NBER WORKING PAPER SERIES

GAINS FROM PRODUCT VARIETY:
EVIDENCE FROM A LARGE DIGITAL PLATFORM

Erik Brynjolfsson
Long Chen
Xijie Gao

Working Paper 30802
http://www.nber.org/papers/w30802

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022

We thank Seth Benzell, Ruyu Chen and Yu Jeffrey Hu for excellent comments and suggestions. Financial support from the Stanford Digital Economy Lab and Luohan Academy is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w30802.ack

Gains from Product Variety: Evidence from a Large Digital Platform
Erik Brynjolfsson, Long Chen, and Xijie Gao
NBER Working Paper No. 30802
December 2022
JEL No. O0

ABSTRACT

E-commerce sales have grown rapidly worldwide, massively increasing the availability of new products. We examine data from the largest digital platform in China and find that the number of book titles almost doubled, prices fell somewhat, and most new books are sold to consumers with unusual tastes. Demand for these niche products was significantly more inelastic than that of mass products. Embedding the estimates of demand elasticity into a two-segment CES framework, we find the welfare gain from increased variety was about 40 times the gain from lower prices and that rural consumers enjoyed the largest gains.

Erik Brynjolfsson
Stanford Digital Economy Laboratory
353 Jane Stanford Way, Office 136
Stanford, CA 94305
and NBER
erik.brynjolfsson@gmail.com

Long Chen
Luohan Academy
Hangzhou
Zhejiang Province
China
chenlong.chl@luohanacademy.com

Xijie Gao
Luohan Academy
Hangzhou
Zhejiang Province
China
xijie.gxj@luohanacademy.com

Latest version of this paper is available at https://www.brynjolfsson.com/research

1 Introduction

The Internet and platform technologies have facilitated coordination among large sets of buyers and sellers, transforming the retail market. As a result, an increasing fraction of retail trade has shifted online. According to Euromonitor, in 2020, e-commerce sales as a percentage of total retail sales exceeded 27% in China and 20% in the U.S. By integrating markets and reducing the cost of distribution, online platforms provide a large and growing set of product choices to consumers, potentially increasing consumer welfare significantly.

The term "The Long Tail" was coined to describe the phenomenon where niche products account for a larger share of sales in online markets, reflecting the fact that these products benefit disproportionately from lower search costs (Anderson, 2004). In particular, digital platforms cater to the needs of consumers with unusual tastes by showcasing and distributing a large selection of product choices. For instance, Brynjolfsson et al. (2003) found that Amazon stocked 57 times more book titles than a typical physical bookstore.[1] In 2000, there were around 5 million SKUs on Amazon, and the number of products increased 100 times to an astonishing 500 million SKUs by 2016.[2] On the largest digital platform in China, the overall number of items offered more than tripled in just four years, rising from about 1 billion items in 2015 to more than 3 billion in 2019.

While there has been growing attention to the gains from new products in the economy since the seminal work on the topic by Bresnahan and Gordon (1996), there is little recent empirical evidence on the rapidly growing role of niche products. In this paper, we use detailed data from the largest e-commerce platform in China to document large increases in product variety in recent years. We find that niche products are qualitatively different from mass products, resulting in disproportionate welfare gains for consumers.

[1]Similarly, in 2018, there were 75 million SKUs on Walmart e-commerce site, whereas an offline Walmart Supercenter stocks only 120,000 different items on its shelves.

[2]Source: https://www.bigcommerce.com/blog/amazon-timeline-infographic/.

In order to quantify the effects, we obtain detailed transaction data for three categories of books from the e-commerce platform. An advantage of examining books is that each title is classified using a unique International Standard Book Number (ISBN). When measured by the number of distinct ISBNs available, we find the number of products increased by 98% from 2015 to 2019.[3]

Despite the large increase in product variety, the market share of the top products remains largely unchanged. Specifically, we define the top 1000 ISBNs as mass products, and find that they account for just under 60% of market share in both years. However, they only account for 0.6% of total ISBNs available in 2015 and 0.3% of ISBNs in 2019, which implies that most of the increase in product variety has been among niche products. In fact, we find that the average market share of a new ISBN is only about 29% that of a surviving ISBN. There are also fewer listings selling new ISBNs, on average 2.8 for a new ISBN and 7.2 for a surviving ISBN.

To reflect the fact that most of the increase in product variety is among niche products, we posit a simple two-segment constant elasticity of substitution (CES) framework which allows heterogeneity in demand elasticity for mass products and niche products. The advantage of CES demand is that we can infer the surplus generated by new products from the expenditure shares. Conditional on the level of expenditure shares of new products, the gains will be larger when new products are less substitutable than existing products. If the demand of niche products is less elastic compared to mass products, neglecting the difference in demand elasticities would underestimate the gains from variety.

We estimate demand elasticity for mass and niche products separately using an exogenous driver of variation of prices and quantities sold. We consider a simple log-linear demand system, and coefficient on the logarithm of price is the elasticity estimate. The variation used to identify the elasticity comes from the price variation

[3] While this is an impressive increase in product variety, it is less than the rate of increase overall on the platform, in part reflecting the introduction of entirely new product categories. Thus, our estimates of the gains from greater product variety during this period are likely to be conservative.

across different counties. A direct ordinary least squares (OLS) regression may suffer from endogeneity because seller might adjust the price of a book in response to temporary county-level demand shocks. Accordingly, we use an instrumental variables approach, where the instrument is the shipping fee that consumers pay. As shipping fee is a part of the price, the two are highly correlated. The identifying assumption is that sellers do not change shipping fees in one county in response to a demand shock in that county. Based on our interviews with multiple sellers regarding their decision-making process for shipping fees, this assumption is likely to hold.

We find that the large increase in product variety between 2015 to 2019 generates enormous welfare gains. Our estimates indicate that demand is less elastic for niche products: in 2015, a one percent increase in price translates into a 1.9 percent decrease in quantities sold for mass products but only a 1.5 percent decrease for niche products. Using the estimates of demand elasticity, we quantify the gains from variety by applying the two-segment CES framework to the data. The gain from increased variety is about 120% total expenditure on books in 2019, which is about 40 times larger than the gain derived from price effect alone. To put the number into perspective, for the three categories of books in our sample, the consumer gains from variety is about 1.45 billion Yuan. This welfare gain is about 30% higher compared to an approach which does not distinguish mass and niche products.

We further explore the geographic heterogeneity by estimating gains from variety at the county-level. Accounting heterogeneity in demand elasticity reveals that rural consumers enjoy larger gains from variety. This can be due to the fact that rural consumers have limited access to product variety via traditional brick-and-mortar stores, thus benefiting more from the rise of online channel.

The remainder of the paper proceeds as follows. Section 2 presents related economic literature. Section 3 discusses the data and documents the pattern of rising product variety. Section 4 develops a framework to measure consumer welfare gains from increased product variety. Section 5 estimates demand elasticities for mass and niche products. Section 6 reports the gains from product variety by applying the framework to the data. Section 7 presents our conclusions and some ideas for future research.

2 Related literature

Our research is related to several strands of literature that evaluate gains from rising product choices. We follow the tradition in the macro and trade literature that uses a CES demand structure to study the gains from variety. Krugman (1979) pioneers the use of love-of-variety models to study how countries could gain from trade through the import of new varieties. Following Krugman's work, Broda and Weinstein (2006) estimate the value to U.S. consumers of the expanded import varieties between 1972 and 2001 to be 2.6 percent of GDP. Our emphasis on the rise of niche products is closely related to Neiman and Vavra (2019), which shows consumers can enjoy the gains from selection if the newly available niche products can better match their tastes.

Our analysis also relates to a large literature in industrial organization studying the gains from varieties.[4] As noted above, Brynjolfsson et al. (2003) find significant gains to consumer welfare (up to $1.03 billion in 2000) due to the increase in access to book varieties provided by Amazon.com. They estimate the gains to consumers from increased variety to be 7 to 10 times larger than the gain derived from the competitive price effect. These gains have since been dubbed the "Long Tail" benefit of online retail by Anderson (2004). More recently, Quan and Williams (2018) emphasizes the gains from online variety depend critically on the extent to which demand varies across geographies and on how traditional brick-and-mortar retailers respond to those local tastes. After accounting for this heterogeneity, they find the

[4]See, for example, Hausman (1996), Petrin (2002).

variety effect to be about equal in size to the price effect. Our work studies the rise of product varieties on the e-commerce platform over time, where the "Long Tail" benefit increases substantially. We find rural consumers benefit more from varieties, which highlights the importance of e-commerce in narrowing the gap in the access to product variety between rural and urban regions.

This paper is also related to a strand of the industrial organization literature studying product design and obfuscation (Johnson and Myatt, 2006; Ellison and Ellison, 2009; Bar-Isaac et al., 2012). Our estimates of demand elasticity suggest a relatively modest level of price competition on the e-commerce platform, reflecting the large scope of product differentiation even within a seemingly-standardized book market. A seller could use advertising, marketing and product design decisions to influence the consumers' valuations and soften price competition.

Finally, we find the average demand elasticity decreases over time in our sample, which raises concerns over rising market power in the product market. Understanding the trend of product market power has been a focus of recent research, such as in the work of De Loecker et al. (2020), Edmond et al. (2018) and Eeckhout and Veldkamp (2021). Although the primary aim of our research is not to evaluate changes in market power, it is important to realize that the fraction of retail that takes place on the online platforms has been growing rapidly. Therefore, the future evolution of the aggregate market power is increasingly influenced by online competition.

3 Data

We obtain weekly transaction data from the largest Chinese e-commerce platform in 2015 and 2019 for three categories of books: 1) foreign language and linguistics, 2) reference books and encyclopedia, and 3) philosophy and religion.[5] We focus on these three categories because the way the platform catalogued the books evolved

[5] According to open data sources, the total size of the Chinese book market is around 102 Billion Yuan in 2019, and 70% of sales come from online channel. See http://en.openbook.com.cn/EN/Report?reportId=1 for details.

Table 1: Product Variety in 2015 and 2019

(a) Rise of Product Variety

	Year	No. Items	No. ISBNs	Avg No. items/ISBN	Sales share
All 2015 goods	2015	973,354	167,116	5.73	1.00
All 2019 goods	2019	1,313,715	330,597	3.96	1.00
Common 2015-2019	2015	51,206	87,880	9.12	0.94
Common 2015-2019	2019	51,206	87,880	7.21	0.57
2015 not in 2019	2015	922,148	79,236	1.96	0.06
2019 not in 2015	2019	1,262,509	242,717	2.79	0.43

(b) Sales Share of Mass Products

Year	Top 100 ISBNs	Top 1,000 ISBNs	Top 10,000 ISBNs
2015	0.29	0.56	0.85
2019	0.32	0.58	0.87

over time. The nature of the products in these three categories did not experience changes in cataloguing rules over time, which makes it possible to more directly study the gains from product variety.

A product is defined by an item ID which is listing specific. For each item, we gather weekly transaction records at item-week-county level. We observe both sales revenues and sales quantities. The revenue information contains the amount consumers actually pay, which includes the shipping fee. We compute the weekly average price as the ratio of revenues to quantities sold.

In addition to the transaction information, we also observe the listing information of the product, including the item title, seller's information, and most importantly the International Standard Book Number (ISBN). ISBN is a numeric commercial book identifier that uniquely identifies each edition and variation of a book – for instance, the paperback, e-book and hardcover edition of the same book title will have different ISBNs. Thus, the ISBN helps us to identify the same product across different items.

3.1 Rise of Product Variety

The first panel of Table 1 reports the product variety in 2015 and 2019. When measured by distinct ISBNs available, the number of products increased by 98%, from 167,116 in 2015 to 330,597 in 2019. The number of new ISBNs in 2019 is about three times that of the surviving ISBNs. In addition, there are typically multiple items within one ISBN, reflecting different sellers with different ways of positioning the same book. In 2015, there were about 5.7 different items for each identical ISBN. The number of items within an ISBN declined over time. As a result, when measured by the number of distinct items, the increase in product variety is not as great as the increase in ISBNs. Lastly, compared to surviving products, new products tend to be niche products. The average market share of a new ISBN is only about 29% that of a surviving ISBN. There are also fewer items selling new ISBNs, on average 2.8 items for a new ISBN and 7.2 items for a surviving ISBN.

Figure A1 in Appendix shows the product sales distribution in 2015 and in 2019. Products are ranked in terms of their annual sales, which is then plotted against sales rank using a logarithmic scale. Unlike the sales distribution documented by Brynjolfsson et al. (2010) at Amazon, which was a power law, the sales distribution at here does not have a constant slope. Instead, the concave slope is better fitted by a log-normal distribution.[6]

Despite the large increase in the product variety and expansion of the market, the sales *shares* of top products remain largely unchanged. The second panel of Table 1 reports the sales shares of top products. In 2015, we see a market that already has a large share of niche products and this number grows further by 2019. Top 100 ISBNs have a sales share around 30%, top 1,000 ISBNs have a sales share around 60% and top 10,000 ISBNs have a sales share close to 90%. We label an item as a mass product if its ISBN falls into top 1000 ISBNs, while it is a niche product if its ISBN is ranked below 1,000. Brynjolfsson et al. (2010) study the long-tail phenomena using Amazon sales data on books. They find the sales share from niche products is close to 40%, which is consistent with our cutoff at 1,000. In addition, the set of mass products is redefined separately for each year, reflecting the changes

[6]Figure A2 fits the actual sales distribution with a log-normal distribution. It confirms the actual sales distribution is better fitted by the log-normal distribution.

in demand for individual titles over time.

3.2 Product Differentiation with an ISBN

Despite the fact that we have ISBNs as our product identifiers, there remains sizable product differentiation across items even when they have the same ISBN.[7] Within an identical ISBN, items are sold by different sellers with different item titles, pictures, logistics services, prices, free gifts, and cumulative orders. This type of within-ISBN differentiation is common on the platform, which makes items with identical ISBN far from perfect substitutes.

To formally evaluate the product differentiation *within* an ISBN, we consider the correlation between price and sales across items within a same ISBN using the following regression

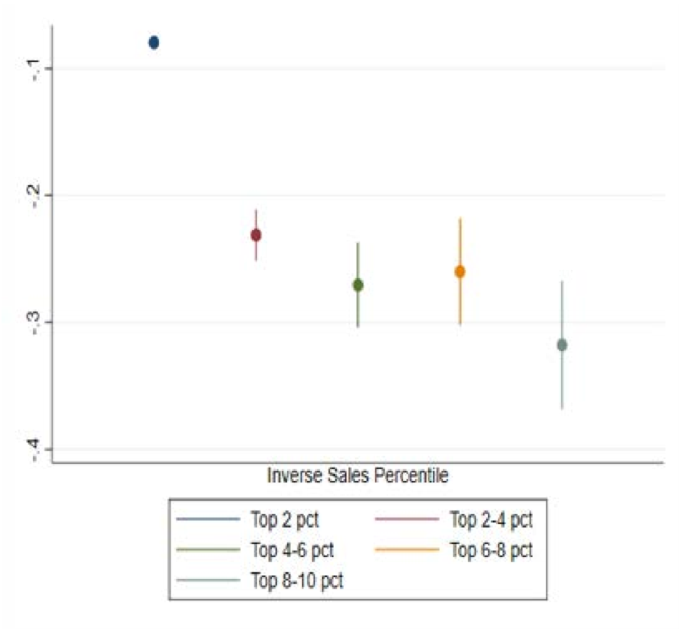
$$log\left(q_{itl}\right) = \alpha_{ntl} + \beta log\left(p_{itl}\right) + \epsilon_{itl} \tag{1}$$

where q_{itl} and p_{itl} are the quantity sold and price of item i at week t to the county l. α_{ntl} are the ISBN-week-county level fixed effects. The coefficient β identifies how price variation across items affects their sales for a given ISBN-week-county, which can be interpreted as a measure of product differentiation within an ISBN.

Figure 1 reports the within-ISBN product differentiation for different product groups in 2015. The product groups are determined by their annual sales. We find a significant degree of within-ISBN product differentiation. When products become more mainstream, the within-ISBN differentiation level increases, which suggests that sellers engage in product differentiation in more competitive mainstream markets.

[7] Figure A3 in Appendix illustrates the potential product differentiation within an ISBN using a search result from the platform.

Figure 1: Product Differentiation in 2015



Notes: This figure reports the within-ISBN correlation between price and quantity sold, which can be viewed as a measure of within-ISBN product differentiation. We rank all ISBNs in terms of annual sales, and divide top 10% of ISBNs into five product groups. For each group, β estimate from the regression (1) is shown.

4 Estimate Gains from Variety

We estimate the gains from product variety using a CES specification for consumer demand. To highlight the difference between mass and niche products, we consider a two-segment CES utility function, which is given as

$$U_t = \left[C_{m,t}^{\frac{\sigma-1}{\sigma}} + C_{n,t}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

and

$$C_{m,t} = \left[\sum_{i=1}^{M_t} (\phi_{i,t} C_{i,t})^{\frac{\sigma_M - 1}{\sigma_M}} \right]^{\frac{\sigma_M}{\sigma_M - 1}} \tag{3}$$

$$C_{n,t} = \left[\sum_{i=1}^{N_t} (\phi_{i,t} C_{i,t})^{\frac{\sigma_N - 1}{\sigma_N}} \right]^{\frac{\sigma_N}{\sigma_N - 1}} . \tag{4}$$

The aggregate consumption (utility) is given by a CES combination of $C_{m,t}$ and $C_{n,t}$, which in turn are produced by combining all the products within the mass and niche product categories. We allow the demand elasticities to differ between mass and niche products. Product i here refers to an item in our data as there is product differentiation even within an ISBN.

Let $P_{i,t}$ and $C_{i,t}$ be the price and consumption of product i at time t. Given the budget constraint $\sum_i P_{i,t}C_{i,t} = E_t$, consumer welfare is the aggregate expenditure normalized by aggregate price index

$$U_t = \frac{E_t}{P_t}. \tag{5}$$

Denote by $s_{M,t}$ ($s_{N,t}$) the expenditure share on mass (niche) products at time t. Conditional on the same expenditure, welfare change can be identified using the change in price index

$$1 + \pi^{CES}_{t,t+1} = \frac{P_{t+1}}{P_t} = \Pi_{i\in\{M,N\}} \left(\frac{P_{i,t+1}}{P_{i,t}}\right)^{\omega_{i,t,t+1}} \tag{6}$$

where

$$\omega_{i,t,t+1} = \frac{(s_{i,t+1} - s_{i,t})/\left(ln\left(s_{i,t+1}\right) - ln\left(s_{i,t}\right)\right)}{\sum_{j\in\{M,N\}}(s_{j,t+1} - s_{j,t})/\left(ln\left(s_{j,t+1}\right) - ln\left(s_{j,t}\right)\right)}. \tag{7}$$

The change in aggregate price index is a weighted average of price changes of mass and niche products. Within mass and niche products, there is product turnover. Let $\Omega^{\star}_{M,t,t+1}$ and $\Omega^{\star}_{N,t,t+1}$ denote the set of common mass and niche products that have transactions in both periods. For the common products, the aggregate price index is again a weighted average of product-level price changes.

$$CPPI_{M,t+1} = \Pi_{i \in \Omega^{\star}_{M,t,t+1}} \left(\frac{P_{i,t+1}}{P_{i,t}} \right)^{\omega_{i,t,t+1}} \tag{8}$$

$$CPPI_{N,t+1} = \Pi_{i \in \Omega^{\star}_{N,t,t+1}} \left(\frac{P_{i,t+1}}{P_{i,t}} \right)^{\omega_{i,t,t+1}} \tag{9}$$

where $CPPI_{M,t+1}$ and $CPPI_{N,t+1}$ are the price index for the common products, the weight of each product $\omega_{i,t,t+1}$ is defined in the same way as in equation 7 and $s_{i,t}$ is the share of total period t expenditure on continuing products allocated to product i.

The above price index does not account for entry of new products. Following the existing literature (Feenstra, 1994; Broda and Weinstein, 2010), we can correct for the product turnover using the expenditure share on new products and exiting products. Conceptually, by assuming a CES utility function, we can infer the surplus generated by new products from the expenditure shares. Let $s^{entry}_{M,t+1}$ ($s^{entry}_{N,t+1}$) denote the share of total period $t+1$ expenditure on mass (niche) products allocated to new products, and $s^{exit}_{M,t}$ ($s^{exit}_{N,t}$) denote the share of total period t expenditure on mass (niche) products allocated to exiting products. The correction for the price index can be expressed as

$$\frac{P_{M,t+1}}{P_{M,t}} = CPPI_{M,t+1} \cdot CEE_{M,t+1} = CPPI_{M,t+1} \cdot \left(\frac{1 - s^{entry}_{M,t+1}}{1 - s^{exit}_{M,t}} \right)^{\frac{1}{\sigma_M - 1}} \tag{10}$$

$$\frac{P_{N,t+1}}{P_{N,t}} = CPPI_{N,t+1} \cdot CEE_{N,t+1} = CPPI_{N,t+1} \cdot \left(\frac{1 - s^{entry}_{N,t+1}}{1 - s^{exit}_{N,t}} \right)^{\frac{1}{\sigma_N - 1}} . \tag{11}$$

The correction term CEE reduces the inflation when expenditure share on new products is larger than the expenditure share on exiting products. It happens when there are more varieties on the market over time, or equivalently the entering products are more attractive to consumers. In addition, the value of new products also depends on the demand elasticities, σ_M and σ_N. When elasticity becomes larger, the value of new products becomes smaller.

The demand elasticities σ_M and σ_N can differ for many reasons. First, as the market size is different for products with different levels of popularity, demand elasticity of mass and niche products tend to be different. For example, a large market size may attract more sellers to enter the market, which increases the competition and demand elasticity. Also, there is a potential salience effect. The mass products are more likely to enter the consideration set of consumers making them more sensitive to price differences. If we neglect the elasticity difference between mass and niche products, it will most likely underestimate the gains from variety by mistakenly treating niche products as just as substitutable as existing products.

To convert the change in price index into consumer surplus, we ask if price stays at the initial level, how much more consumers have to spend to achieve the same level of utility at period $t+1$

$$\Delta CS = E^{\star}(U_{t+1}, P_t) - E_{t+1} = \left(\frac{P_t}{P_{t+1}} - 1\right) E_{t+1}. \tag{12}$$

Under the CES specification, the gains from new variety can be fully characterized by the change in price index. As the change in aggregate price index depends on both Common Products Price Index ($CPPI$) and product entry (CEE), the gains from variety can be isolated by imposing $CPPI$ unchanged, and we calculate the gains from variety as

$$\left(\frac{1}{CEE_{M,t+1}^{\omega_{M,t,t+1}} CEE_{N,t+1}^{\omega_{N,t,t+1}}} - 1\right) E_{t+1}. \tag{13}$$

5 Demand Elasticity Estimation

We use a log-linear demand system to estimate the own-price elasticity and omit the cross-price elasticity. While the BLP estimates allow for flexible substitution patterns, the estimation can be numerically unstable in setting such as ours. Knittel and Metaxoglou (2014) document the implementation procedures, which combine various optimization algorithms, starting values, and tolerances of the fixed-point iterations, often converge to local optima, and the resulting elasticity estimates exhibit a substantial amount of variation.[8] Given that our goal is to estimate demand elasticity separately for mass and niche products, our estimates are simple and computationally tractable.

We denote the quantities as q and prices as p. For an item i sold to county l at week t in the quarter q, we estimate log-linear regressions

$$log(q_{itl}) = \alpha_{it} + \beta_s log(p_{itl}) + \gamma_{ql} + \epsilon_{itl} \tag{14}$$

where α_{it} are item-week fixed effects. The coefficient on prices β_s measures the price elasticities separately for products within different categories (mass or niche). The variation used to identify the elasticity comes from the price variation across different counties. In addition, the high-dimensional fixed effects at item-week level guarantee that the demand response we study comes from the exact same product of the same seller within a short period of time. We control for quarter-county level fixed effects γ_{ql}.

The direct OLS regression may suffer from standard endogeneity concerns: sellers could adjust prices in response to county-level demand shocks. On the e-commerce platform, although the listing price may be same across consumers, sellers could still discriminate by issuing coupons that target particular markets and consumers.

If sellers tend to increase the price when there is a positive demand shock (or vice-versa), we would expect the OLS estimates to be biased towards zero. To minimize the endogeneity concern, we apply an Instrumental Variable (IV) approach. The instrument we use is the shipping fee that consumers pay on the transaction. As

[8] Knittel and Metaxoglou (2014) find the own-price elasticity for the product with highest market share can vary by a factor of 2 to 4.

the shipping fee is included in product sales, it is a part of prices. The identifying assumption is that sellers are unlikely to change shipping fees in response to a temporary county-level demand shock.

The rationale behind the IV is two-fold. Firstly, on the platform, each seller sells a large number of products every day. The average number of items listed per seller on a particular day is more than 8,000.[9] Thus, it is unlikely that sellers would individually change shipping fees in response to county-level demand shock. Secondly, the way that shipping fee is set by sellers also mitigates the concern of endogeneity. Each seller only has a few of shipping templates, and each specifies how shipping fee is calculated based on shipping destination and logistics service providers. Each item is associated with one shipping template, and the shipping fee will be calculated automatically once a transaction takes place. The fact that one shipping template is shared by a large number of items makes the identifying assumption likely to hold. Furthermore, according to several sellers we interviewed, in cases when they change the shipping fee of an item, it is usually in terms of switching between two templates, rather than adjusting shipping template in response to location-specific demand shock of the item.

The first panel of Table 2 reports the first stage regression between shipping fee and prices. As expected, the amount of within item-week price variation that can be explained by shipping fee is large, about 19% in 2015 and 10% in 2019.

The second panel of Table 2 shows the our estimates of demand elasticities using weekly data in 2015. The first three columns show the results using OLS regressions. The last three columns show the IV regressions where we instrument prices using shipping fees. Several observations can be made. First, the demand is more elastic for mass products. In 2015, a one percent increase in prices translates into a 1.9 and a 1.5 percent decrease in sales for mass and niche products respectively. Second, for each specification, the IV estimates are larger compared to the OLS

[9] In 2015, we extract the number of items listed for each seller in the last day of each month and compute the average number of items listed per seller.

Table 2: Demand Elasticities in 2015 and 2019

(a) Price and Shipping Fee

	2015			2019		
	All	Mass	Niche	All	Mass	Niche
log(1+shipping fee)	0.160	0.130	0.189	0.143	0.140	0.159
	(0.002)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)
Item-Week FE	YES	YES	YES	YES	YES	YES
Quarter-County FE	YES	YES	YES	YES	YES	YES
R^2-within	0.188	0.091	0.327	0.101	0.067	0.149
Obs	3,479,469	2,397,106	1,082,011	8,317,964	5,907,659	2,409,954

(b) Demand Elasticities in 2015

	OLS	OLS: Mass	OLS: Niche	IV	IV: Mass	IV: Niche
log(weekly price)	-1.439	-1.423	-1.283	-1.756	-1.931	-1.510
	(0.069)	(0.084)	(0.036)	(0.059)	(0.075)	(0.033)
Item-Week FE	YES	YES	YES	YES	YES	YES
Quarter-County FE	YES	YES	YES	YES	YES	YES
R^2-within	0.079	0.076	0.110	0.022	0.013	0.050
Obs	3,479,469	2,397,106	1,082,011	3,479,469	2,397,106	1,082,011

(c) Demand Elasticities in 2019

	OLS	OLS: Mass	OLS: Niche	IV	IV: Mass	IV: Niche
log(weekly price)	-1.226	-1.308	-1.034	-1.402	-1.395	-1.124
	(0.093)	(0.116)	(0.046)	(0.105)	(0.103)	(0.039)
Item-Week FE	YES	YES	YES	YES	YES	YES
Quarter-County FE	YES	YES	YES	YES	YES	YES
R^2-within	0.050	0.051	0.089	0.007	0.004	0.016
Obs	8,317,964	5,907,659	2,409,954	8,317,964	5,907,659	2,409,954

Notes: This table reports the estimates of demand elasticity. The first panel reports the results of first-stage regression between weekly price and shipping fee. The second and third panel report the demand elasticities in 2015 and 2019 respectively. In each of these two panels, column (1)-(3) show the OLS estimates for an average product, mass product, and niche product separately, whereas column (4)-(6) show corresponding elasticity estimates using IV regression, where we instrument price using shipping fee that consumers pay. As the regression contains high-dimension FEs, we report R^2-within of each regression which computes the R^2 of the regression where every variable has already been demeaned with respect to all the fixed effects. Standard errors are clustered at county-level and shown below the estimates in parentheses.

estimates. Sellers do seem to be more likely to increase prices when there is a positive demand shock, which generates a standard attenuation bias in the OLS estimates. By instrumenting prices using the shipping fees, we address this bias.

The third panel of Table 2 reports the elasticity estimates in 2019. Demand elasticities declined compared to 2015. A one percent increase in prices translates into a 1.4 and a 1.1 percent decrease in demand for mass and niche products. As the number of products proliferate, sellers may engage in product differentiation via positioning and marketing more extensively.[10,11]

6 Gains from product variety

6.1 Basic gains from variety

We quantify the gains from product variety by applying the framework in section 4 to sales data for books on the platform. As gains from variety decrease with demand elasticity, to be conservative, we use the higher elasticity estimates in 2015 for our benchmark results.

The first row in the first panel of Table 3 reports the Common Products Price Index $CPPI$ to be 0.967.[12] It implies that price index decreases by 3.3% from 2015 to 2019 for the continuing products. As a result, consumer surplus increases by 3% . However, once we correct for product turnover in the price index, as in the second row of the panel, we find a much sharper decline in the price index: the turnover-adjusted price in 2019 is 56% lower than that of 2015. Given the large

[10]In Appendix A.2, we explore the change in product differentiation over time. In particular, we compare the estimate of demand elasticity with the price coefficient obtained from an alternative regression, where cross-item price variation is taken into consideration. The gap between the two price coefficients serves as a measure of within-ISBN product differentiation.

[11]The decline in demand elasticities could also be driven by the change in platform design and customer composition. The early adopters of e-commerce may have characteristics and share different tastes (Moore, 2002) compared to the later adopters. In addition, changes in the search and recommendation algorithm, and other aspects of platform design could also affect demand elasticity.

[12]The $CPPI$ takes the overall inflation in China during the sample period into consideration. From 2015 to 2019, the overall consumer price index increases by 8.9%.

increase in the product variety, consumers reallocate their expenditure towards new products, which results in significantly lower cost of living. The large decline in prices translates into large increase in consumer surplus reflecting the gains from variety. The gain from increased variety is about 120% total expenditure on books in 2019, which is about 40 times larger than the gain derived from price effect alone.

To put the number into perspectives, for the three categories of books in our sample, the total expenditure in 2019, E_{2019}, is about 1.2 billion Yuan while the gains from variety are about 1.45 billion Yuan. For all categories of books on the platform, E_{2019} is about 40 billion Yuan.[13] If there is a similar entry pattern, given the estimates of demand elasticity, the gains would be 48 billion Yuan for books overall. At current exchange rates, that is about 6.7 Billion U.S. dollars.

The third and fourth row in the first panel of Table 3 show the gains from variety for different values of σ_M. Even when the demand for mass products becomes much more elastic $\sigma_M = 2.5$, the gains from variety are still large given the low elasticity for niche products. This highlights that the value of niche products is largely underestimated if we do not correctly identify the demand elasticity for niche products.

We also compare our estimates with the more traditional approach of a standard one-segment CES demand. The second row in the second panel of Table 3 shows the gains from variety using the existing approach. With an average elasticity $\sigma = 1.8$, price index declines by 50% between 2015 and 2019, implying a gain of consumer welfare of $0.94E_{2019}$. The traditional approach assumes that demand elasticities for mass and niche products are identical, but since we observe the sharp increase of niche products over time and niche products tend to have more inelastic demand, this leads to an underestimate of the gains from product variety of about 30%.

One concern of our approach is that it may generate spurious product entry and exit which could potentially bias the welfare estimates upwards. This is because if

[13] Source: http://www.xinhuanet.com/tech/20220422/d1fe616d664a47a2a0039d4860af9431/c.html.

Table 3: Gains from Variety

(a) Gains from Variety with Heterogeneous Demand Elasticities

Elasticity	Change in Price Index	Change in CS	Gains from Variety
Common Products Price Index	0.967	$0.03E_{2019}$	
$\sigma_M = 1.9$, $\sigma_N = 1.5$	0.438	$1.28E_{2019}$	$1.21E_{2019}$
$\sigma_M = 1.5$, $\sigma_N = 1.5$	0.357	$1.80E_{2019}$	$1.71E_{2019}$
$\sigma_M = 2.5$, $\sigma_N = 1.5$	0.484	$1.07E_{2019}$	$1.00E_{2019}$
$\sigma_M = 1.8$, $\sigma_N = 1.8$	0.520	$0.92E_{2019}$	$0.86E_{2019}$

(b) Gains from Variety with One-Segment CES

Elasticity	Change in Price Index	Change in CS	Gains from Variety
Common Products Price Index	0.971	$0.03E_{2019}$	
$\sigma = 1.8$	0.500	$1.00E_{2019}$	$0.94E_{2019}$
$\sigma = 1.4$	0.258	$2.88E_{2019}$	$2.76E_{2019}$
$\sigma = 2.2$	0.624	$0.60E_{2019}$	$0.55E_{2019}$

Notes: The first panel of this table reports gains from variety by applying the two-segment CES approach to the book sales data. The second panel of this table reports gains from variety by applying the traditional one-segment CES approach to the book sales data. In each panel, the second column shows change in price index between 2015 to 2019, where overall inflation in China during the period is taken into account. The third column reports the corresponding change in consumer welfare. The last column isolates the gains from variety when imposing the CPPI to be unity.

an item changes its category from one year to another, we treat it as exiting from one category and entering into another category. To gauge the magnitude of such bias, the last row in first panel of Table 3 reports the case when σ_M and σ_N are equal to average elasticity 1.8. The gain from variety is smaller compared to that in the one-segment CES approach, where the latter has no concern for such bias. The results show that the large gains from variety do not come from the spurious entry and exit. If anything, the spurious entry pattern slightly reduces the estimated gains.

6.2 Heterogeneity across markets

The effects of digital marketplace may differ across markets. Consumers in rural regions may have poorer access to new varieties via traditional brick-and-mortar stores and thus benefit more from the online channel. In this section, we explore geographic heterogeneity and estimate gains from variety at the county-level.

We construct a county-level measure of urbanization using population density of the county, where we retrieve population data from the 2010 National Census and calculate the area of each county using geographic information published by the planning authority.[14]

We examine how the gains from variety vary across counties. The first panel of Figure 2 plots the gains against the urbanization measure, which confirms a strong negative correlation between the two. The result is consistent with our expectation that the development of online marketplace provides rural counties with the much-needed access to variety, delivering larger benefits. Table A2 in Appendix reports the regression output when we regress gains from variety on the level of urbanization, income per capita, and distance to Hangzhou.[15] In addition to the urbanization level, counties with higher incomes tend to derive lower gains from variety while remote counties, measured by the distance to the e-commerce hub Hangzhou, enjoy higher benefits from variety.

One concern of the result above is that we impose the assumption that all counties have the same demand elasticity. Given the large heterogeneity across markets, the demand elasticity can vary for many reasons. In particular, rural consumers have more limited choices offline and therefore they rely more on the online channel. This is likely to translate into a lower demand elasticity. In addition, there can be taste heterogeneity as rural consumers tend to be poorer and demand fewer niche products. We explore these possibilities by dividing the sample into four groups based on population density, and estimate demand elasticity separately for each group.

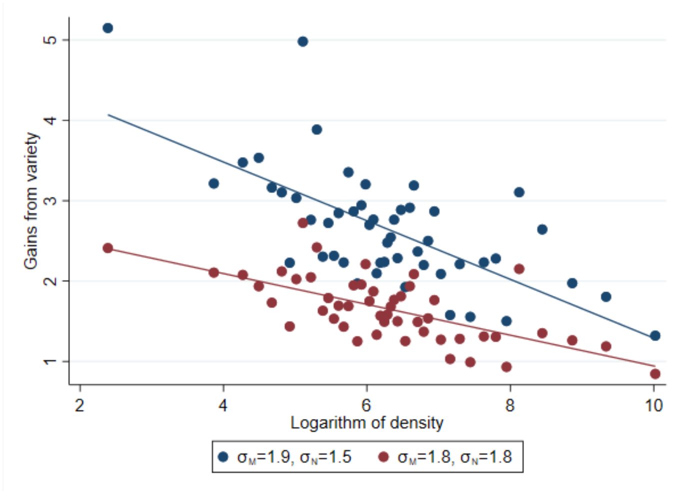
The second panel of Figure 2 reports the estimates of demand elasticity for different groups of counties in 2015, with group 1 as the least populated counties and group 4 as the most populated counties. For mass products, the demand elasticity

[14] The county-level population and geographic information are made available by Gao et al. (2021) via `https://www.scidb.cn/en/detail?dataSetId=849628989872930816`.

[15] Hangzhou is the hub and enter of e-commerce in China. Distance to Hangzhou is another proxy of e-commerce development.

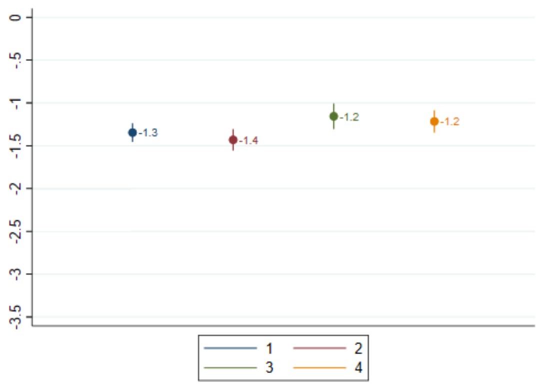
Figure 2: Heterogeneity across Markets

(a) Gains from Variety and Urbanization



(b) Heterogeneity in Demand Elasticities across Counties

Mass products



Niche products

Notes: This figure shows the heterogeneity across markets. The first panel shows the binscatter plot between county-level gains from variety and population density. 1,733 counties are included. We estimate the gains from variety using the method documented in section 4, with two sets of elasticity estimates. The gains on the vertical axis are normalized by county-level expenditure in 2019. The second panel plots the estimate of demand elasticity for different groups of counties in 2015. We equally divide the weekly-transaction data into four groups, with group 1 as the least populated counties and group 4 as the most populated counties. The left sub-panel reports elasticity estimates for mass products and the right sub-panel reports estimates for niche products. Standard errors are clustered at county-level.

increases strongly against urbanization. This is consistent with the idea that more populated regions have many competing channels for mass products, thus demand is very elastic. Rural consumers are more reliant on online market for mass products, resulting in a less elastic demand. For the niche products, demand elasticity is similar across counties, which can be driven by both taste heterogeneity and alternative shopping channels.[16] The results suggest that gains from variety for rural consumers would be even higher if we account for the heterogeneity of demand elasticities across counties.

7 Concluding Remarks

One of the biggest potential benefits from online platforms is that they facilitate the entry of new products, particularly those addressing niche needs that would otherwise go unaddressed. This paper studies the magnitude and nature of new product variety at the largest platform in China, and quantifies their welfare implications for consumers. Using sales data of three categories of books in the largest e-commerce platform in China, we find the amount of product variety, measured by the number of distinct ISBNs, almost doubled between 2015 and 2019. In the meantime, the sales *shares* of top 1000 ISBNs remained largely unchanged. This suggests the new products tend to be niche products and individually account for small market shares, though they are collectively important.

We develop a two-segment CES framework and find that the demand for niche products tends to be less elastic. Embedding the estimates of demand elasticities into the framework, we find the gain from product variety is very large. In fact, consumers would need to spend 120% more in order to achieve the same level of

[16] The heterogeneity in demand elasticity across regions is in line with Brynjolfsson et al. (2009), in which they show that Internet retailers face significant competition from brick-and-mortar retailers when selling mass products, but are virtually immune from competition when selling niche products. However, the greater benefits for rural consumers contrasts with findings by Dolfen et al. (2022) that earlier waves of e-commerce disproportionately benefited American consumers in more densely-populated areas. The difference may reflect differences in the penetration of e-commerce – we examine a country, time period and product category where e-commerce is more developed.

utility in 2019 if there were no increase in product variety. Our results highlight not only the explosion of new products, but also the importance of accounting for the fact that most new products are niche products with relatively inelastic demand. Welfare gain is about 30% higher compared to an alternative approach which does not distinguish mass and niche products.

When we examine geographic heterogeneity, we find that rural consumers enjoy larger gains from variety, and accounting heterogeneity in demand elasticity would benefit them disproportionately. This is consistent with the idea that rural consumers have limited access to product variety at traditional brick-and-mortar stores and thus benefit more from the rise of choices in the online channel.

Many related research questions remain for future research. For example, we need to better understand the competition of sellers and the nature of product creation. We find that despite the multiple listings within an ISBN, demand for even mass products is still not very elastic. This implies even in a standardized book market, sellers differentiate via marketing and product positioning. Besides, our framework does not explicitly model the search and information frictions in the process of consumer decision making. In a dynamic setting where these types of frictions are incorporated, product differentiation could also be a source of obfuscation which increases the search costs and reduces consumer welfare.

References

Anderson, C. (2004). The long tail. *Wired Magazine*, 12(10):170–177.

Bar-Isaac, H., Caruana, G., and Cuñat, V. (2012). Search, design, and market structure. *American Economic Review*, 102(2):1140–60.

Bresnahan, T. F. and Gordon, R. J. (1996). *The economics of new goods*, volume 58. University of Chicago Press.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585.

Broda, C. and Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *American Economic Review*, 100(3):691–723.

Brynjolfsson, E., Hu, Y., and Rahman, M. S. (2009). Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science*, 55(11):1755–1765.

Brynjolfsson, E., Hu, Y., and Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management science*, 49(11):1580–1596.

Brynjolfsson, E., Hu, Y. J., and Smith, M. D. (2010). The longer tail: The changing shape of amazon's sales distribution curve. *Available at SSRN 1679991*.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

Dolfen, P., Einav, L., Klenow, P. J., Klopack, B., Levin, J. D., Levin, L., and Best, W. (2022). Assessing the gains from e-commerce. *American Economic Journal: Macroeconomics*, 15(1).

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How costly are markups? Technical report, National Bureau of Economic Research.

Eeckhout, J. and Veldkamp, L. (2021). Data and market power. Technical report, Columbia University Working Paper Sep.

Ellison, G. and Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, pages 157–177.

Gao, L., Song, D., Yang, Y., Xu, C., and Yang, X. (2021). District/County Level Population Census GIS Datasets in China (1953-2010).

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. *University of Chicago Press*, pages 207–248.

Johnson, J. P. and Myatt, D. P. (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review*, 96(3):756–784.

Knittel, C. R. and Metaxoglou, K. (2014). Estimation of random-coefficient demand models: two empiricists' perspective. *Review of Economics and Statistics*, 96(1):34–59.

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of international Economics*, 9(4):469–479.

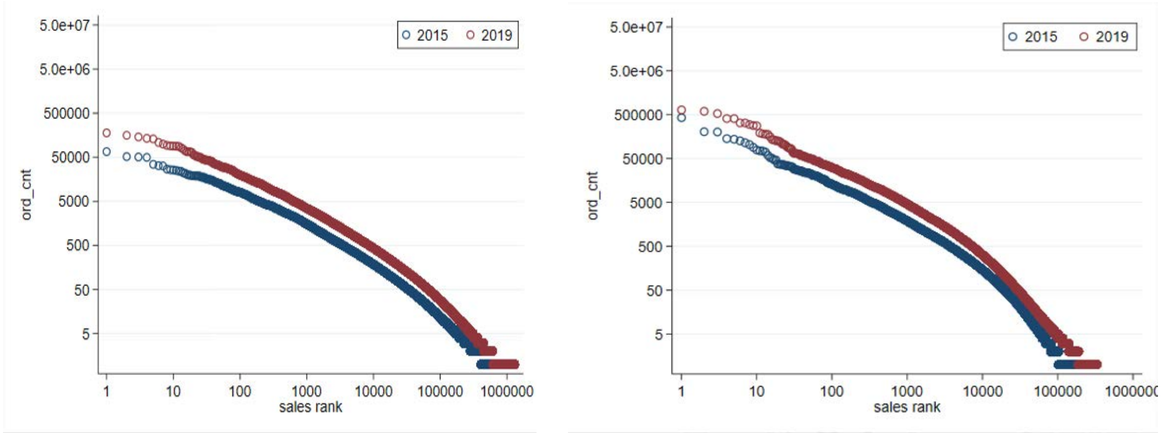
Moore, G. A. (2002). Crossing the chasm.

Neiman, B. and Vavra, J. S. (2019). The rise of niche consumption. Technical report, National Bureau of Economic Research.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729.

Quan, T. W. and Williams, K. R. (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49(4):877–913.

Figure A1: Product Sales Distribution: 2015 and 2019



(a) Item-level (b) ISBN level

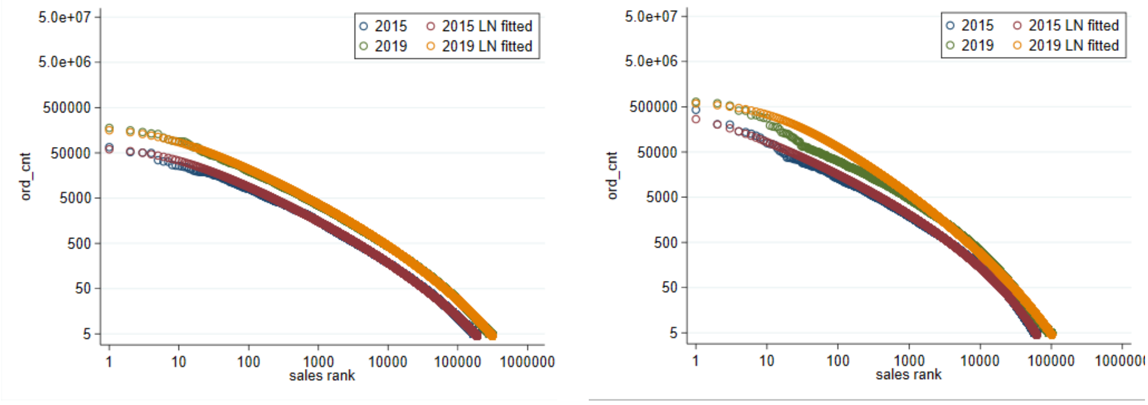
Notes: These two graphs plot the product sales distribution. Products are ranked in terms of their annual sales, which is then plotted against sales rank using the logarithmic scale. The left panel reports the sales distribution when sales are aggregated to item-year level, while the right panel shows the distribution when sales are aggregated to ISBN-year level.

A Appendix

A.1 Product sales distribution

Figure A1 shows the product sales distribution in 2015 and 2019. Several characteristics are noticeable simply by inspecting the plot. First, unlike the sales distribution documented by Brynjolfsson et al. (2010) at Amazon, which was a power law, the sales distribution at here does not have a constant slope. Instead, the concave slope is better fitted by a log-normal distribution. Figure A2 shows the product sales distribution and fitted log-normal distribution. The fitted log-normal distribution is closely aligned with the actual sales distribution, especially for the item-level product sales distribution. Second, the overall market size becomes much larger over time. The sales in 2019 exceed sales in 2015 by 77 percent for an average item and 21 percent for an average ISBN. Third, the graph also confirms that overall product variety increases dramatically over time. As noted above in Table 1, when measured by the number of distinct ISBNs (items) available, the number of product increases by 98 (35) percent.

Figure A2: Fitting Product Sales Distribution



(a) Item-level

(b) ISBN level

Notes: These two graphs plot the product sales distribution and fitted lognormal distribution. We truncate the product sales distribution from the below and keep the products with annual sales larger than 5. For each year and product classification, we then estimate a truncated lognormal distribution and have the fitted lognormal distribution plotted along with the actual sales distribution.

A.2 Product Differentiation and the Decline in the Demand Elasticity

Figure A3 illustrates the potential product differentiation within an ISBN using a search result from the platform. Within an identical ISBN, items are sold by different sellers with different item titles, pictures, logistics services, prices, free gifts, and cumulative orders.

In the paper, we use the elasticity estimates in 2015 to calculate the gains from variety. As shown in Table 2, the demand is more inelastic in 2019. One explanation for the decline in price elasticity is an increase in the level of product differentiation. In the main regression, we estimate demand elasticities by focusing on the cross-county variation in prices for a given item-week pair. To evaluate whether the products become more differentiated over time, we also consider the following regression

$$log(q_{itl}) = \alpha_{nt} + \beta_s^{\star} log(p_{itl}) + \gamma_{ql} + \epsilon_{itl} \tag{A1}$$

Figure A3: Product Differentiation within an ISBN: an example

Notes: This figure shows the search result of one popular ISBN on December 27, 2021. Even within one identical ISBN, items from different sellers are different in many aspects, including price, cumulative order, number and quality of free gifts, and quality of delivery services.

Table A1: Product Differentiation

(a) 2015

	OLS	OLS: Mass	OLS: Niche	OLS	OLS: Mass	OLS: Niche
log(weekly price)	-1.439	-1.423	-1.283	-0.917	-0.895	-0.929
	(0.069)	(0.084)	(0.036)	(0.028)	(0.031)	(0.027)
Item-Week FE	YES	YES	YES			
ISBN-Week FE				YES	YES	YES
Quarter-County FE	YES	YES	YES	YES	YES	YES
R^2-within	0.079	0.076	0.110	0.067	0.064	0.086
Obs	3,479,469	2,397,106	1,082,011	3,826,690	2,479,727	1,346,573

(b) 2019

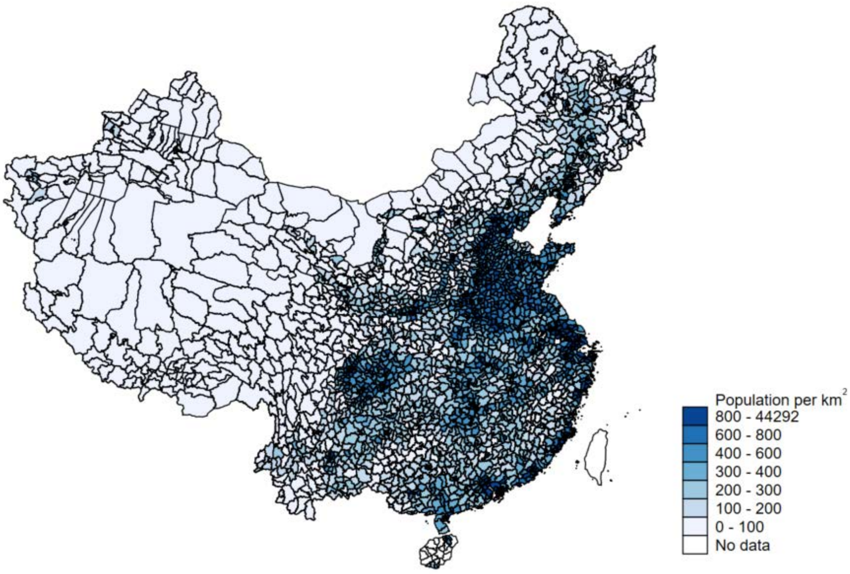
	OLS	OLS: Mass	OLS: Niche	OLS	OLS: Mass	OLS: Niche
log(weekly price)	-1.226	-1.308	-1.034	-0.716	-0.707	-0.756
	(0.093)	(0.116)	(0.046)	(0.034)	(0.038)	(0.031)
Item-Week FE	YES	YES	YES			
ISBN-Week FE				YES	YES	YES
Quarter-County FE	YES	YES	YES	YES	YES	YES
R^2-within	0.050	0.051	0.089	0.039	0.038	0.068
Obs	8,317,964	5,907,659	2,409,954	8,585,396	5,958,315	2,626,725

Notes: This table studies the degree of product differentiation by comparing price coefficients from two different regression specifications. In each panel, the first three columns report price coefficient using specification 14, whereas the last three columns show the price coefficient using specification A1. As the regression contains high-dimension FEs, we report R^2-within of each regression. R^2-within computes the R^2 of the regression where every variable has already been demeaned with respect to all the fixed effects. Standard errors are clustered at county-level and shown below the estimates in parentheses.

where α_{nt} are ISBN-week fixed effects. The $\beta_s^\star$ is estimated by using cross-item and county data within an ISBN-week. Compared to the main regression, the difference in the β estimates can be viewed as a measure of product differentiation.

Table A1 shows the results. The first three columns show the estimates when only within item-week price variation is considered. In contrast, the last three columns report the estimates when we allow for cross-item price variation. Once we allow for within-ISBN variations, the price coefficient of mass products becomes slightly smaller than that of the niche products, which suggests a higher degree of product differentiation *within* mass products. Besides, the degree of product differentiation seems to increase since the reduction in price coefficient is larger in percentage terms in 2019. If indeed products become more differentiated over time, it will further increase the gains from variety, and the standard approach that neglects the niche

Figure A4: Population Density across Counties



Notes: This graph plots the county-level population density on the map of China. The county-level population and geographic information are made available by 2010 National Census and planning authority respectively.

nature of new products will underestimate gains from variety by more.

A.3 Additional Results on Heterogeneity across Markets

Figure A4 illustrates the population density on the map of China. The dispersion in population density is large, with eastern and coastal areas being more populated, and those regions are also those with higher incomes. For the vast western counties, the population density is much lower, with the number of people per square kilometer below 100.

Table A2 reports the regression output when we regress county-level gains from variety against the population density, income per capita, and distance to e-commerce hub Hangzhou. The table shows that rural and counties with lower incomes tend to derive larger gains from variety. Remote counties, measured by the distance to Hangzhou, tend to enjoy higher benefits from variety.

Table A2: Gains from Variety across Counties

	$\sigma_M = 1.9$, $\sigma_N = 1.5$			$\sigma_M = 1.8$, $\sigma_N = 1.8$		
	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	-0.365	-0.282	-0.249	-0.192	-0.145	-0.108
	(0.060)	(0.061)	(0.083)	(0.028)	(0.028)	(0.035)
log(distance to HZ)		0.616	0.600		0.345	0.311
		(0.108)	(0.151)		(0.049)	(0.064)
log(GDP p.c)			-0.636			-0.462
			(0.167)			(0.071)
R^2	0.021	0.039	0.042	0.027	0.054	0.075
Obs	1,733	1,733	1,307	1,733	1,733	1,307

Notes: This table shows regression output when regressing gains from variety on the logarithm of population density at the county level. In addition to population density, we consider income and distance to Hangzhou as explanatory variables. The number of counties declines to 1,307 due to the unavailability of county-level income data. Standard errors are shown below the estimates in parentheses.