficients of H_m depend only on $N, \{q^i\}, K, \alpha, I$ and not on initial visibility levels $\{v_0^i\}$;*
3. *Under our model it holds that*

$$\sum_{i \leq I} \mathbb{E}[p_{t+2}^i] - \sum_{i \leq I} \mathbb{E}[p_{t+1}^i] = \sum_{m=1}^{K-1} \frac{(V_t)^{K-m}}{(V_t+1)^K} \cdot \mathbb{E}[H_m(r_t^1, \ldots, r_t^N)].$$

[44] Previously we have argued that sellers with the same quality level can be combined. The point here is that the fixed probability choice rule has the special feature that sellers with unequal quality levels can also be combined, so long as they constitute the highest (or lowest) quality sellers.

Equation (A.4) can be seen as a special case of this lemma, once we write $r_t^1(1-r_t^1)(0.75-0.5r_t^1)$ as a homogeneous polynomial in $r_t^1,\dots,r_t^N$ using the fact that they sum to 1. For a general choice rule, it is not straightforward to write down H_m explicitly. The crux of our argument is thus to find a way to express H_m in a compact way, and use properties of the logit choice rule to verify it has non-negative coefficients. We defer the technical proof of this lemma to later, and first demonstrate how to use it to prove Proposition 2.

The strategy is the same as before: We claim that if $V_0 \geq \frac{1}{2}K(T+2)^2$, then for each m and each $t \leq T$ it holds that

$$\frac{(V_t)^{K-m}}{(V_t+1)^K} \cdot \mathbb{E}[H_m(r_t^1,\dots,r_t^N)] > \frac{\widetilde{V_t}^{K-m}}{(\widetilde{V_t}+1)^K} \cdot \mathbb{E}[H_m(\widetilde{r_t^1},\dots,\widetilde{r_t^N})]. \tag{A.6}$$

Once this is proved, then we can show by induction that $\sum_{i\leq I}\mathbb{E}[p_{t+2}^i] > \sum_{i\leq I}\mathbb{E}[\widetilde{p_{t+2}^i}]$ for every I and every $t \leq T$. Thus expected quality decreases in the first $T+2$ periods.

To prove the above inequality (A.6), we focus on the most difficult case $m=1$ and $t=T$. It will be clear that the same argument applies to every m and $t \leq T$. To bound the ratio $\frac{\mathbb{E}[H_1(\widetilde{r_T^1},\dots,\widetilde{r_T^N})]}{\mathbb{E}[H_1(r_T^1,\dots,r_T^N)]}$, we again consider a given purchase history $b_1,\dots,b_T$. The associated visibility levels satisfy $\frac{\widetilde{r_T^i}}{r_T^i} \leq e^{\frac{\epsilon T}{V_0}}$ as shown before. Since H_1 is a homogeneous polynomial with degree $2K-1$ and non-negative coefficients, we deduce that for the same b-sequence

$$\frac{H_1(\widetilde{r_T^1},\dots,\widetilde{r_T^N})}{H_1(r_T^1,\dots,r_T^N)} \leq e^{\frac{\epsilon(2K-1)T}{V_0}}.$$

Moreover, while scaling can affect the probability of any given b-sequence, the magnitude of this probability increase (in ratio terms) is

$$\prod_{t=0}^{T-1}\frac{P^{b_t}(\widetilde{r_t^1},\dots,\widetilde{r_t^N})}{P^{b_t}(r_t^1,\dots,r_t^N)} \leq \prod_{t=0}^{T-1} e^{\frac{\epsilon K t}{V_0}} = e^{\frac{\epsilon\frac{K}{2}(T^2-T)}{V_0}},$$

where the inequality uses the fact that $P^i(r_t^1,\dots,r_t^N) = K\sum_{i_1,\dots,i_K:\ i_1=i}\prod_{k=1}^K r_t^{i_k}\cdot g_{i_1,i_{-1}}$ is a homogeneous polynomial with degree K and non-negative coefficients.

Hence we arrive at the following bound:

$$\frac{\mathbb{E}[H_1(\widetilde{r_T^1},\dots,\widetilde{r_T^N})]}{\mathbb{E}[H_1(r_T^1,\dots,r_T^N)]} \leq e^{\frac{\epsilon(2K-1)T}{V_0}} \cdot e^{\frac{\epsilon\frac{K}{2}(T^2-T)}{V_0}} = e^{\epsilon\cdot\frac{\frac{K}{2}T^2+\frac{3K-2}{2}T}{V_0}}.$$

To prove (A.6) it suffices to show that

$$\log\frac{(V_T)^{K-1}}{(V_T+1)^K} - \log\frac{(\widetilde{V_T})^{K-1}}{(\widetilde{V_T}+1)^K} > \epsilon\cdot\frac{\frac{K}{2}T^2+\frac{3K-2}{2}T}{V_0},$$

which is equivalent to

$$\log \frac{\widetilde{V_T}}{V_T} > \epsilon \cdot \frac{\frac{K}{2}T^2 + \frac{3K-2}{2}T}{V_0} + K \log \frac{(V_T+1)\widetilde{V_T}}{(\widetilde{V_T}+1)V_T}$$

On the RHS, $\log \frac{(V_T+1)\widetilde{V_T}}{(\widetilde{V_T}+1)V_T} = \log(1 + \frac{\widetilde{V_T}-V_T}{(\widetilde{V_T}+1)V_T}) = \log(1 + \frac{\epsilon V_0}{(\widetilde{V_T}+1)V_T}) \leq \frac{\epsilon V_0}{(\widetilde{V_T}+1)V_T} \leq \frac{\epsilon}{V_0}$. So the entire RHS is bounded above by $\epsilon \cdot \frac{\frac{K}{2}T^2+\frac{3K-2}{2}T+K}{V_0}$. In comparison, the LHS is $\log \frac{\widetilde{V_T}}{V_T} = \log(1 + \frac{\epsilon V_0}{V_0+T}) \geq \frac{\epsilon V_0}{V_0+T} - (\frac{\epsilon V_0}{V_0+T})^2 \geq \frac{\epsilon V_0}{V_0+T} - \epsilon^2$. Thus we just need

$$\epsilon \cdot \frac{V_0}{V_0+T} - \epsilon^2 > \epsilon \cdot \frac{\frac{K}{2}T^2 + \frac{3K-2}{2}T + K}{V_0}.$$

Since $V_0 \geq \frac{K}{2}(T+2)^2 > \frac{K}{2}T^2 + \frac{3K-2}{2}T + K$, we have $\frac{\frac{K}{2}T^2+\frac{3K-2}{2}T+K}{V_0} \leq \frac{\frac{K}{2}T^2+\frac{3K-2}{2}T+K+T}{V_0+T}$. Thus the preceding displayed inequality holds whenever $\epsilon \cdot \frac{V_0}{V_0+T} - \epsilon^2 > \epsilon \cdot \frac{\frac{K}{2}T^2+\frac{3K}{2}T+K}{V_0+T}$, which can be simplified to $\frac{V_0-\frac{K}{2}T^2-\frac{3K}{2}T-K}{V_0+T} > \epsilon$. Again using $V_0 \geq \frac{K}{2}(T+2)^2$, we have $\frac{V_0-\frac{K}{2}T^2-\frac{3K}{2}T-K}{V_0+T} \geq \frac{\frac{K}{2}(T+2)^2-\frac{K}{2}T^2-\frac{3K}{2}T-K}{\frac{K}{2}(T+2)^2+T} = \frac{\frac{K}{2}(T+2)}{\frac{K}{2}(T+2)^2+T} > \frac{1}{T+3}$.

We conclude that (A.6) holds for any scaling factor $1+\epsilon$ with $\epsilon \leq \frac{1}{T+3}$. Because a larger scaling factor can be decomposed into a sequence of small scalings involving $\epsilon \leq \frac{1}{T+3}$, the same inequality (A.6) holds for all scalings. This completes the proof that any upward scaling of initial visibility levels decreases welfare in the first $T+2$ periods, modulo Lemma 2 which we prove below.

Proof of Lemma 2. Throughout, let $\phi = (i_1, \ldots, i_K)$ denote a generic sample of size K. We have $p^i_{t+1} = K \sum_{\phi:\ i_1=i} \left(\prod_{k=1}^K r_t^{i_k} \cdot g_{i_1,i_{-1}}\right)$. Thus

$$(V_t)^K \cdot p^i_{t+1} = K \sum_{\phi:\ i_1=i} \left(\prod_{k=1}^K v_t^{i_k} \cdot g_{i_1,i_{-1}}\right).$$

Likewise at the next period it holds that

$$(V_t+1)^K \cdot p^i_{t+2} = K \sum_{\phi:\ i_1=i} \left(\prod_{k=1}^K v_{t+1}^{i_k} \cdot g_{i_1,i_{-1}}\right).$$

Conditioning on period t visibility shares r_t (which we use as a shorthand for the vector $r_t^1, \ldots, r_t^N$), and taking the difference of the previous two equations, we obtain

$$(V_t+1)^K \cdot \mathbb{E}[p^i_{t+2} \mid r_t] - (V_t)^K \cdot p^i_{t+1} = K \sum_{\phi:\ i_1=i} g_{i_1,i_{-1}} \cdot \left(\mathbb{E}\left[\prod_{k=1}^K v_{t+1}^{i_k} - \prod_{k=1}^K v_t^{i_k} \mid r_t\right]\right). \tag{A.7}$$

We now study the difference $\prod_{k=1}^K v_{t+1}^{i_k} - \prod_{k=1}^K v_t^{i_k}$ carefully and seek to express it in a tractable way. For each seller j, there is probability p^j_{t+1} that it is chosen in period $t+1$, in which case $v^j_{t+1} = v^j_t + 1$

while the visibility of other sellers are unchanged from period t. Thus with probability p^j_{t+1}, we have

$$\prod_{k=1}^{K} v_{t+1}^{i_k} - \prod_{k=1}^{K} v_t^{i_k} = \left(\prod_{k:\ i_k=j} (v_t^{i_k}+1) - \prod_{k:\ i_k=j} v_t^{i_k} \right) \cdot \prod_{k:\ i_k\neq j} v_t^{i_k}$$
$$= \sum_{\emptyset\subsetneq A\subset\{k:\ i_k=j\}} \prod_{k\notin A} v_t^{i_k},$$

where the set A here ranges across all non-empty subsets of those k such that $i_k = j$. The last equality holds by expanding the product $\prod_{k:\ i_k=j}(v_t^{i_k}+1)$.

It follows that for a given sample $\phi = (i_1, \ldots, i_K)$,

$$\mathbb{E}\left[\prod_{k=1}^{K} v_{t+1}^{i_k} - \prod_{k=1}^{K} v_t^{i_k} \mid r_t\right] 1 = \sum_{j=1}^{N} p^j_{t+1} \left(\sum_{\emptyset\subsetneq A\subset\{k:\ i_k=j\}} \prod_{k\notin A} v_t^{i_k} \right)$$
$$= \sum_{j=1}^{N} p^j_{t+1} \left(\sum_{\emptyset\subsetneq A\subset\{1,\ldots,K\}} 1_{\{i_k=j\ \forall k\in A\}} \prod_{k\notin A} v_t^{i_k} \right),$$

where we introduce the indicator function $1_{\{i_k=j\ \forall k\in A\}}$ to expand the inner summation to all non-empty subsets A of $\{1, \ldots, K\}$. When plugged back into (A.7), this allows us to interchange the order of summation and deduce

$$\begin{aligned}(V_t+1)^K \cdot \mathbb{E}[p^i_{t+2} \mid r_t] - (V_t)^K \cdot p^i_{t+1} &= K \sum_{\phi:\ i_1,\ldots,i_K:\ i_1=i} g_{i_1,i_{-1}} \sum_j p^j_{t+1} \sum_{A\neq\emptyset} 1_{\{i_k=j\ \forall k\in A\}} \prod_{k\notin A} v_t^{i_k} \\ &= K \sum_j p^j_{t+1} \sum_{A\neq\emptyset} \left(\sum_{\phi:\ i_1=i \& i_k=j\ \forall k\in A} g_{i_1,i_{-1}} \prod_{k\notin A} v_t^{i_k} \right).\end{aligned} \tag{A.8}$$

The innermost summation above is over all samples ϕ such that $i_1 = i$ and $i_k = j$ for each $k \in A$.

To further simply the RHS of (A.8), we distinguish between $j = i$ and $j \neq i$ in the outermost summation. First consider $j = i$, then there are two sub-cases: either $1 \in A$ or $1 \notin A$. If $1 \in A$, then for each $1 \leq m \leq K$, there are $\binom{K-1}{m-1}$ possible sets A that contain 1 and have size m. By symmetry, for each such set A the sum $\sum_{\phi:\ i_1=i \& i_k=i\ \forall k\in A} g_{i_1,i_{-1}} \prod_{k\notin A} v_t^{i_k}$ is the same as $A = \{1, \ldots, m\}$, and equals $\sum_{\phi:\ i_1=\cdots=i_m=i} g_{i_1,i_{-1}} \prod_{k>m} v_t^{i_k}$. Summing across m, we deduce that the contribution of $j = i$ and $1 \in A$ to the RHS of (A.8) is

$$p^i_{t+1} \sum_{m=1}^{K} K \binom{K-1}{m-1} \left(\sum_{\phi:\ i_1=\cdots=i_m=i} g_{i_1,i_{-1}} \prod_{k>m} v_t^{i_k} \right) = p^i_{t+1} \sum_{m=1}^{K} \binom{K}{m} \left(\sum_{\phi:\ i_1=\cdots=i_m=i} m \cdot g_{i_1,i_{-1}} \prod_{k>m} v_t^{i_k} \right), \tag{A.9}$$

where we used the identity $K\binom{K-1}{m-1} = m\binom{K}{m}$. Now recall the identity for total choice probability: $\sum_{k=1}^{K} g_{i_k,i_{-k}} = 1$. When $i_1 = \cdots = i_m = i$ we have $g_{i_k,i_{-k}} = g_{i_1,i_{-1}}$ for each $k \leq m$. Thus $m \cdot g_{i_1,i_{-1}} =$

$1-\sum_{\ell>m} g_{i_\ell,i_{-\ell}}$. Summing over $i_{m+1},\dots,i_K$, we then have

$$\sum_{\phi:\ i_1=\cdots=i_m=i} m\cdot g_{i_1,i_{-1}}\prod_{k>m} v_t^{i_k} = \left(\sum_{\phi:\ i_1=\cdots=i_m=i}\prod_{k>m} v_t^{i_k}\right) - \left(\sum_{\phi:\ i_1=\cdots=i_m=i}\sum_{\ell>m} g_{i_\ell,i_{-\ell}}\prod_{k>m} v_t^{i_k}\right)$$
$$= (V_t)^{K-m} - \sum_{\phi:\ i_1=\cdots=i_m=i}(K-m)g_{i_{m+1},i_{-(m+1)}}\prod_{k>m} v_t^{i_k},$$

where the second equality holds because $i_{m+1},\dots,i_K$ are unrestricted and symmetric.

Plugging this back into (A.9), we deduce

$$\begin{aligned}
&Kp_{t+1}^i\sum_{1\in A}\left(\sum_{\phi:\ i_1=i\&i_k=j\ \forall k\in A} g_{i_1,i_{-1}}\prod_{k\notin A} v_t^{i_k}\right)\\
=&\left(p_{t+1}^i\sum_{m=1}^{K}\binom{K}{m}(V_t)^K\right) - p_{t+1}^i\sum_{m=1}^{K}(K-m)\binom{K}{m}\sum_{\phi:\ i_1=\cdots=i_m=i} g_{i_{m+1},i_{-(m+1)}}\prod_{k>m} v_t^{i_k}\\
=&p_{t+1}^i\left((V_t+1)^K-(V_t)^K\right) - p_{t+1}^i\sum_{m=1}^{K-1}K\binom{K-1}{m}\sum_{\phi:\ i_1=\cdots=i_m=i} g_{i_{m+1},i_{-(m+1)}}\prod_{k>m} v_t^{i_k}.
\end{aligned}\tag{A.10}$$

Next we study the RHS of (A.8) for $j=i$ and $1\notin A$. In this case A is a non-empty subset of $\{2,\dots,K\}$, so there are $\binom{K-1}{m}$ such sets A with size m ($1\le m\le K-1$). By symmetry, for each such set A the sum $\sum_{\phi:\ i_1=i\&i_k=i\ \forall k\in A} g_{i_1,i_{-1}}\prod_{k\notin A} v_t^{i_k}$ is the same as $A=\{2,\dots,m+1\}$, and is equal to $\sum_{\phi:\ i_1=\cdots=i_{m+1}=i} g_{i_1,i_{-1}}\prod_{k\notin\{2,\dots,m+1\}} v_t^{i_k}$. This is also equal to $\sum_{\phi:\ i_1=\cdots=i_{m+1}=i} g_{i_{m+1},i_{-(m+1)}}\prod_{k\notin\{1,\dots,m\}} v_t^{i_k}$ given that $i_1=i_{m+1}=i$. Summing across m,

$$\begin{aligned}
&Kp_{t+1}^i\sum_{1\notin A, A\neq\emptyset}\left(\sum_{\phi:\ i_1=i\&i_k=i\ \forall k\in A} g_{i_1,i_{-1}}\prod_{k\notin A} v_t^{i_k}\right)\\
=&p_{t+1}^i\sum_{m=1}^{K-1}K\binom{K-1}{m}\sum_{\phi:\ i_1=\cdots=i_{m+1}=i} g_{i_{m+1},i_{-(m+1)}}\prod_{k>m} v_t^{i_k}.
\end{aligned}\tag{A.11}$$

Summing (A.10) and (A.11), we obtain the total contribution of $j=i$ to the RHS of (A.8):

$$\begin{aligned}
&Kp_{t+1}^i\sum_{A\neq\emptyset}\left(\sum_{\phi:\ i_1=i\&i_k=i\ \forall k\in A} g_{i_1,i_{-1}}\prod_{k\notin A} v_t^{i_k}\right)\\
=&p_{t+1}^i\left((V_t+1)^K-(V_t)^K\right) - p_{t+1}^i\sum_{m=1}^{K-1}K\binom{K-1}{m}\left(\sum_{\phi:\ i_1=\cdots=i_m=i\&i_{m+1}\neq i} g_{i_{m+1},i_{-(m+1)}}\prod_{k>m} v_t^{i_k}\right).
\end{aligned}\tag{A.12}$$

The inner summation in the second term above is now over samples that satisfy $i_1=\cdots=i_m=i$ and i_{m+1} *different from* i. These samples are precisely the difference between those considered in the second

term on the RHS of (A.10), relative to those covered by the RHS of (A.11). By considering possible values of i_{-m+1}, we can further write

$$\begin{aligned}\sum_{\phi:\ i_1=\cdots=i_m=i\&i_{m+1}\neq i} g_{i_{m+1},i_{-(m+1)}} \prod_{k>m} v_t^{i_k} &= \sum_{j\neq i} \sum_{\phi:\ i_1=\cdots=i_m=i\&i_{m+1}=j} g_{i_{m+1},i_{-(m+1)}} v_t^j \prod_{k>m+1} v_t^{i_k} \\ &= (V_t)^{K-m} \sum_{j\neq i} \sum_{\phi:\ i_1=\cdots=i_m=i\&i_{m+1}=j} g_{i_{m+1},i_{-(m+1)}} r_t^j \prod_{k>m+1} r_t^{i_k}.\end{aligned}$$

Plugging into (A.12) and interchanging the order of summation yields

$$\begin{aligned}&Kp_{t+1}^i \sum_{A\neq\emptyset} \left(\sum_{\phi:\ i_1=i\&i_k=i\ \forall k\in A} g_{i_1,i_{-1}} \prod_{k\notin A} v_t^{i_k} \right) \\ =&p_{t+1}^i \left((V_t+1)^K - (V_t)^K\right) - \sum_{j\neq i} p_{t+1}^i r_t^j \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \left(\sum_{\phi:\ i_1=\cdots=i_m=i\&i_{m+1}=j} g_{i_{m+1},i_{-(m+1)}} \prod_{k>m+1} r_t^{i_k} \right) \\ =&p_{t+1}^i \left((V_t+1)^K - (V_t)^K\right) - \sum_{j\neq i} p_{t+1}^i r_t^j \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \left(\sum_{\phi:\ i_1=j\&i_2=\cdots=i_{m+1}=i} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right),\end{aligned} \tag{A.13}$$

where the last step "swaps" i_{m+1} with i_1.

Finally we study the contribution of $j \neq i$ to the RHS of (A.8). In this case we only need to consider those sets A that do not contain 1, for otherwise $i_1 = i$ is inconsistent with $i_k = j$ for $k \in A$, and the inner sum would be zero. For each $1 \leq m \leq K-1$, there are exactly $\binom{K-1}{m}$ such sets A with size m. By symmetry, for each such set A the sum $\sum_{\phi:\ i_1=i\&i_k=j\ \forall k\in A} g_{i_1,i_{-1}} \prod_{k\notin A} v_t^{i_k}$ is the same as $A = \{2,\ldots,m+1\}$, and is equal to $\sum_{\phi:\ i_1=i\&i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k\notin\{2,\ldots,m+1\}} v_t^{i_k} = \sum_{\phi:\ i_1=i\&i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} v_t^i \prod_{k>m+1} v_t^{i_k}$,

Summing across m yields that for each $j \neq i$,

$$\begin{aligned}&Kp_{t+1}^j \sum_{A\neq\emptyset} \left(\sum_{\phi:\ i_1=i\&i_k=j\ \forall k\in A} g_{i_1,i_{-1}} \prod_{k\notin A} v_t^{i_k} \right) \\ =&p_{t+1}^j v_t^i \sum_{m=1}^{K-1} K \binom{K-1}{m} \left(\sum_{\phi:\ i_1=i\&i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k>m+1} v_t^{i_k} \right) \\ =&p_{t+1}^j r_t^i \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \left(\sum_{\phi:\ i_1=i\&i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right).\end{aligned} \tag{A.14}$$

Observe that the RHS here is very much similar to the second term on the RHS of (A.13). The difference is that "i" and "j" are swapped everywhere: $p_{t+1}^i r_t^j$ changes now to $p_{t+1}^j r_t^i$, and $\sum_{\phi:\ i_1=j\&i_2=\cdots=i_{m+1}=i}$ changes to $\sum_{\phi:\ i_1=i\&i_2=\cdots=i_{m+1}=j}$. This similarity is key to the subsequent argument.

We can now plug both (A.13) and (A.14) back into (A.8) to deduce

$$
\begin{aligned}
&(V_t+1)^K \cdot \mathbb{E}[p_{t+2}^i \mid r_t] - (V_t)^K \cdot p_{t+1}^i \\
=&p_{t+1}^i \left((V_t+1)^K - (V_t)^K\right) - \sum_{j\neq i} p_{t+1}^i r_t^j \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \left(\sum_{\phi:\ i_1=j \& i_2=\cdots=i_{m+1}=i} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right) \\
&+ \sum_{j\neq i} p_{t+1}^j r_t^i \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \left(\sum_{\phi:\ i_1=i \& i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right)
\end{aligned}
$$

Rearranging then gives

$$
\begin{aligned}
&(V_t+1)^K \cdot \left(\mathbb{E}[p_{t+2}^i \mid r_t] - p_{t+1}^i\right) \\
=&-\sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \sum_{j\neq i} p_{t+1}^i r_t^j \left(\sum_{\phi:\ i_1=j \& i_2=\cdots=i_{m+1}=i} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right) \\
&+ \sum_{m=1}^{K-1} K \binom{K-1}{m} (V_t)^{K-m} \sum_{j\neq i} p_{t+1}^j r_t^i \left(\sum_{\phi:\ i_1=i \& i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right).
\end{aligned} \tag{A.15}
$$

We now define

$$
\begin{aligned}
H_m^i(r_t^1,\ldots,r_t^N) =& K \binom{K-1}{m} \sum_{j\neq i} p_{t+1}^j r_t^i \left(\sum_{\phi:\ i_1=i \& i_2=\cdots=i_{m+1}=j} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right) \\
&- K \binom{K-1}{m} \sum_{j\neq i} p_{t+1}^i r_t^j \left(\sum_{\phi:\ i_1=j \& i_2=\cdots=i_{m+1}=i} g_{i_1,i_{-1}} \prod_{k>m+1} r_t^{i_k} \right),
\end{aligned} \tag{A.16}
$$

which is a homogeneous polynomial with degree $2K-m$ if we think of p_{t+1}^i and p_{t+1}^j as polynomials of $r_t^1,\ldots,r_t^N$ with degree K. Then (A.15) translates into

$$
(V_t+1)^K \cdot \left(\mathbb{E}[p_{t+2}^i \mid r_t] - p_{t+1}^i\right) = \sum_{m=1}^{K-1} (V_t)^{K-m} \cdot H_m^i(r_t^1,\ldots,r_t^N).
$$

If we sum across those "top" sellers $i \leq I$, then

$$
(V_t+1)^K \cdot \left(\sum_{i\leq I} \mathbb{E}[p_{t+2}^i \mid r_t] - \sum_{i\leq I} p_{t+1}^i \right) = \sum_{m=1}^{K-1} (V_t)^{K-m} \cdot \sum_{i\leq I} H_m^i(r_t^1,\ldots,r_t^N).
$$

Taking another expectation yields the desired recursion in the lemma:

$$
\sum_{i\leq I} \mathbb{E}[p_{t+2}^i] - \sum_{i\leq I} \mathbb{E}[p_{t+1}^i] = \sum_{m=1}^{K-1} \frac{(V_t)^{K-m}}{(V_t+1)^K} \cdot \sum_{i\leq I} H_m^i(r_t^1,\ldots,r_t^N).
$$

To complete the proof of the lemma it remains to show that for each $1 \leq m \leq K-1$ and $1 \leq I \leq N-1$, the polynomial $\sum_{i \leq I} H_m^i(r_t^1, \ldots, r_t^N)$ has non-negative (and not all zero) coefficients. From (A.16) we have

$$\begin{aligned}\sum_{i \leq I} H_m^i(r_t^1, \ldots, r_t^N) =& K\binom{K-1}{m} \sum_{i \leq I < j} p_{t+1}^j r_t^i \left(\sum_{\phi:\ i_1 = i \& i_2 = \cdots = i_{m+1} = j} g_{i_1, i_{-1}} \prod_{k > m+1} r_t^{i_k} \right) \\ &- K\binom{K-1}{m} \sum_{i \leq I < j} p_{t+1}^i r_t^j \left(\sum_{\phi:\ i_1 = j \& i_2 = \cdots = i_{m+1} = i} g_{i_1, i_{-1}} \prod_{k > m+1} r_t^{i_k} \right).\end{aligned} \tag{A.17}$$

It thus suffices to show that for any pair of sellers $i < j$, the polynomial

$$p_{t+1}^j r_t^i \left(\sum_{\phi:\ i_1 = i \& i_2 = \cdots = i_{m+1} = j} g_{i_1, i_{-1}} \prod_{k > m+1} r_t^{i_k} \right) - p_{t+1}^i r_t^j \left(\sum_{\phi:\ i_1 = j \& i_2 = \cdots = i_{m+1} = i} g_{i_1, i_{-1}} \prod_{k > m+1} r_t^{i_k} \right)$$

has non-negative coefficients. In fact we can prove a stronger statement, which is that for fixed $i < j$, fixed m and fixed $i_{m+2}, \ldots, i_K$, the polynomial

$$p_{t+1}^j r_t^i \cdot g_{i,j,\ldots,j,i_{m+2},\ldots,i_K} - p_{t+1}^i r_t^j \cdot g_{j,i,\ldots,i,i_{m+2},\ldots,i_K} \tag{A.18}$$

has non-negative coefficients (ignoring the common multiplier $\prod_{k>m+1} r_t^{i_k}$). Above, $g_{i,j,\ldots,j,i_{m+2},\ldots,i_K}$ is the choice probability of seller i in a sample consisting of i, m copies of j and the remaining $i_{m+2}, \ldots, i_K$. Likewise for $g_{j,i,\ldots,i,i_{m+2},\ldots,i_K}$.

Recall that

$$\begin{aligned} p_{t+1}^i &= K \sum_{j_1 = i} g_{j_1, j_{-1}} \prod_{k=1}^{K} r_t^{j_k} = K r_t^i \sum_{j_2, \ldots, j_K} g_{i, j_{-1}} \prod_{k>1} r_t^{j_k} \\ p_{t+1}^i &= K \sum_{j_1 = i} g_{j_1, j_{-1}} \prod_{k=1}^{K} r_t^{j_k} = K r_t^j \sum_{j_2, \ldots, j_K} g_{j, j_{-1}} \prod_{k>1} r_t^{j_k} \end{aligned}$$

Thus the difference in (A.18) can be written as

$$K r_t^i r_t^j \sum_{j_2, \ldots, j_K} \prod_{k>1} r_t^{j_k} \cdot (g_{j,j_{-1}} \cdot g_{i,j,\ldots,j,i_{m+2},\ldots,i_K} - g_{i,j_{-1}} \cdot g_{j,i,\ldots,i,i_{m+2},\ldots,i_K})$$

The final step of the proof is to show that under logit choice,

$$g_{j,j_2,\ldots,j_K} \cdot g_{i,j,\ldots,j,i_{m+2},\ldots,i_K} \geq g_{i,j_2,\ldots,j_K} \cdot g_{j,i,\ldots,i,i_{m+2},\ldots,i_K} \tag{A.19}$$

for any $i < j$, any $m \geq 1$ and any $i_{m+2}, \ldots, i_K, j_2, \ldots, j_K$. Indeed, by the logit choice rule we have

$$\frac{g_{i,j_2,\ldots,j_K}}{g_{j,j_2,\ldots,j_K}} = \frac{e^{\alpha q^i}/(e^{\alpha q^i} + \sum_{k>1} e^{\alpha q^{j_k}})}{e^{\alpha q^j}/(e^{\alpha q^j} + \sum_{k>1} e^{\alpha q^{j_k}})} < e^{\alpha(q^i - q^j)}$$

because $e^{\alpha q^i} + \sum_{k>1} e^{\alpha q^{j_k}} > e^{\alpha q^j} + \sum_{k>1} e^{\alpha q^{j_k}}$ when $q^i > q^j$. On the other hand,

$$\frac{g_{i,j,\ldots,j,i_{m+2},\ldots,i_K}}{g_{j,i,\ldots,i,i_{m+2},\ldots,i_K}} = \frac{e^{\alpha q^i}/(e^{\alpha q^i} + me^{\alpha q^j} + \sum_{k>m+1} e^{\alpha q^{i_k}})}{e^{\alpha q^j}/(e^{\alpha q^j} + me^{\alpha q^i} + \sum_{k>m+1} e^{\alpha q^{i_k}})} \geq e^{\alpha(q^i - q^j)}$$

because $e^{\alpha q^i} + me^{\alpha q^j} + \sum_{k>m+1} e^{\alpha q^{i_k}} \leq e^{\alpha q^j} + me^{\alpha q^i} + \sum_{k>m+1} e^{\alpha q^{i_k}}$ when $q^i > q^j$ and $m \geq 1$.

Therefore, under any logit choice rule, the inequality (A.19) holds strictly. It follows that the polynomial in (A.18), and thus the polynomial $\sum_{i \leq I} H_m^i$ in (A.17), has non-negative (and not all zero) coefficients. This proves the lemma and Proposition 2. ■

A.3 Proof of Proposition 3

For the model with learning, we first show that long run efficiency is achieved *only if* consumers "experiment." We do this by considering the "max choice" rule which chooses the highest quality seller in the sample with probability 1. This corresponds to the limiting case of logit choice, as $\alpha \to \infty$. We show below that Proposition 3 does *not* hold under max choice.

Consider a profile of true qualities $q^1 > \cdots > q^N$, but suppose in the first N periods each seller is chosen exactly once, with signals leading to posterior beliefs ranked in the opposite way: $\widehat{q_N^1} < \cdots < \widehat{q_N^N} < q^N$, which occurs with positive probability. Then under max choice, so long as the belief $\widehat{q_t^N}$ about seller N remains above $\widehat{q_N^N}$ at every $t \geq N$, the lowest quality seller N will continue to be chosen in period $t+1$ (for each $t \geq N$) with probability $1 - (1 - r_t^N)^K = 1 - (\frac{V_0 - v_0^N}{V_0 + t})^K$.

Because the signals about seller N's quality are i.i.d. with mean $q^N > \widehat{q_N^N}$, it is not difficult to show (using arguments similar to the proof of the Law of Large Numbers) that there is positive probability c that $\widehat{q_t^N} \geq \widehat{q_N^N}$ for every $t \geq N$. Hence the probability that seller N is chosen in *every* period after period N is at least

$$c \cdot \prod_{t \geq 0} \left(1 - (\frac{V_0 - v_0^N}{V_0 + t})^K\right).$$

This probability is strictly positive because $\sum_{t \geq N} (\frac{V_0 - v_0^N}{V_0 + t})^K$ is finite for every $K \geq 2$. Hence there is positive probability that seller N gets predominant market share in the long run.

As discussed in the main text, this kind of inefficiency can persist under max choice because max choice does not "experiment" with sellers whose current expected qualities are inferior (but whose true qualities may be very high). Without sufficient experimentation, beliefs about true qualities do not necessarily converge to the truth, making it possible for a lower quality seller to be perceived to have higher quality even in the long run.

We now turn to the proof of Proposition 3, demonstrating that sufficient experimentation is enough

to guarantee long run efficiency. We first show that along almost every history, for each seller i, either v_t^i remains bounded as $t \to \infty$ (i.e. seller i is only chosen finitely many times), or $\widehat{q_t^i}$ converges to q^i (i.e., beliefs are correct in the limit). Indeed, let ω denote a point in the probability space, which represents a generic infinite history of $\{v_t^i, \widehat{q_t^i}\}$. For each seller i define a sequence of random variables $Y_1^i(\omega), Y_2^i(\omega), \ldots$, where $Y_m^i(\omega)$ equals $Z_t = q^i + \epsilon_t$ where t is the period in which seller i is chosen for the m-th time.[45] If such a period t does not exist then let $Y_m^i(\omega)$ be drawn randomly from $q^i + \mathcal{N}(0, \frac{1}{\tau})$. By construction, $Y_1^i, Y_2^i, \ldots$ are i.i.d. random variables with the distribution $q^i + \mathcal{N}(0, \frac{1}{\tau})$, so by the Strong Law of Large Numbers we have

$$\lim_{m\to\infty} \frac{1}{m}(Y_1^i + \cdots + Y_m^i) = q^i \quad \text{for almost every } \omega.$$

Now observe that if t is the period in which seller i is chosen for the m-th time, then the belief at the end of period t is

$$\widehat{q_t^i} = \frac{\tau_0 \cdot \widehat{q_0} \cdot + \tau(Y_1^i + \cdots + Y_m^i)}{\tau_0 + m\tau},$$

which is approximately equal to $\frac{1}{m}(Y_1^i + \cdots + Y_m^i)$ for m large. Hence, along any history where the number m is unbounded, beliefs $\widehat{q_t^i}$ converge to the truth q^i, as we desire to show.

A corollary of the previous result is that along almost every history, all beliefs $\{\widehat{q_t^i}\}_{i,t}$ are bounded. Indeed, if seller i is only chosen finitely many times, then $\widehat{q_t^i}$ is eventually constant. And if seller i is chosen infinitely often, then $\widehat{q_t^i}$ converges to q^i. In either case, a convergent sequence is necessarily bounded.

We now use this boundedness property to show that, *under the logit choice rule*, each seller i is chosen infinitely often along almost every history. To do this, we fix a seller i and a pair of positive integers M and T, and show *there is zero probability that seller i is not chosen after the first T periods and that $|\widehat{q_t^j}| \leq M$ for all j, t.* Once this is proved, taking the countable union over i, M, T then implies there is zero probability that some seller is not chosen infinitely often and that all beliefs are bounded. Hence there will be zero probability that some seller is not chosen infinitely often, period.

To establish the preceding italicized claim, we fix any history in the first T periods, and show there is zero conditional probability that seller i is not chosen afterwards and that all future beliefs belong to $[-M, M]$. This conditional probability can be written as the product of conditional probabilities below:

$$\prod_{t\geq T} \mathbb{P}\{\text{seller } i \text{ not chosen in period } t+1 \ \& \ |\widehat{q_{t+1}^j}| \leq M \ \forall j \quad | \quad v_t^i = v_T^i \ \& \ |\widehat{q_\tau^j}| \leq M \ \forall j \ \forall \tau \leq t\}.$$

For each $t \geq T$, we condition on the event that seller i is not chosen from period $T+1$ to period t. Thus in this event, the belief $\widehat{q_t^i}$ is the same as the earlier belief $\widehat{q_T^i}$, which has been fixed. On the other hand, beliefs about all sellers are bounded above by M. Thus the choice probability of (any copy of) seller i out of a sample of size K is at least $\eta = \frac{e^{\alpha \widehat{q_T^i}}}{Ke^{\alpha M}}$, which is a strictly positive constant. Because the expected number of seller i in the sample is Kr_t^i, it follows that the total conditional probability that

[45] Formally, t is the smallest time such that $s_t^i = m$.

seller i will be chosen in period $t+1$ is at least $\eta K r_t^i$, which in turn is bounded below by $\frac{\eta K v_0^i}{V_0+t}$. Thus the conditional probability that seller i is not chosen in period $t+1$ (and beliefs remain bounded by M) is at most $1-\frac{\eta K v_0^i}{V_0+t}$. Since $\sum_{t\geq T}\frac{\eta K v_0^i}{V_0+t}=\infty$, we have $\prod_t(1-\frac{\eta K v_0^i}{V_0+t})=0$. So there is zero probability that seller i is not chosen after period T and all beliefs are bounded by M. This proves our earlier claim.

We now know that almost surely $v_t^i\to\infty$ for each seller i, and thus $\widehat{q_t^i}\to q^i$. In what follows we first show that the relative visibility of seller 1, r_t^1, almost surely converges. We will then show the limit must be 1 almost surely. To prove the convergence of r_t^1, we will again construct a sub-martingale. However, unlike before, r_t^1 itself need not be a sub-martingale because the ranking of expected qualities $\widehat{q_t^i}$ may not coincide with the ranking of true qualities q^i (so that seller 1 may not be "favored" in the choice stage). To address this issue, the key insight is that when $\widehat{q_t^i}$ are not ranked in the same order as the true qualities q^i, it must be the case that for some i, the distance between the belief $\widehat{q_t^i}$ and the truth q^i is relatively big. The signal produced in the next period will tend to reduce this distance, motivating the following construction:

$$X_t=(r_t^1)^{K-1}-C\sum_{i=1}^{n}\left((\widehat{q_t^i}-q^i)^2+\frac{1}{s_t^i\tau+\tau_0}\right), \tag{A.20}$$

where C is a large positive constant to be determined later. As we show below, the term $(\widehat{q_t^i}-q^i)^2+\frac{1}{s_t^i\tau+\tau_0}$ decreases (in expectation) with learning, and this decrease is larger in magnitude than possible decrease in $(r_t^1)^{K-1}$, making X_t a sub-martingale. The $(K-1)$-th power applied to r_t^1 turns out to be important for our argument, although any larger exponent would also work.[46]

To check X_t is a sub-martingale, we first compute $\mathbb{E}[(\widehat{q_{t+1}^i}-q^i)^2+\frac{1}{s_{t+1}^i\tau+\tau_0}\mid\mathcal{F}_t]$, where we use $\mathcal{F}_t$ to denote the natural filtration containing the history in the first t periods. With probability $1-p_{t+1}^i$ seller i is not chosen in period $t+1$, so the expression $(\widehat{q_{t+1}^i}-q^i)^2+\frac{1}{s_t^i\tau+\tau_0}$ is the same as in period t. With probability p_{t+1}^i however, s_{t+1}^i increases to s_t^i+1, and

$$\widehat{q_{t+1}^i}=\frac{\left(s_t^i\tau+\tau_0\right)\cdot\widehat{q_t^i}+\tau\cdot(q^i+\epsilon_{t+1})}{\left(s_t^i\tau+\tau_0\right)+\tau},$$

and so

$$\widehat{q_{t+1}^i}-q^i=\frac{s_t^i\tau+\tau_0}{(s_t^i+1)\tau+\tau_0}\cdot(\widehat{q_t^i}-q^i)+\frac{\tau}{(s_t^i+1)\tau+\tau_0}\cdot\epsilon_{t+1}.$$

Since $\epsilon_{t+1}\sim\mathcal{N}(0,\frac{1}{\tau})$ is independent of $\mathcal{F}_t$, the expectation of $(\widehat{q_{t+1}^i}-q^i)^2$ in this case is

$$\left(\frac{s_t^i\tau+\tau_0}{(s_t^i+1)\tau+\tau_0}\right)^2\cdot(\widehat{q_t^i}-q^i)^2+\left(\frac{\tau}{(s_t^i+1)\tau+\tau_0}\right)^2\cdot\frac{1}{\tau}.$$

[46] Intuitively, applying the a-th power for $a>1$ is a convex transformation that would maintain the sub-martingale property.

Summarizing the above, we have

$$
\begin{aligned}
&\mathbb{E}[(\widehat{q^i_{t+1}} - q^i)^2 + \frac{1}{s^i_{t+1}\tau + \tau_0} \mid \mathcal{F}_t] - (\widehat{q^i_t} - q^i)^2 - \frac{1}{s^i_t\tau + \tau_0} \\
=& - p^i_{t+1} \cdot [\left(1 - \left(\frac{s^i_t\tau + \tau_0}{(s^i_t+1)\tau + \tau_0}\right)^2\right) \cdot (\widehat{q^i_t} - q^i)^2 \\
&\qquad + \frac{1}{s^i_t\tau + \tau_0} - \frac{1}{(s^i_t+1)\tau + \tau_0} - \left(\frac{\tau}{(s^i_t+1)\tau + \tau_0}\right)^2 \cdot \frac{1}{\tau}] \\
\leq& - p^i_{t+1} \cdot \left[\frac{\tau}{(s^i_t+1)\tau + \tau_0} \cdot (\widehat{q^i_t} - q^i)^2\right].
\end{aligned}
\tag{A.21}
$$

The last inequality above uses $1-\left(\frac{s^i_t\tau+\tau_0}{(s^i_t+1)\tau+\tau_0}\right)^2 \geq 1-\left(\frac{s^i_t\tau+\tau_0}{(s^i_t+1)\tau+\tau_0}\right) = \frac{\tau}{(s^i_t+1)\tau+\tau_0}$, and $\frac{1}{s^i_t\tau+\tau_0} - \frac{1}{(s^i_t+1)\tau+\tau_0} = \frac{\tau}{(s^i_t\tau+\tau_0)\cdot((s^i_t+1)\tau+\tau_0)} \geq \frac{\tau}{((s^i_t+1)\tau+\tau_0)^2}$.

Plugging (A.21) into (A.20), we obtain

$$
\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \geq \mathbb{E}[(r^1_{t+1})^{K-1} \mid \mathcal{F}_t] - (r^1_t)^{K-1} + \sum_{i=1}^{n} p^i_{t+1} \cdot \frac{C\tau}{(s^i_t+1)\tau + \tau_0} \cdot (\widehat{q^i_t} - q^i)^2. \tag{A.22}
$$

We next study the difference $\mathbb{E}[(r^1_{t+1})^{K-1} \mid \mathcal{F}_t] - (r^1_t)^{K-1}$. If $p^1_{t+1} \geq r^1_t$, then as we calculated before $\mathbb{E}[r^1_{t+1} \mid \mathcal{F}_t] - r^1_t = \frac{p^1_{t+1} - r^1_t}{V_t+1} \geq 0$. By Jensen's inequality, it follows that

$$
\mathbb{E}[(r^1_{t+1})^{K-1} \mid \mathcal{F}_t] \geq \left(\mathbb{E}[r^1_{t+1} \mid \mathcal{F}_t]\right)^{K-1} \geq (r^1_t)^{K-1}.
$$

In this case we immediately deduce from (A.22) that $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \geq 0$.

Suppose instead that $p^1_{t+1} < r^1_t$. Then the expected quality $\widehat{q^1_t}$ cannot be the highest across sellers (otherwise seller 1 is chosen with probability at least $\frac{1}{K}$ out of each sample, leading to total choice probability at least r^1_t). Thus we can let $j > 1$ be a seller such that $\widehat{q^j_t} \geq \widehat{q^1_t}$. Since $q^1 - q^j \geq q^1 - q^2$, either we have $\widehat{q^1_t} - q^1 \leq -\frac{q^1-q^2}{2}$, or $\widehat{q^j_t} - q^j \geq \frac{q^1-q^2}{2}$. In the former case we make use of the fact that $p^1_{t+1} \geq (r^1_t)^K$, which is the probability that the sample consists of K copies of seller 1. Thus

$$
\begin{aligned}
p^1_{t+1} \cdot \frac{C\tau}{(s^1_t+1)\tau + \tau_0} \cdot (\widehat{q^1_t} - q^1)^2 &\geq (r^1_t)^K \cdot \frac{C \cdot (q^1 - q^2)^2 \cdot \tau}{4(v^1_t\tau + \tau + \tau_0)} \\
&= \frac{(r^1_t)^{K-1}}{V_t} \cdot \frac{C(q^1 - q^2)^2}{4} \cdot \frac{v^1_t\tau}{v^1_t\tau + \tau + \tau_0} \\
&\geq \frac{(r^1_t)^{K-1}}{V_t} \cdot \frac{C(q^1 - q^2)^2}{4} \cdot \frac{v^1_0\tau}{v^1_0\tau + \tau + \tau_0}.
\end{aligned}
$$

If instead $\widehat{q^j_t} - q^j \geq \frac{q^1-q^2}{2}$, then we make use of the fact that $p^j_{t+1} \geq (r^1_t)^{K-1} r^j_t$, because $K(r^1_t)^{K-1} r^j_t$ is the probability of a sample consisting of $K-1$ copies of seller 1 and a copy of seller j, and out of this

sample seller j is chosen with probability at least $\frac{1}{K}$ because $\widehat{q_t^j} \geq \widehat{q_t^1}$. Hence in this case we have

$$\begin{aligned} p_{t+1}^j \cdot \frac{C\tau}{(s_t^j+1)\tau+\tau_0} \cdot (\widehat{q_t^j} - q^j)^2 &\geq (r_t^1)^{K-1} r_t^j \cdot \frac{C \cdot (q^1-q^2)^2 \cdot \tau}{4(v_t^j\tau+\tau+\tau_0)} \\ &= \frac{(r_t^1)^{K-1}}{V_t} \cdot \frac{C(q^1-q^2)^2}{4} \cdot \frac{v_t^j\tau}{v_t^j\tau+\tau+\tau_0} \\ &\geq \frac{(r_t^1)^{K-1}}{V_t} \cdot \frac{C(q^1-q^2)^2}{4} \cdot \frac{v_0^j\tau}{v_0^j\tau+\tau+\tau_0}. \end{aligned}$$

Therefore, if we choose a sufficiently large constant C such that $\frac{C(q^1-q^2)^2}{4} \cdot \frac{v_0^i\tau}{v_0^i\tau+\tau+\tau_0} \geq K-1$ for every $1 \leq i \leq N$, then whenever $p_{t+1}^1 < r_t^1$ we would have

$$\sum_{i=1}^n p_{t+1}^i \cdot \frac{C\tau}{(s_t^i\tau+1)\tau+\tau_0} \cdot (\widehat{q_t^i} - q^i)^2 \geq \frac{(K-1)\cdot(r_t^1)^{K-1}}{V_t} > \frac{(K-1)\cdot(r_t^1)^{K-1}}{V_t+1}.$$

Now observe that we always have $r_{t+1}^1 \geq \frac{V_t}{V_t+1} \cdot r_t^1$. So

$$\mathbb{E}[(r_{t+1}^1)^{K-1} \mid \mathcal{F}_t] - (r_t^1)^{K-1} \geq \left((\frac{V_t}{V_t+1})^{K-1} - 1 \right) \cdot (r_t^1)^{K-1} \geq -\frac{K-1}{V_t+1} \cdot (r_t^1)^{K-1},$$

where the second inequality is based on Bernoulli's inequality $(1+x)^n \geq 1+nx$ (valid for all $x \geq -1$). Hence, combining the previous two estimates, we deduce from (A.22) that $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \geq 0$ also holds even if $p_{t+1}^1 < r_t^1$.

This analysis shows that the process $\{X_t\}$ is a sub-martingale. Since X_t is bounded above by 1, it follows that X_t converges almost surely to a limit random variable X_∞. But since we have already shown that almost surely $v_t^i \to \infty$ and $\widehat{q_t^i} \to q^i$, we can see from (A.20) that in fact $(r_t^1)^{K-1}$ converges almost surely to X_∞. It follows that r_t^1 converges almost surely to a limit random variable r_∞^1.

To complete the proof we argue that r_∞^1 is almost surely equal to 1. This part of the argument is very similar to the case with known qualities, modulo some additional details that we explain below. Specifically, suppose r_∞^1 is less than 1 with positive probability, then there exists $\epsilon > 0$ such that $\mathbb{P}\{r_\infty^1 < 1-\epsilon\} > \epsilon$. Fixing this ϵ, we can choose T such that for every $t \geq T$, $\mathbb{P}\{r_t^1 < 1-\epsilon\} > \epsilon$. Next recall that we have shown that beliefs $\widehat{q_t^i}$ converge almost surely to the true qualities q^i. Thus the belief differences $\widehat{q_t^1} - \widehat{q_t^j}$ between seller 1 and another seller $j > 1$ have an almost surely positive limit $q^i - q^j$. Again since almost sure convergence implies convergence in probability, the probability that all these differences $\widehat{q_t^1} - \widehat{q_t^j}$ exceeds $\frac{q^1-q^2}{2}$ converges to 1 as $t \to \infty$. So for any $\delta > 0$, we can make T even larger to ensure that $\mathbb{P}\{\widehat{q_t^1} - \widehat{q_t^j} > \frac{q^1-q^2}{2} \ \forall j > 1\} > 1-\delta$ for every $t \geq T$.

For these fixed ϵ (chosen above), δ (determined below), and T (chosen based on ϵ, δ), we consider an arbitrary period $t \geq T$. With probability at least $1-\delta$ the belief of seller 1's quality is highest, and in fact exceeds the second highest belief by a margin of $\frac{q^1-q^2}{2}$. Thus in any sample of K sellers where seller 1 appears k times, the total probability of choosing seller 1 is at least $\frac{k}{k+(K-k)\cdot e^{-\alpha\frac{q^1-q^2}{2}}}$, which is

in turn larger than $\frac{k(1+\eta)}{K}$ for some $\eta > 0$ that depends only on $q^1 - q^2, \alpha, K$. In this event our previous calculations in Equation (A.3) imply that $p^1_{t+1} \geq r^1_t(1 + \eta(1 - r^1_t))$. In particular $p^1_{t+1} \geq r^1_t$, and (A.2) gives that 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 non-negative with probability $1 - \delta$.

Moreover, in this event with probability $1-\delta$, there is a sub-event with probability at least $\epsilon - \delta$ such that $1 - r^1_t > \epsilon$. In this case $p^1_{t+1} \geq r^1_t(1 + \eta\epsilon)$ and (A.2) gives the stronger estimate $\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t}) \geq \frac{\eta\epsilon}{V_t}$. On the other hand, in the complementary event with probability δ we may not have $p^1_{t+1} \geq r^1_t$, but in any case (A.2) gives $\mathbb{E}[\log(r^1_{t+1}) - \frac{1}{v^1_{t+1}} \mid r^1_t] - (\log(r^1_t) - \frac{1}{v^1_t}) \geq \frac{-1}{V_t}$. Putting it together, we deduce

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

Therefore, if we initially choose δ to be so small that $c = (\epsilon - \delta)\eta\epsilon - \delta > 0$, then for all sufficiently large t it holds that

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

Telescoping this inequality leads to the same contradiction as in the proof of Proposition 1. Hence r^1_∞ must be equal to 1 almost surely, completing the proof.

A.4 Short Run Welfare with Learning

In this appendix we derive an analogue of Proposition 2 for the model with learning. Specifically, we consider N sellers with equal initial visibility level v_0, as well as equal prior mean $\widehat{q_0} = 0$ and prior variance $\sigma_0^2 = \frac{1}{\tau_0}$ about their true qualities $\{q^i\}$. Each purchase from seller i produces an independent normal signal about q^i with signal variance $\sigma^2 = \frac{1}{\tau}$. We are interested in how an increase in the number of sellers affects consumer welfare in earlier periods.

A subtle issue is that, unlike the baseline setting with known qualities, here it is no longer possible to study an increase in N by relating it to an equivalent increase in initial visibility v_0. The reason is that with uncertainty, the model with $2N$ sellers is not the replication of the model with N sellers due to qualities being independently drawn. Thus, the following result states the welfare comparison directly in terms of an increase in N:

Proposition 4. *Fix $v_0, K, \alpha, \tau_0, \tau$ in the model with learning. Then for every positive integer T there exists $\underline{N}(T)$ such that whenever $N \geq \underline{N}(T)$, expected quality received by the consumer in each of the periods $2 \sim T$ decreases with N.*

Proof. It suffices to study the expected quality in period T. To compute this, we adopt the subjective perspective, and average the period $T - 1$ belief of period T seller's quality, across different histories. This approach leads to the correct answer due to the Law of Iterated Expectations.

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

- the search sample $(i^1_t, \ldots, i^K_t)$ in each period $1 \leq t \leq T$;

- in each period $1 \leq t \leq T$ an index $k(t) \in \{1, \ldots, K\}$ 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 $Z_t = q^{i_t^{k(t)}} + \mathcal{N}(0, \frac{1}{\tau})$.

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.

A key observation is that the likelihood of any such history can be explicitly written as the following product:

$$L(\mathcal{H}) = \left(\prod_{t=1}^{T} \prod_{k=1}^{K} \frac{v_0 + s_{t-1}^{i_t^k}}{N v_0 + t - 1} \right) \cdot \left(\prod_{t=1}^{T-1} \frac{exp(\alpha \widehat{q_{t-1}^{i_t^{k(t)}}})}{\sum_{k=1}^{K} exp(\alpha \widehat{q_{t-1}^{i_t^k}})} \right) \cdot l\left(Z_1, \ldots, Z_{T-1} \mid \{i_t^{k(t)}\}_{t \leq T-1} \right). \quad \text{(A.23)}$$

The first part above captures the probability of generating each K-sample (based on initial visibility and sales). The second part is the probability of choosing the particular seller out of the sample (based on logit rule applied to beliefs). The last part, $l(Z_1, \ldots, Z_{t-1} \mid \cdot)$, represents the probability of seeing the signal realizations Z_t (based on the prior and normal signal likelihoods). These likelihoods are the weights we will 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

$$f(\mathcal{H}) = \sum_{j=1}^{K} \widehat{q_{T-1}^{i_T^j}} \cdot \frac{exp(\alpha \widehat{q_{T-1}^{i_T^j}})}{\sum_{k=1}^{K} exp(\alpha \widehat{q_{T-1}^{i_T^k}})}. \quad \text{(A.24)}$$

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 $f(\mathcal{H}) = 0$, so we can ignore such histories in computing the above integral.

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

The other $K-1$ sellers sampled in period T are all distinct from those previous $K \cdot (T-1)$, and from each other.

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

$$\frac{K(T-1)(v_0+1)\left(\prod_{s=0}^{KT-2}(N-s)v_0\right)}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$$

To understand this expression, note that in order 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 $\frac{(N-1)v_0}{Nv_0}$, i_1^3 with probability $\frac{(N-2)v_0}{Nv_0}$ and so on. Similar conditional probabilities apply to all sampled sellers before period T, as well as all but one of the sellers sampled in period T. The remaining term $\frac{K(T-1)(v_0+1)}{NV_0+T-1}$ 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.[47]

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 chosen once before period T, but is also part of the sample in period T. This signal $Z = q^i + \mathcal{N}(0, \frac{1}{\tau})$ leads to belief $\widehat{q^i_{T-1}} = \frac{\tau Z}{\tau_0+\tau}$ by Bayes rule. Since $q^i \sim \mathcal{N}(0, \frac{1}{\tau_0})$, it is easy to see that the distribution of the belief $\widehat{q^i_{T-1}}$ is normal with mean 0 and variance $(\frac{\tau}{\tau_0+\tau})^2 \cdot (\frac{1}{\tau_0} + \frac{1}{\tau}) = \frac{\tau}{\tau_0(\tau_0+\tau)}$. For the remaining $K-1$ sellers sampled in period T, their beliefs are zero as in the prior.

Thus, given the samples and given the "special" belief $\widehat{q} = \widehat{q^i_{T-1}}$, the believed quality in period T can be computed as

$$f(\mathcal{H}) = \widehat{q} \cdot \frac{exp(\alpha\widehat{q})}{K-1+exp(\alpha\widehat{q})}.$$

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

$$\eta = \mathbb{E}\left[\widehat{q} \cdot \frac{exp(\alpha\widehat{q})}{K-1+exp(\alpha\widehat{q})} \mid \widehat{q} \sim \mathcal{N}(0, \frac{\tau}{\tau_0(\tau_0+\tau)})\right] > 0.$$

This is positive intuitively because information, which does not hurt under random choice, must be strictly beneficial under the logit choice rule which favors higher quality. Formally, it can be proved by observing that $\widehat{q} \cdot \frac{exp(\alpha\widehat{q})}{K-1+exp(\alpha\widehat{q})} + (-\widehat{q}) \cdot \frac{exp(\alpha(-\widehat{q}))}{K-1+exp(\alpha(-\widehat{q}))} > 0$ whenever $\widehat{q} \neq 0$.

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

$$\eta \cdot \frac{K(T-1)(v_0+1)\left(\prod_{s=0}^{KT-2}(N-s)v_0\right)}{\prod_{t=1}^{T}(Nv_0+t-1)^K}.$$

[47] 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 the sellers sampled before period T are all distinct, and thus completely symmetric.

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

$$\frac{P(N)}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$$

for some polynomial P with degree $KT-1$ and positive leading coefficient.

(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 will show that the contribution of these histories to period T expected quality can be written as

$$\frac{Q(N)}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$$

for some polynomial Q with degree at most $KT-2$. Intuitively, this is because 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 relabelled accordingly. But due to ex-ante symmetry, the resulting set of histories contributes the same to 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 relabelled into each other, and then do a weighted sum with weights given by the number of relabellings associated with each collection.

The benefit of this approach is that modulo relabelling, we are essentially concerned with the patterns of repetition among KT sampled firms. The number of such patterns depends on K, T but not on N, and so does the number of collections that cannot be relabelled into each other (the latter number is K^{T-1} times bigger 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 relabellings 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 $\frac{c}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$ for some constant c independent of N, then the weighted sum of such contributions would have the desired form $\frac{Q(N)}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$ with $deg(Q) \leq KT-2$.

Now, for a fixed collection, we know that sales evolution has been determined. Thus the first part on the RHS of (A.23) is fixed, and has the form $\frac{c_1}{\prod_{t=1}^{T}(Nv_0+t-1)^K}$ for some constant c_1. Using (A.23) and (A.24), the contribution of this collection can be written as

$$\frac{c_1}{\prod_{t=1}^{T}(Nv_0+t-1)^K}\cdot\int_{Z_1,\ldots,Z_{T-1}}\left(\prod_{t=1}^{T-1}\frac{exp(\alpha\widehat{q_{t-1}^{i_t^{k(t)}}})}{\sum_{k=1}^{K}exp(\alpha\widehat{q_{t-1}^{i_t^k}})}\right)\cdot l\left(Z_1,\ldots,Z_{T-1}\right)\cdot\left(\sum_{j=1}^{K}\widehat{q_{T-1}^{i_T^j}}\cdot\frac{exp(\alpha\widehat{q_{T-1}^{i_T^j}})}{\sum_{k=1}^{K}exp(\alpha\widehat{q_{T-1}^{i_T^k}})}\right),$$

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 a finite constant c_2 independent of N, as we desire to show.[48]

To complete the proof of Proposition 4, we put together the 3 kinds of histories studied above. The previous analysis allows us to deduce that the expected quality in period T can be written as

$$\frac{R(N)}{S(N)},$$

where $R(N) = P(N) + Q(N)$ is a polynomial with degree $KT - 1$ and leading coefficient $\alpha > 0$, and $S(N) = \prod_{t=1}^{T}(Nv_0 + t - 1)^K$ is a polynomial with degree KT and leading coefficient $\beta > 0$. Thus, the derivative of this expected quality with respect to N is

$$\frac{R'(N)S(N) - S'(N)R(N)}{S(N)^2}.$$

The numerator $R'(N)S(N) - S'(N)R(N)$ is a polynomial with degree $2KT - 2$ and leading coefficient $-\alpha\beta < 0$. Hence for sufficiently large N, this derivative is negative. In other words, when N is sufficient large the expected quality in early periods decreases with N. ■

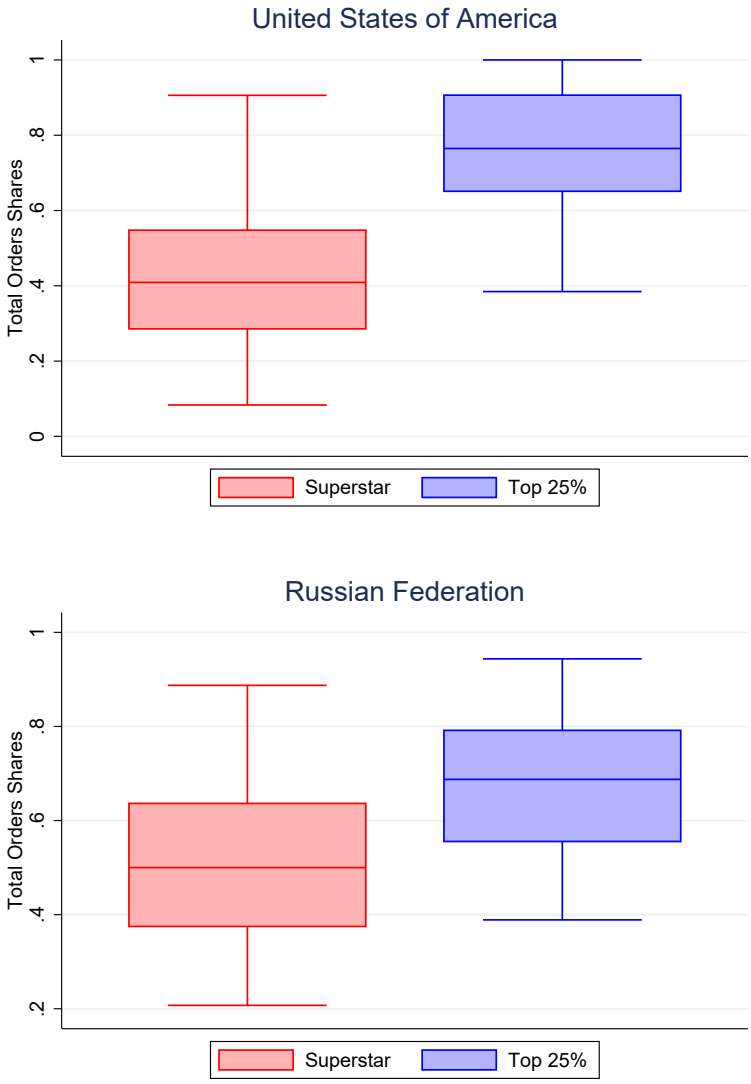
[48] To see that the integral is finite, we interpret it as the expectation of the following function of beliefs:

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

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

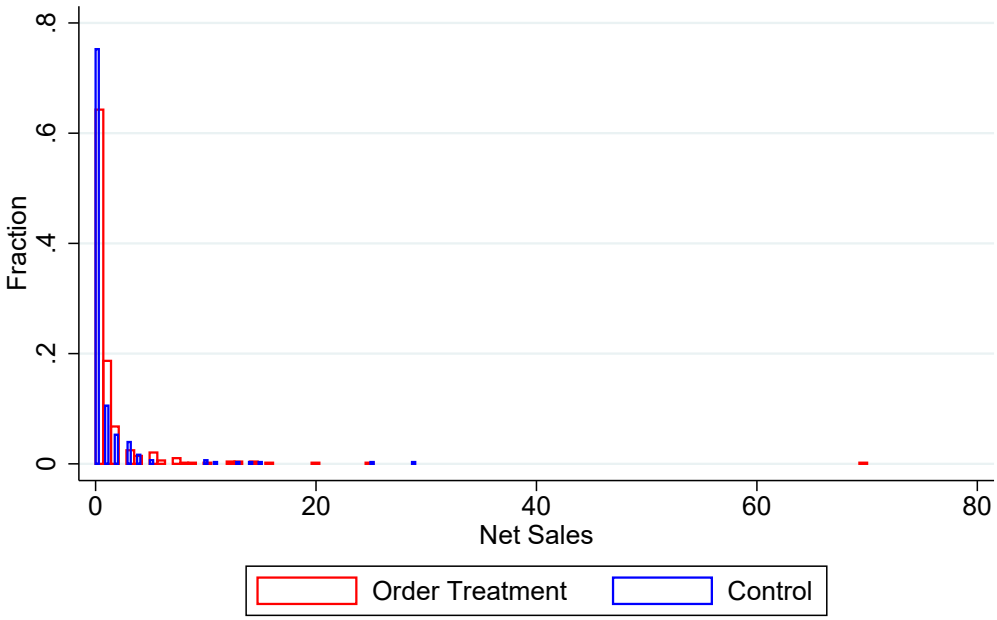
B Figures and Tables

Figure B.1: Six-Month Sales Performance Within Identical Variety By Country



Note: This figure plots the distribution of the total share of cumulative orders for top listings across groups by country. The number of cumulative orders is constructed from the transaction records from March to August 2018. "Superstar" indicates the listing that has the highest cumulative orders within its group for a given country. "Top 25%" indicates listings that have the top 25% cumulative orders within its group. We limit to groups with at least 4 listings and at least 1 with positive sales to a given country.

Figure B.2: Endline Net Sales Distribution



Note: This figure plots the endline distribution of cumulative orders (net of our own orders). The blue bars indicate the control group; the red bars indicate the treatment groups, i.e. T1 and T2.

Table B.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
LostPackage	0.0
NoPackageDamage	8.0
ShipDeliveryTimeLag	6.8
ServiceQualityIndex	**5.8**
ReplyWithinTwoDays	5.8

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

Table B.2: Correlation between Quality and Star Rating

	Dependent: Star Rating					
	(1)	(2)	(3)	(4)	(5)	(6)
ProductQualityIndex	0.029	0.170				
	(0.048)	(0.114)				
ShippingQualityIndex			0.082*	0.098*		
			(0.044)	(0.055)		
ServiceQualityIndex					0.034**	0.036*
					(0.017)	(0.020)
Constant	4.803***	4.804***	4.795***	4.793***	4.793***	4.793***
	(0.019)	(0.020)	(0.019)	(0.019)	(0.016)	(0.017)
Observations	409	409	422	422	624	624
Rsquare	0.001	0.317	0.008	0.319	0.006	0.210
Group FE	No	Yes	No	Yes	No	Yes

Note: This table presents results from regressing listing star ratings on their quality indices. Standard errors are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table B.3: The Dependence of New Orders on Current Cumulative Orders

Dummy=1 if having an order in the following week	(1)	(2)
Log Orders	0.092***	0.102***
	(0.001)	(0.002)
Constant	-0.012***	-0.032***
	(0.004)	(0.005)
Observations	15180	15180
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.

Table B.4: Summary Statistics of Listings In the Children's T-shirts Market on AliExpress: All Groups

	Observations	Mean	Std Dev	Median	5th Pctile	95th Pctile
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

Note: This table reports the same summary statistics as in Table 2 but containing all variety groups collected in May 2018.

Table B.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	5.95	5.46	5.63	-0.48	-0.32	0.16	1.28
	(4.10)	(2.57)	(3.72)	(0.29)	(0.35)	(0.29)	(0.26)
Cumulative Orders	0.90	0.73	0.81	-0.18*	-0.09	0.09	0.91
	(1.26)	(1.18)	(1.20)	(0.10)	(0.11)	(0.11)	(0.34)
Total Feedback	0.46	0.38	0.65	-0.08	0.19	0.27*	1.88
	(1.21)	(1.37)	(1.88)	(0.11)	(0.13)	(0.15)	(0.17)
Positive Rating Rate	0.95	0.96	0.91	0.01	-0.04	-0.05	0.83
	(0.21)	(0.17)	(0.28)	(0.04)	(0.04)	(0.05)	(0.36)
Free Shipping Dummy	0.50	0.45	0.48	-0.05	-0.02	0.03	0.21
	(0.50)	(0.50)	(0.50)	(0.04)	(0.04)	(0.05)	(0.65)
Shipping Price	0.74	0.76	0.69	0.02	-0.05	-0.06	0.25
	(1.01)	(0.89)	(1.03)	(0.08)	(0.09)	(0.09)	(0.61)

Note: This table checks whether the order and review treatments are correlated with listing characteristics collected prior to the treatment. The first three columns report the mean and standard deviation of the variables for each treatment group. Columns (4)-(6) show the difference between the two groups and the standard errors of the difference. The column heading indicates which groups are being compared. The last column tests whether the three treatment groups have the same mean. $***$: p ¡ 0.01; $**$: p ¡ 0.05; $*$: p ¡ 0.1.

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

	All Destinations		English-speaking		United States	
Order	0.023	0.026*	0.014**	0.015***	0.016***	0.017***
	(0.020)	(0.015)	(0.005)	(0.005)	(0.003)	(0.003)
ReviewXPostReview	0.005	-0.014	0.021	0.016	0.017	0.014
	(0.024)	(0.027)	(0.019)	(0.018)	(0.017)	(0.016)
Observations	10270	10270	10270	10270	10270	10270
Group FE	No	Yes	No	Yes	No	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	No	No	No	No	No	No
Clustered SE at listing level	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 endline number of cumulative orders, calculated using the transaction data collected in August 2018. The independent variable is the order treatment dummy that equals one for all products in the treatment groups (T1 and T2) and zero for the control group. Column 1 reports the average treatment effect, and Columns 2 to 6 report the quantile treatment effects. Standard errors clustered at listing level are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table B.7: Treatment Effects on Ranking

	(1)	(2)
	EnterFirst15Pages	
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	10270	10270
Group FE	No	Yes
Week FE	Yes	Yes
Baseline Controls	Yes	Yes
Clustered SE at listing level	Yes	Yes

Note: This table reports the treatment effects of the experimentally generated orders and reviews on listing ranks using the 13-week panel of the experiment sample. The dependent variable is a dummy variable that equals one if the listing enters the first 15 pages in the no-group search. 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 review on shipping and product quality. "PostReview" is a dummy that equals one for the weeks after the reviews were given (i.e., from week 7 onward). "MonthX" is a dummy variable that equals one for the X-th month after treatment. Standard errors clustered at listing level are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table B.8: Heterogeneous Treatment Effect: Quality

	(1)	(2)	(3)	(4)
Order	0.301	0.327*	0.479	0.797
	(0.236)	(0.180)	(0.392)	(0.635)
OrderXServiceQualityIndex	-0.157	-0.075		
	(0.228)	(0.203)		
ServiceQualityIndex	0.290**	0.093		
	(0.119)	(0.125)		
OrderXStdStar			-0.041	0.132
			(0.160)	(0.329)
StdStarRating			0.172**	0.050
			(0.068)	(0.179)
Constant	0.978***	0.889***	0.759	0.002
	(0.320)	(0.230)	(0.542)	(0.758)
Observations	784	784	168	168
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 total number of orders net of our own, measured using the transaction data collected in August 2018. 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. The standardized quality measures are constructed by standardizing individual quality metrics first and taking their average within each quality type. See 2.2 for details about the quality metrics. Standard errors are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

Table B.9: Seller Actions After Treatment

Panel A: Price

	AdjustPrice		CutPrice		RaisePrice		ΔLogPrice	
Treatment	0.026	0.020	0.046	0.048	0.030	0.021	0.001	-0.003
	(0.039)	(0.038)	(0.038)	(0.037)	(0.038)	(0.036)	(0.010)	(0.009)
Constant	0.456***	0.459***	0.333***	0.332***	0.345***	0.351***	0.005	0.007
	(0.031)	(0.030)	(0.030)	(0.029)	(0.030)	(0.028)	(0.008)	(0.007)
Observations	693	693	693	693	693	693	721	721
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Panel B: Shipping Cost

	AdjustShippingCost		CutShippingCost		RaiseShippingCost		ΔLogShippingCost	
Treatment	0.018	0.012	0.019	0.014	-0.006	-0.009	-0.028	-0.025
	(0.033)	(0.030)	(0.029)	(0.027)	(0.028)	(0.026)	(0.025)	(0.026)
Constant	0.223***	0.227***	0.150***	0.153***	0.150***	0.151***	-0.004	-0.006
	(0.026)	(0.023)	(0.023)	(0.021)	(0.022)	(0.020)	(0.020)	(0.020)
Observations	692	692	692	692	692	692	720	720
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Panel C: Product Description and Introduction of New Listings

	ChangeTitle		ChangeDescription		ChangeNumPictures		LogNewListings	
Treatment	0.001	0.000	-0.008	-0.007	-0.004	-0.008	-0.092	-0.092
	(0.011)	(0.011)	(0.019)	(0.019)	(0.014)	(0.014)	(0.080)	(0.067)
Constant	0.020**	0.021**	0.076***	0.075***	0.038***	0.040***	3.043***	3.043***
	(0.008)	(0.009)	(0.015)	(0.015)	(0.011)	(0.011)	(0.063)	(0.052)
Observations	769	769	790	790	768	768	764	764
Group FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: This table presents regression results on sellers' responses after treatment. AdjustPrice is a dummy that equals one for listings that have adjusted their prices by more than 5% within 13 weeks after treatment. CutPrice, RaisePrice, AdjustShipping-Cost, CutShipingCost, RaiseShippingCost are dummy variables defined in a similar way. ChangeTitle is a dummy that equals one for listings that have updated their product titles within the 13 weeks after treatment. ChangeDescription is a dummy that equals one for listings that have updated their product descriptions within the 13 weeks after treatment; and a set of descriptions include website pictures, pattern type, material, fit, gender, sleeve length, collar, clothing length, item type, color, etc. HaveNewListings is dummy that equals one for a listing if the store to which it belongs has introduced new listings within 13 weeks after treatment; and LogNewListings is the log number of those new listings. Standard errors are in the parentheses. *** indicates significance at 0.01 level, ** 0.5, * 0.1.

C Details on Data and the Experiment

C.1 Measuring Service, Shipping and Product Quality

In order to examine the relationship between quality and growth dynamics in the presence of search and information frictions, we collect a rich set of quality measures on service, shipping and product via (i) direct communication with the sellers and (ii) actual purchases of the products. For each t-shirt variety (of the same design), our quality grading is conducted on all small listings, all medium-size listings with sales between 6 and 50, and the superstar listing (with the largest sales quantity of the variety).

Service Quality. First, we visited the homepage of each store and sent the following message via the platform to engage in pre-transaction service (i.e., inquiry about a particular product):

"Hi, I am wondering if you could help me choose a size that fits my kid, who is 5 years old, 45lbs 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!"

We then constructed a measure of service quality based on whether the message was replied to within two days, which was true for 69% of the listings.

Shipping Quality. To capture the quality of shipping, we recorded the date of purchase, the date of shipment, the date of delivery, carrier name, and the condition of package. The numbers of days between the date of purchase and the date of shipping, the number of days between the date of shipping and the date of delivery, whether the package was delivered successfully, and whether the package was broken upon delivery are used as alternative measures of shipping quality. The medium numbers of days between purchase and shipping and between shipping and delivery are 3 and 12, respectively. Again, there are considerable variations, especially in sellers' turnaround to ship the products.

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 clothing retail business and was invited to grade the quality of the t-shirts.

Each t-shirt was given an anonymous identification number and the owner was asked to grade the t-shirt on 9 quality dimensions, following standard grading criteria used in the textile and garment industry as shown in Panel A of Figure 2. These dimensions include obvious quality defect, fabric durability, fabric softness, wrinkle test, seams (straightness and neatness), outside stray threads, inside loose stitches, pattern smoothness, and trendiness. Most of these metrics, except trendiness, capture differences across t-shirts that are vertical in nature. For example, at equal prices, consumers would prefer T-shirts with more durable fabric, straight seams, and no loose stray threats. The quality examiner grades each t-shirt along the first dimension based on a 0 or 1 scale, and along the other eight dimensions based on a 1 to 5 scale, with higher numbers denoting higher quality. The identification system ensured that the examiner had no information on the purchase price, popularity, and retailer of the t-shirts and whether the t-shirts belonged to our treatment or control group.

In addition, the examiner was asked to price each t-shirt based on her willingness to pay and willingness to sell, respectively. These two additional metrics would reflect not only product quality but also local consumer preferences assessed based on the examiner's retail experience.

T-shirts within the same variety were grouped together for assessment to make sure the grading could better capture within-variety variations. The examiner also conducted two rounds of evaluation that took place several weeks apart to ensure consistency in grading. Panel B of Figure 2 shows examples of the grading and variations across different quality dimensions.

The mean scores vary from 2.6 to 4.2 across different quality metrics. On average, t-shirts scored the worst on inside stray threads and the best on straight seams. Dimensions that record the greatest variations in scores are outside and inside stray threads and pattern smoothness within t-shirt varieties, and pattern smoothness, trendiness, and outside stray threads across varieties.

C.2 The Review Treatments

In our randomized experiment, we group small listings into three groups of different order and review treatments: a control group C without any order and review treatment, T1 which receives 1 order randomly generated by the research team and a star rating, T2 which receives 1 order and 1 detailed review on shipping and product quality in addition to the star rating.

To generate the content of the product and shipping 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 key words. 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 leave positive reviews to all listings unless there are obvious quality defect or shipping problems, in which case no review is provided. Of all the orders placed, about 8 percent have obvious quality defect or shipping issue.

D Details on the Simulation Procedure

We implement the Method of Simulated Moments according to the following procedure.

D.1 Recover Marginal Cost

In the first step, We use the data 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 demand derivative $\frac{\partial D_i}{\partial p_i}(\mathbf{p}, \mathbf{r}, \mathbf{s})$ based on equation (8) and (**??**).

D.2 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 D.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 steps:

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.

D.3 Simulate One Period

In each period, we use the weighted sampling without replacement to generate the consumer's search sample of size K and draw idiosyncratic preference ε from the I.I.D. type I extreme value distribution. Based on the average reviews, we calculate the expected quality and the expected utility of purchasing from each seller in the search sample. Then, we simulate the purchasing decision, the realized experience for the consumer, and the review he/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 sales. In addition, we allow the sellers

to update their prices by solving the static pricing problem at the frequency that matches with the observed frequency of price adjustment.

D.4 Simulate Moments

Starting from the initialized market, we repeat step D.3 for $T = 10000$ times so that the market share based on cumulative orders reaches an invariant distribution. Then, we simulate forward for another ΔT periods to produce moments from a stabilized market. Specifically, we calculate the distribution of cumulative revenue for the sellers, the regression coefficient of log price and quality, and the regression coefficient of the market share of cumulative orders on quality in the final period, i.e. $t = T + \Delta T$. And we calculate the dependence of seller's new order on cumulative orders using simulated data from period $T + 1$ to $T + \Delta T$.

D.5 Weighting Matrix and Objective Function

We bootstrap our data sample moments 1000 times and construct the weighting matrix W. The objective function used for optimization is

$$Q(\theta) = -\frac{1}{2} \left(g_0 - \gamma_m(\theta)\right)' W \left(g_0 - \gamma_m(\theta)\right),$$

where g_0 is the data moments vector, $\gamma_m(\theta)$ is the simulated moments vector based on $m = 100$ simulations, and $\theta = (s_0, \sigma, \rho, K, \gamma)$ is the vector of parameters.