The Alibaba Effect: Spatial Consumption Inequality and the Welfare Gains from e-Commerce

Jingting Fan, Lixin Tang, Weiming Zhu, Ben Zou*

December 2015

Abstract

Studies show that differential access to varieties of goods contributes to inequality in living standards across cities. By eliminating the fixed cost of entry for firms, e-Commerce might disproportionately improve smaller cities' access to varieties, and reduce this inequality. Using unique data from China's leading e-Commerce platform, we first document a negative relationship between online purchasing intensity and market size. We then build a multi-region general-equilibrium model to quantify the welfare gains from e-Commerce. With an average of 1.62 percent, the welfare gains are 0.94 percentage point higher for cities in the 1st population quintile than those in the 5th quintile.

Keywords: Spatial consumption inequality, Gains from e-Commerce, Economic integration.
JEL Code: R13 L81 F15

*Jingting Fan and Weiming Zhu, University of Maryland; Lixin Tang, Shanghai University of Finance and Economics; Ben Zou, Michigan State University. Contacts: Fan, fan@econ.umd.edu; Tang, tang.lixin@mail.shufe.edu.cn; Zhu, weimingzhu@rhsmith.umd.edu; Zou, benzou@msu.edu. We thank Ali Research for providing access to the data for this research. For helpful comments we thank Roger Betancourt, Jonathan Dingel, Nuno Limao, Asha Sundaram, and participants at the 2015 Midwest International Trade and Economic Theory Conference and the 10th Meeting of the Urban Economics Association. Any remaining errors are our own.

1 Introduction

Recent studies have documented that differences in prices and availability of consumption goods can be an important source of inequality in living standards across cities of different sizes (see, for example, Handbury and Weinstein, 2014; Couture, 2015; Hottman, 2014). One reason why cities do not offer the same bundle of consumption goods is the costs associated with serving new markets, including both the fixed cost of setting up stores and distributional channels, and the variable cost of shipping goods from producers to destination markets. Due to high shipping cost and low demand, the expected profit from entering small and remote markets might not cover the fixed investment, preventing firms from entering. Because of this, consumers in small markets might have access to a smaller number of varieties compared to their counterparts in big and well-connected cities, resulting in spatial inequality in consumption. This problem is likely to be more pronounced in developing countries, where, as a result of poor infrastructure, low productivity in the retail industry, and widespread entry regulations, firms might be even more reluctant to set up stores in small markets.[1]

This paper studies the extent to which the rise of e-Commerce reduces this spatial inequality in consumption. With e-Commerce, transactions are made online and goods are directly shipped to consumers. This eliminates the need to set up distributional network and brick-and-mortar stores, allowing firms to reach consumers in cities that otherwise would not be served. Since consumers in smaller cities had been more constrained in their consumption choices, they benefit disproportionately more from the improved access to varieties of goods. This force is gaining empirical relevance, thanks to the rapid growth in e-Commerce in the past decade. According to *emarketer.com*, a market research company in the retail industry, total e-Commerce sales in the US are projected to increase from $259 billion in 2010 to more than $400 billion by 2017. The online retail industry is expanding rapidly in the developing world as well.

In this paper we focus on China, which offers a suitable setting for studying the role of e-Commerce in alleviating spatial inequality in consumption. On the one hand, economic activities, including retailing, are very unevenly distributed in China. Large retailers are disproportionately concentrated in big cities, leaving many small cities and the vast rural area unserved.[2] On the other hand, in recent years e-Commerce has been growing rapidly in China. As shown in Figure 1, the total volume of sales in online markets increased by more than 25 times between 2008 and 2014. Online sales accounted for about 10% of total retail sales in 2014. Despite its large size and the staggering growth, the welfare effects of e-Commerce on consumers in different cities have not been examined. This paper aims to fill this gap.

Our study employs unique data from Taobao, a subsidiary of Alibaba Inc. and the dominant online marketplace that accounts for more than 80 percent of online retail sales in China. We first document a strong negative relationship between market size and the share of expenditures spent online in total

[1] Using bar-code level data, Atkin and Donaldson (2014) shows that the unit-distance shipping costs in a sample of African countries could be more than 5 times as high as that for within the US. Using a calibrated model, Lagakos (forthcoming) argues that due to the lack of complementarity assets such as home-owned cars, big-box retail chains with high labor productivity are less prevalent in poor countries.

[2] A recent joint study by National Bureau of Statistics of China and Ali Research shows that per-capita retail space of chain stores in cities on the east coast is four times as large as that in cities in the hinterland.

consumption expenditure shares (hereafter "online expenditure share"). Our analysis is at the city level.[3] The primary measure for market size is population. We show that the elasticity of online expenditure share with respect to population is -0.12. Everything else equal, the online expenditure share is 2.1% percentage points higher for residents in cities in the smallest quintile (with an average population of 0.8 million) than those in the largest quintile (with an average population of 8.1 million). The results are robust to controlling for a host of demand and supply shifters for online shopping, including cities' social-economic, demographic, and geographic characteristics, as well as access to broadband internet and transportation infrastructure. When we estimate this population elasticity separately by product category, we find that the elasticity is negative and statistically different from zero only for tradable goods, but not for locally-traded goods and services. This suggests that our results are unlikely to be driven by unobservable heterogeneity in taste for online shopping that is negatively correlated with city size.

Motivated by this empirical finding, we then develop a theoretical framework to evaluate the effect of e-Commerce on spatial inequality in consumer welfare. We begin with a simple model that illustrates the key insights. In the model, residents in all cities have constant elasticity of substitution (CES) preference for *one* online good and *one* offline good. While the price of the online good is the same everywhere, the price of the offline good differs across cities. The online expenditure share therefore contains information on the prices of the offline good in different markets, which we can extract through the model. Since in the data, smaller cities have larger online expenditure shares, the offline good in those cities must be less desirable. The availability of the online good reduces the inequality associated with the price differences of the offline good.

The simple model overlooks the fact that cities have differential access to the online market due to their geographic locations. More importantly, it rules out the possibility that firms may endogenously choose whether to serve a city and if so, whether to sell in the online or the offline market. Because of these firm-level decisions, e-Commerce affects consumer welfare not only directly through the online market, but also indirectly through its impacts on the offline market. To address these shortcomings, we build a multi-region general equilibrium model with realistic geographic features, endogenous firm entry and endogenous choice of distributional channels. If a firm chooses to serve a city through the offline channel, it needs to pay a fixed cost to set up a store. If it chooses to sell online, it saves the fixed cost, but incurs a higher variable cost for delivery. This higher variable cost can be interpreted as associated with individualized shipping and other inconveniences of doing business online. The trade-off between variable and fixed costs captures the spatial distribution of benefits from e-Commerce.

We calibrate the model to the Chinese economy by matching several salient facts on income, firm size, and online and offline sales across cities. The calibration of the model implies large differences in access to varieties of goods related to population sizes: the real income of an average city in the largest quintile is 95.0% higher than that of an average city in the smallest quintile.

The calibrated model is then used to perform counter-factual experiments. We calculate the welfare

[3] In this paper, a city refers to a prefecture in China. China is divided into 345 prefectures. A prefecture typically consists of several urban centers and the rural areas in between these urban centers.

gains of e-Commerce to each city by comparing an economy in which the online channel is shut down with the observed economy in which online sales account for about 8% of the total retail sales in an average city.[4] We find that the welfare gains for an average city is about 1.62%. Residents in smaller cities gain more than those in bigger cities: under the most conservative calibration, the average welfare gains are 1.21% for cities in the largest population quintile and 2.15% for cities in the smallest quintile. Therefore, e-Commerce can reduce the inequality across Chinese cities, although the reduction is small compared to real income inequality in the calibrated model.

We study the scope for future development of e-Commerce, modeled as further reductions in online delivery costs, to improve consumer welfare. Specifically, we reduce online delivery costs such that the overall online expenditure share is twice and three times of its 2013 level, respectively. These numbers are within the reach in the near future according to various estimates based on the rapid growth in the online retail industry. We find that the continued growth of e-Commerce will further benefit consumers in all cities and reduce the real income inequality across cities. For example, when the share of online expenditure is tripled, the cumulative gains from e-Commerce will be 3.0% for consumers in the largest quintile of cities, and 6.5% for those in the smallest quintile. Therefore, further development of e-Commerce can significantly reduce the inequality in living standards between cities.

Our paper contributes to a new and rapidly growing literature on consumption inequality across space. Recent studies show that there are substantial spatial differences in access to varieties.[5] This paper contributes to this literature in two aspects. First, it introduces the online shopping channel to a literature that has so far predominantly been focusing on traditional retailing. Since e-Commerce has become increasingly common throughout the world, studies that do not take this into account are likely to miss a large part of the differential access to varieties across space. This paper complements existing studies by quantifying the inequality-mitigation effects of e-Commerce. The second contribution is in terms of measurement. By focusing mostly on the US, the literature has so far neglected developing countries, where such spatial inequality is likely to be more severe. This may be explained by the lack of high-quality bar-code-level data covering a large sample of cities.[6] This paper uses a revealed-preference approach to measure the spatial inequality in access to consumption goods by combining online sales data with a model. To our knowledge, this is the first paper to perform this exercise.

This paper also contributes to the literature that studies the welfare effects of e-Commerce. Existing studies have identified specific channels through which e-Commerce benefits consumers.[7] Most of these studies do not have a spatial dimension, neither do they speak to the distributional effects of e-Commerce. An exception is Forman et al. (2009), which uses data from top-selling books from Ama-

[4]We arrive at the figure of 8% using the 2013 data on total sales on Taobao (See Section2.1).

[5]This includes studies on tradable good, especially groceries (Broda and Weinstein, 2006; Handbury and Weinstein, 2014; Handbury, 2013; Hottman, 2014) and non-tradable goods and services such as restaurants and local amenities (Couture, 2015; Diamond, 2015).

[6]Notable exceptions include Faber (2014) and Atkin et al. (2015), both of which use detailed category level price data from Mexico. These two papers, however, do not focus on spatial inequality.

[7]For example, online markets might offer products at lower prices (Brynjolfsson and Smith, 2000; Clay et al., 2002; Brown and Goolsbee, 2002), improve the match quality between consumers and products (Goldmanis et al., 2010; Glenn and Ellison, 2014), and increase the number of product varieties available to consumers (Brynjolfsson et al., 2003).

zon.com, and finds that when a store opens locally, people substitute away from online purchases. This study is supportive of the key channel highlighted in our paper. However, with a focus on books, their results might not be generalizable to the overall welfare gains from e-Commerce. We provide a more complete picture by using a comprehensive dataset on online sales. Additionally, Forman et al. (2009) does not explicitly model the entry decision and the channel choice by firms and therefore cannot quantify the general equilibrium welfare effects, which might be important given the size of the industry. We incorporate firms' decisions as well as the general equilibrium effects in our quantitative framework. In doing so, we contribute to this literature by proposing a tractable quantitative model that allows realistic geographic features and is thus amenable to data.[8]

A few papers have tried to measure the overall impact of e-Commerce by using information on internet access and usage. For example, Goolsbee and Klenow (2006) examines the value of the internet to consumers based on time spent on computers; Tran (2014) studies the impacts of internet access on offline retailers. These studies do not have direct measures of online consumption. To the best of our knowledge, this paper is the first to use a comprehensive online sales dataset from a leading e-Commerce platform that allows us to measure the aggregate welfare gains.

2 Background and Data

2.1 Retail Industry and e-Commerce in China

E-Commerce has been growing rapidly in China. As shown in Figure 1, between 2008 and 2014, the total value of online retail sales has been growing at an annual rate of 64 percent. Online retail sales as a share of total retail sales increased from 1.1 percent in 2008 to 10 percent in 2014. Taobao has been the dominant online retail platform in China, accounting for 82 percent of the total online retail sales.[9]

The rise of e-Commerce in China has an important geographic dimension. Although bigger and richer cities on the East Coast have the highest per capita online retail consumption, it is the smaller and less developed cities that spend a higher proportion of their income online. According to a survey performed by the McKinsey Global Institute, while residents in tier-1 and tier-2 cities, including China's largest and most internationalized cities, spend 18 percent and 17 percent of their disposable income online, residents in much smaller tier-3 and tier-4 cities spend 21 percent and 27 percent, respectively.

China's relatively less developed offline retail industry helps explain both the rapid rise of e-Commerce and its spatial distribution. The offline retail industry in China is dominated by small-scale local retailers which provide limited numbers of varieties—the top 100 retail chains in China collectively accounted for only 9 percent of retail sales in 2012. Furthermore, existing large retailers are traditionally concentrated

[8]Our model exploits the similarity between online sales and exports. Some earlier studies have shown this similarity. Hortaçsu et al. (2009) documents that distance continues to matter in online purchase, and there is also evidence of home-market effect; Lendle et al. (Forthcoming) looks at international purchase on eBay, and find that the effects of distance to be smaller than the estimates from the off-line gravity estimation; Lendle et al. (2013) documents the relationship between firm size distribution and exporting behavior, in a way that is similar to Eaton et al. (2011). None of the existing studies connect online sales with urban consumption inequality and examine the welfare impacts.

[9]This number includes sales on Taobao.com, a mainly C2C marketplace, and Tmall.com, its smaller B2C platform.

in large and rich cities. For example, when foreign retail chains such as Walmart and Carrefour entered China, their first stores were predominately in the biggest metropolitan areas. This is in sharp contrast to the retail industry in developed countries. In the US, for example, the top 100 retail chains account for about 35% of total retail sales. Those retail chains penetrate many small to medium cities and suburban areas. Walmart, for one, established its roots in mid-size cities.

The different spatial development of traditional retailing in China might be related to the costs of setting up and operating retail stores. Setting up a large bricks-and-mortar store requires an upfront investment that can only be justified by a sufficiently-large local demand. With per capita income in China still a fraction of that in the developed countries and costs of doing business high, it remains unprofitable for retail chains to enter many smaller, less prosperous markets. In absence of large chain stores to connect producers with consumers, in order to reach consumers in small markets, producers have to establish connections with local retailers or set up their own storefronts, which can be prohibitively costly for many producers.

E-Commerce is thus a more economical way for many producers to reach customers in small markets. By signing up with large e-Commerce platforms such as Taobao, a producer might find it profitable to sell a small number of items to a small number of consumers in markets that it would not find profitable to enter via the offline channel. Shopping online is therefore particularly attractive for residents in small cities because it can significantly increase their access to varieties. Indeed, in a survey reported in Dobbs et al. (2013), when asked about what attracts them to buy online, 55 percent of the respondents in tier-3 cities cited "access to varieties" as a main reason, compared with 31 percent in tier-1 cities, and 44 percent in tier-2 cities.[10]

2.2 Data and Sample

The data used in this paper comes from several sources. The first piece of data is from Taobao. Taobao is a subsidiary of Alibaba catering to retail consumers and includes both a B2C platform (tmall.com) and a C2C platform (taobao.com). Both platforms sell hundreds of millions of unique products and taobao.com is the world's 11th most visited website.[11] We obtain confidential data of the total sales and purchases by city-category in 2013 from both platforms.[12]

We supplement the sales data from Taobao with data on city-level characteristics. Specifically, from the 2010 census tabulations and te 2013 Regional Statistical Yearbook, we construct variables related to factors affecting the demand and supply of online shopping. These factors include city-level demographic characteristics such as education, age, and gender composition, income and consumption levels, industrial composition, and shares of households with broadband connection and smartphones. We construct the distance to highways and railways of each city from the transportation network database

[10] Ali Research and McKinsey Global Institute. http://i.aliresearch.com/img/20150324/20150324203753.pdf, Page 27.

[11] http://www.alexa.com/siteinfo/taobao.com#trafficstats, visited on 12/15/2015.

[12] The categories are defined by Taobao. We start with 139 product categories and consolidate them into 81 product categories. The original list of categories can be found here: https://list.taobao.com/browse/cat-0.htm. The consolidated list of categories are reported in Table A.

developed by Baum-Snow et al. (2015).[13] We obtain other city-level geographic characteristics, such as the longitude, latitude, and ruggedness of a city from the China Historical GIS database. We also obtain a database of housing price from a leading real estate market research company in China.

Our benchmark regression is performed at the city level. To test the channel behind the correlation between market size and online expenditure share, we also use city-category level expenditure. For this, we construct city-by-category expenditure shares from China's Urban Household Survey (UHS).[14] The product categories in UHS are different from those used by Taobao. We manually match Taobao categories to UHS categories.

Our baseline sample includes 337 cities.[15] Table 1 provides the summary statistics of the key characteristics. In a typical city in 2013, Taobao accounted 7.9 percent of total retail sales. An average city had a population of 2.7 million (natural log value is 14.8) and per capita income of 13,780 *yuan* (2,153 US dollars).

3 Market Size and Online Expenditure Share

3.1 Baseline Results

Figure 2 provides a first look at the correlation between market size and online expenditure share. With each blue solid dot representing a city, the figure shows a clear negative correlation between log population and log online expenditure as a share of total retail sales. The slope of the bold fitted line is -0.116 with a standard error of 0.026.

To test this correlation more formally, we run the following regression:

$$\ln OnlineShare_i = \beta_0 + \beta_1 \ln Pop_i + \mathbf{X}_i \cdot \beta_2 + \epsilon_i, \tag{1}$$

where $OnlineShare_i$ is the total expenditure on Taobao as a share of city i's total retail sales, Pop_i is the city's population in 2010, $\mathbf{X}_i$ is a vector of city-level characteristics, and ϵ_i is the error term.[16] In the baseline, $\mathbf{X}_i$ includes basic information such as log average income level, average educational attainment, urbanization rate, and in some specifications province fixed effects. In the next subsection we add more covariates to $\mathbf{X}_i$ to highlight our preferred interpretation of the mechanism.

Table 2 shows the baseline results. Column 1 reproduces the simple correlation as presented in Figure 2. The coefficient of -0.116 implies that for each percent increase in population, the share of online

[13] To obtain city-level average distance to railroads/highways, we compute the minimum distance to railroads/highways for each *county* within the city, and then use the population-weighted average county-level distance as the city-level distance.

[14] The UHS is conducted by the Bureau of National Statistics of China, and it tracks a panel of households in different cities for their daily expenditures. The weight of households is only representative at the provincial level, and the category-level per capita expenditures are only available by province. We use province-category-level expenditures, together with predictors including average income and demographics to impute the category-level expenditures for each city. We provide details on this procedure in Section 3.3.

[15] 8 cities are dropped due to missing values in baseline covariates.

[16] The results are the same, after conversion, when we use $OnlineShare_i$ as the outcome variable. We choose to use a log-log specification mainly to be consistent with the category-level analysis introduced later. In the category-level analysis, the coefficient in the log-log specification is unit free which makes comparison across categories possible.

purchases drop by 0.116%. When log population increases by one standard deviation from the mean, the average online expenditure share decreases from 7.9% to 7.1%. Alternatively, everything else equal, residents in cities in the smallest quintile (with an average population of 84 thousand) spends 2.1% percentage points more in terms of online expenditure share, compared with those in the largest quintile (with an average population of 8.1 million).

Columns 2 to 4 add log per capita income, log average years of education, and log share of employment in the non-agricultural sector to the simple regression, respectively. When included in the regression separately, each of these additional control variables has a positive coefficient. The magnitude of the coefficients associated with log population become a little larger than that in Column 1, although the differences are not statistically significant. When the additional controls are included in the regression jointly (Column 5), the coefficient of interest becomes slightly smaller than that in Column 1 (again, the difference is not statistically significant). While the three control variables are highly correlated, it appears that the share of non-agricultural employment is the best predictor for online shopping. Column 6 includes a set of province fixed effects, so that the coefficient is estimated by comparing cities with different sizes within the same province. The R-squared increases substantially compared with that in Column 1, while the coefficient associated with log population does not change much. Finally, we may worry that log population is measured with error so that the coefficient is biased towards zero. In Column 7 we instrument for log population using the log of lagged population from the 2000 census. The coefficient barely budges. Therefore, we use the OLS estimate in Column 6 as our baseline specification.

3.2 Robustness

3.2.1 Additional Covariates

In Table 3 we further examine potential confounding factors that can explain this negative correlation between population and online expenditure share, by adding additional covariates that are related to alternative explanations. All columns include the same set of covariates as in Column 6 of Table 2.

Since both e-Commerce and traditional retail depend crucially on transportation, it is ambiguous how connectivity to transportation infrastructure affects the share of retail consumption spent online. Column 1 of Table 3 includes population-weighted distances to the nearest highway and railway. It turns out that being close to a railway is negatively correlated with online expenditure share while distance to a highway is not very important. The coefficient associated with log population remains quantitatively similar.

Online shopping needs access to internet. In Column 2 we include log percentage of households with broadband internet connections. Internet penetration is positively correlated with online expenditure share. The coefficient associated with log population drops by about 20%, but it remains highly statistically significant and economically meaningful.

Column 3 includes cities' average housing prices. High housing price increases the cost of running a bricks-and-mortar store, and might increase the share of online purchase. Since housing price is likely correlated with population, it is important to purge out its effects. We have to drop 49 cities due to the lack of housing data. Our regression suggests that housing prices indeed have a strong positive correlation with the online expenditure share, and its inclusion increases the coefficient associated with population significantly.

In Column 4, we include a dummy indicating whether a city is the provincial capital. Given the hierarchical system of Chinese cities, the provincial capital is often far more developed than other cities in the same province. The provincial capital dummy is likely to capture other social economic performances in addition to those we are able to control for. The positive coefficient for the provincial capital dummy, though only marginally significant, is consistent with the hypothesis that overall more developed cities purchase more online. More importantly, the coefficient associated with log population does not change, suggesting that the effect is not driven by the difference between capital and non-capital cities.

In Column 5, we include the log distance-weighted access to sellers on Taobao. Specifically, for city i, this term is defined as $ln(\sum_j S_j/d_{ij})$, where S_j is the total value of goods sold on Taobao from sellers in city j, and d_{ij} is the distance between city i and city j (d_{ii} is set to be one). This term intends to capture the access to the online market. Intuitively, cities that are far from sellers are less likely to purchase online because the shipping cost is likely to be higher. We find that the log inverse distance is positively correlated with online expenditure share, confirming the intuition. Furthermore, the coefficient associated with log population becomes larger in magnitude. This is because cities farther away from major Taobao seller cities also tend to be smaller.

In Column 6 we include all the additional covariates from Column 1 to Column 5. The coefficient of interest remains significant and in a similar magnitude (note that we lose about 15% of the observations). Finally, although not reported here, we also experimented with an array of demographic, socio-economic, geographic controls in flexible functional forms. Similar to our findings in Table 3, the coefficient associated with log population remains stable.

3.2.2 Alternative Measures

Our benchmark specification uses the log of the online expenditure as a share of total retail sales as the dependent variable, and uses population as a proxy for market size. We test the robustness of the results to alternative measures for online purchasing intensity and proxies for market size. The first 3 columns of Table 4 experiment with alternative measures of market size. Column 1 uses log population and replicates the baseline results. Column 2 uses log GDP as a proxy for market size, while Column 3 instead uses log total consumption. All three columns yield negative and statistically significant coefficients associated with the market size measure, and the magnitudes are largely comparable.

Column 4 and Column 5 use alternative measures of online purchasing intensity. Specifically, Column 4 uses total consumption instead of total retail sales as the denominator in the online share formula. This measure helps alleviate the concern that total retail sales might be measured with noise, as it is often difficult to classify whether a purchase is for final consumption or resale. Column 5 uses total non-hospitality retail sales, which excludes sales by local restaurants and hotels, as the denominator. This measure addresses the concern that the estimated coefficient might be negative because residents in bigger cities spend more money on local services such as dining out. Again, the coefficients associated with log population remain statistically significant and similar in magnitude.

3.3 Alternative Explanations

While we have controlled for a vector of variables that intend to capture demand shifters for online shopping in a flexible way there might still be concern that the negative population elasticity we have documented may be driven by unobservable heterogeneous preference: residents in small cities may simply prefer online shopping. Our product-category-level data offer a potential test for this alternative explanation. If it is due to unobserved preference that residents in smaller cities have a higher tendency to shop online, such tendency should hold for all product categories. On the contrary, if it is the limited access to varieties in small cities that prompts consumers to shop online, the negative correlation should show up only in the categories that are tradable across cities, but not for locally-traded goods and services.

We run Equation 1 by category using the same specification as in Column 6 of Table 2. Our data cover all 81 categories on Taobao. We measure category-level online expenditure share as the total online purchases in each category as a share of the *aggregate* city-level retail sales. Our log-log specification renders the coefficient to be neutral to the relative size of each category, so a comparison of coefficients across categories is possible.

The estimated coefficients and standard errors associated with log population for all 81 regressions are reported in Appendix Table A. The categories are ranked in the order of the population elasticity (from the very negative to positive). The coefficients are negative and statistically significant for most categories. The categories that have the highest population elasticities are (1) basic building materials, (2) motorcycle parts and accessories, and (3) desktop computers and servers. In contrast, service and locally-tradable goods, which we mark with a flag, mostly have statistically insignificant or even positive coefficients.

These results show that heterogeneous preference for online shopping is unlikely to have driven our results. However, there is still another alternative explanation: residents in bigger cities, who are in general richer than their smaller city counterparts, spend a larger share of their income on non-tradable goods such as services. If this is the case, then even if residents in all cities have the same preference for online shopping, because most products sold on Taobao are tradable, there would be a negative correlation between city size and our measure of online expenditure share, which uses the *aggregate* consumption as the denominator.

We believe this non-homotheticity in preference is unlikely to be driving our results for two reasons. First, we already control for a rich array of covariates such as income, education, and urbanization rate, which capture the differences in consumption structures. Our results are robust to these controls for taste shifters. We also show in Table 4 that our result is robust to using an alternative measure of total consumption that excludes sales at restaurants and hotels. Second, if there is still correlation between consumption structure and city size intermediated through income that eludes from our rich set of controls, we should expect positive signs for tradable goods that are consumed mainly by high-income households. An inspection of Table A suggests this is not the case. For example, "musical instruments and parts" and "sports, yoga & bodybuilding supplies," two categories that are probably most likely to be consumed by high-income household by the standard of China, have an elasticity of -0.141 and -0.143, respectively.

To further alleviate this concern, we perform an additional test at the city-category level, using category-specific online expenditure share as the dependent variable(that is, $\ln(\frac{\text{Online Purchase}_{ih}}{\text{Total Consumption}_{ih}})$ as the dependent variable for city i and category h). We construct the denominator using category-leve expenditure share from the Urban Household Survey(UHS). Since the UHS information is at province level, we impute the relevant expenditure shares for each city based on observable province- and city-level characteristics.[17] Since the categorization in the UHS is different from that on Taobao, we manually match Taobao categories to UHS categories to construct the numerator. The dependent variable constructed this way explicitly takes into account the differences in consumption structure. The last row of the Table 5 shows the mean log shares of online expenditure for each category. Not surprisingly, categories with the largest online expenditure shares are "clothing" and "household appliances and service".

Table 5 reports the results for separate regressions for each category, using the same specification as in Column 6 of Table 2. We find negative and sizable coefficients associated with log population for both "clothing" and "household appliances and service" categories. We find small, insignificant, and even slightly positive coefficients for "food," "residence-related," "health-related," and "recreation, education, and culture" categories. These categories mostly include non-tradable or locally-traded goods and have only small shares sold online.[18]

Overall, our empirical analysis finds a robust negative relationship between online expenditure share and market size, which does not seems to be driven by common alternative explanations. What does this negative correlation tell us about the inequality in access to consumption goods across cities? And how does it translates into differential gains from the e-Commerce for residents from cities of different sizes? The next section develops a simple model to illustrate the information revealed by online expenditure share. Section 5 presents a more quantitative approach.

[17] The characteristics include average income, average years of education, and share of non-agricultural employment. The imputation proceeds in four steps. In the first step, we regress per capita expenditure on the predictors for each category at the province level. In the second step, we calculate the differences of these predictors between each city and the average levels of its province. In the third step, we calculate the differences of the expenditures by each category between a city and its province average by multiplying the estimated coefficients from the first step and the differences in predictors from the second step. Finally, city-category expenditures are recovered from province averages and city-specific differences obtained in the third step.

[18] Although its online expenditure share is also small, we find a negative and statistically significant coefficient for the "transportation and communication" category. We postulate that this is because residents in smaller cities, due to a lack of local agents resulting from the small market size, rely more on travel agents on Taobao to buy tickets or recharging cellphone credits. This result is consistent with the channel we highlight in this paper.

4 A Simple Model

Suppose consumers in each city i can purchase two goods, one sold online and the other sold in the local offline market.[19] Consumers haves exogenous taste for these two goods. We assume the utility for consumers living in city i is given by the following CES function

$$U_i = (q_{ii}^{\frac{\sigma-1}{\sigma}} + a_o q_{io}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where q_{ii} and q_{io} are the consumption of goods from the local and the online market, respectively; $\sigma > 1$ is the elasticity of substitution between the online and local goods; a_o is the preference for online good. Since we have shown that the negative relationship between online expenditure share and market size are not driven by differential preference for online shopping, we assume a_o to be constant across cities. This assumption is critical for us to infer the cross-sectional inequality in access to consumption goods in offline markets, but will not affect our measurement of gains from e-Commerce.

Let w_i be the income of consumers in city i, p_{ii} be the price of local good in city i, and p_{io} the price of online good. We assume w_i, p_{ii} and p_{io} to be exogenous in our partial-equilibrium analysis. In the special case that $p_{io} = \infty$, consumers spend all of their income on the local good and there is no e-Commerce. When p_{io} is finite, the optimal consumption decision satisfies the following equation:

$$\frac{a_o q_{io}^{-\frac{1}{\sigma}}}{q_{ii}^{-\frac{1}{\sigma}}} = \frac{p_{io}}{p_{ii}} \tag{3}$$

Using λ_{io} to denote the share of expenditures spent on the online good, it is straightforward to show that

$$\frac{\lambda_{io}}{1-\lambda_{io}} = a_o{}^{\sigma} (\frac{p_{io}}{p_{ii}})^{1-\sigma}. \tag{4}$$

If consumers in city i spend a larger share of expenditure on the online good than consumers in city j, it follows that $\frac{p_{ii}}{p_{io}} > \frac{p_{jj}}{p_{jo}}$. Assuming that $p_{io} = p_{jo} = p_o$ for simplicity, we have $p_{ii} > p_{jj}$, that is, the offline good in city i is more expensive. The intuition is simple: if consumers in a city choose to spend less on the local good relative to the common online good, the local good must be less attractive in that city.[20]

To evaluate the gains from e-Commerce, we hold w_i and p_{ii} constant and evaluate the welfare of consumers in City i while changing the value of p_{io}. Let P_i be the price index for city i, then $P_i^{1-\sigma} =$

[19] In this simple model, we assume that consumers do not value local good produced in other cities, and we are agnostic about the source of the online good.

[20] Of course, without a micro foundation on why consumers choose to purchase goods from online or local markets, this simple model cannot be used to evaluate what "$p_{ii} > p_{jj}$" entails. For example, it could be that there are only limited number of varieties or these varieties are very expensive. In addition, the model is silent on where the goods sold online is from. In the quantitative model in the next section, we take a stand on what drives e-Commerce and use that model to evaluate spatial inequality as implied by online expenditure share, and the gains from e-Commerce.

$p_{ii}^{1-\sigma} + a_o^\sigma p_{io}^{1-\sigma}$, and the welfare of a consumer in city i is:

$$u_i = \frac{w_i}{P_i} = \frac{w_i}{(p_{ii}^{1-\sigma} + a_o^\sigma p_{io}^{1-\sigma})^{\frac{1}{1-\sigma}}}. \tag{5}$$

In the absence of e-Commerce, $p_{io} = \infty$, $P_i = p_{ii}$ and $u_i = \frac{w_i}{p_{ii}}$. The gains from e-Commerce, defined as the percentage change in welfare as a city move from the case without e-Commerce ($p_{io} = \infty$) to the case with e-Commerce (p_{io} is finite), is given by

$$\text{Gains from e-Commerce} = (1 - \lambda_{io})^{\frac{1}{1-\sigma}} - 1. \tag{6}$$

The welfare gain is an increasing function of the online expenditure share, λ_{io}. This result is consistent with the revealed-preference intuition that, if consumers purchase more from the online market, they must benefit more from doing so.

We can use Equation 6 to gauge the effects of e-Commerce on spatial consumption inequality. The average of online expenditure share across the cities in our sample is about 8%. Setting $\sigma = 4$, which is consistent with the estimates by Jensen et al. (2003), the average gains from e-Commerce are $(1 - 0.08)^{-\frac{1}{3}} - 1 = 2.8\%$.[21] Now consider a city with an online expenditure share of 12.3%, which is one standard deviation above the average. The welfare gains for consumers in this city are $(1 - 0.123)^{-\frac{1}{3}} - 1 = 4.4\%$. This value is more than 50% higher than that of the city with the average online expenditure share. Therefore, there is large spatial dispersion across cities in the gains from e-Commerce.

In an influential paper, Arkolakis et al. (2012) show that the insight in this simple model is robust: for a large class of trade models, the gains from international trade can be obtained by two statistics, the share of total expenditure on domestic goods and a trade elasticity.[22] Despite this generality, there are important drawbacks to directly using the simple model to study the welfare effects of e-Commerce. First, the ACR formula is only useful, conditional on the change in the important import penetration ratio ("online expenditure share" in our application) of individual countries ("cities" in our application). The ACR formula cannot be used to make ex-ante evaluation on policy changes, and hence we cannot perform welfare analysis for future growth in e-Commerce.[23] Second, the simple model is not informative on the specific *channels* of welfare gains of e-Commerce. For example, by assuming exogenous local prices and wage, the simple model cannot be used to study how local prices and entry respond to competition from online. Lastly, when we account for the fact that firms can choose to sell both offline and online, the model economy may deviate substantially from the ACR world, and the ACR formula may not apply. We address these drawbacks in the next section using a fully-fledged model.

[21]Note that σ is close to the lower end of the range given by the literature. If a larger value is used, the welfare gains from trade would be smaller. We adopt this value because this is estimated using the US manufacturing data in a setting with heterogeneous firms, which is consistent with the framework we are to develop in the next section.

[22]Costinot and Rodríguez-clare (2013) shows that this formula can be further generalized to allow for multiple sectors and input-output linkages.

[23]In our counter-factual experiments in our full-fledged model in the next section, we target the overall online expenditure share to calibrate a reduction in e-Commerce cost. However, the ACR formula has nothing to say on how the online expenditure share and the real income of *individual* cities would change in response to such a reduction in e-Commerce cost.

5 The Quantitative Framework

The fully-fledged model adapts Helpman et al. (2004) (henceforth HMY). There are N cities, corresponding to a prefecture city in the data, indexed by $i \in \{1, 2, \cdots N\}$, each with population L_i. Each worker in city i supplies one unit of labor inelastically and receives a wage of w_i, which is determined endogenously. Workers' sole source of income is their wages, and they spend all of their wages on consumption good, so total demand in city i is $w_i L_i$.

5.1 Preference

There are 2 sectors, each indexed by $h \in \{T, NT\}$ where N and NT denote the tradable sector and the non-tradable sector, respectively. The preference of a worker in city i is given by

$$U = u_{NT}^{\beta_{NT}} u_T^{\beta_T} \tag{7}$$

where β_{NT} and β_N are preference parameters and

$$u_h = \Big(\sum_{v \in \Omega_h} q(v)^{\frac{\sigma_h - 1}{\sigma_h}} \Big)^{\frac{\sigma_h}{\sigma_h - 1}}, h \in \{T, NT\}. \tag{8}$$

In Equation (8), $q(v)$ is the quantity of variety v in the consumption bundle, Ω_h is the set of all available varieties in sector h from all cities, and σ_h is the elasticity of substitution for sector h. The constant elasticity of substitution (CES) formulation of preference conveys the idea that workers value the varieties of goods within a sector. We omit the index h where doing so does not cause confusion.

5.2 Production and Trade

To enter into production in sector h in city i, a potential entrepreneur pays the fixed cost of entry f_E^h, measured in labor units. Upon entry, the firm receives a cost draw a from a distribution $G(a)$, which enables the firm to use a units of labor to produce one unit of a differentiated good. So the marginal cost of production for a city-i firm with a draw a is aw_i. Following HMY and a vast literature in international trade, we assume the draws follow a Pareto distribution, that is, $G(a) = Pr(\tilde{a} \geq a) = 1 - (\frac{a}{\bar{a}})^{k_h}, 0 < a < \bar{a}$ where $k_h > 1$ is the Pareto shape. We assume the same $G(a)$ distribution for all cities in the calibration exercise.

Upon receiving the productivity draw, firms engage in monopolistic competition. It can sell to any city, including its home city, using one of two different trading technologies, offline retail or online platform. To save notations, we assume firms in the non-tradable sectors also use these two technologies, with the only difference being that the inter-city trade costs are infinite for the non-tradable sector.

5.2.1 Offline Retail

A firm in city j can sell to consumers in another city i, $i \in \{1, 2, \cdots N\}$, (including the home city j) by incurring f_{off} units of labor at the destination as the fixed cost of entry. This term captures the cost associated marketing, site selection, building distributional network, as well as the fixed operation costs of running a store. Alternatively, this cost can be interpreted as the cost of finding a local retailer that can sell the products on behalf of the firm.

Additionally, to sell one unit of the differentiated good in city i, a firm in city j has to ship $\tau^{ji} \geq 1$ units, as is standard in the trade and economic geography literature. It captures not only the transportation costs, but also the other costs that are dependent on quantity. We assume $\tau^{jj} = 1$ for any j so that there is no variable cost in selling the good to the local market.

5.2.2 Online Retail

A firm in city j can also choose to sell to consumers in city i via e-Commerce. E-Commerce allows firms to sell to consumers in other markets without building an offline store or finding a local retail partner. Therefore, there are no fixed costs associated with entry into a new market. Instead, firms can display their product to consumers online. If a purchase is made, the firm can deliver the product through mail service.

In order to sell one unit of the differentiated good in city i, a firm in city j has to ship δ^{ji} units of the differentiated goods. It is convenient to define $e^{ji} = \frac{\delta_{ji}}{\tau_{ji}}$, which is the shipping cost wedge associated with e-Commerce. We assume $e^{ji} > 1$ for any city pair. The variable cost of delivering a product is higher if it is sold online than if it is sold through the offline channel, for two reasons. First, while firms exploit economy of scale by shipping in bulk when they sell goods through brick-and-mortar stores, shipments of goods purchased online tend to be smaller and consequently the unit cost of shipping for online purchases is larger. Second, there might be more severe information friction regarding product qualities in the online market, which hinders online transactions. We capture the information friction with our assumption that $e^{ji} > 1$.

5.3 Demand functions and Firms' Channel Choices

The demand function for variety v of sector h in city i is given by

$$q(v) = A_{ih} p(v)^{-\sigma_h} \tag{9}$$

where the demand shifter A_{ih} is given by $A_{ih} = \frac{\beta_h w_i L_i}{(P_{ih})^{1-\sigma_h}}$. In turn, the city-sector price index P_{ih} is given by $P_{ih} = (\int_{\Omega_{ih}} p(v)^{1-\sigma_h} dv)^{\frac{1}{1-\sigma_h}}$, where Ω_{ih} is the set of sector h varieties in city i.

Consider a firm with cost a in city j pondering to sell to city i. Optimal pricing implies that price is $p = \frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ij}$ if the firm opts to sell offline, and $p = \frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ji} e^{ji}$ if the firm opts to sell online. The expected profits from offline and online sales are, $\frac{1}{\sigma_h} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a \tau^{ji} w_j)^{1-\sigma_h}$ and $\frac{1}{\sigma_h} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a \tau^{ji} e^{ji} w_j)^{1-\sigma_h}$, respectively.

A firm from city j decides to sell to city i through offline retail entry if $a \leq a_{\text{off}}^{ji}$, where the unit-cost cutoff a_{off}^{ji} is given by

$$a_{\text{off}}^{ji} = \frac{\sigma_h - 1}{\sigma_h} \frac{1}{w_j \tau^{ji}} \left(\frac{A_{ih}(1 - {e^{ji}}^{1-\sigma_h})}{\sigma_h f_{\text{off}} w_i} \right)^{\frac{1}{\sigma_h - 1}}. \tag{10}$$

Intuitively, because of the fixed costs associated with offline retail entry, only the more productive firms which benefit more from the lowered shipping costs will choose to do so; smaller and less productive firms will only sell online. Firms with a draw a_{off}^{ji} are indifferent between the two modes.

5.4 Free Entry and Labor Market Clearing Conditions

A key equilibrium condition of the model is the free entry condition, which states that the cost of entry is equal to the expected profit from sales in both the home city and elsewhere. The zero profit condition for potential entrants in sector h of city j is given by

$$\begin{aligned} f_E w_j = \sum_{i \in \{1,2,\cdots,N\}} \Big(& \underbrace{\int_0^{a_{\text{off}}^{ji}} [\frac{1}{\sigma_h} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ji})^{1-\sigma_h} - f_{\text{off}} w_i] dG(a)}_{\text{Expected Profit from offline sales}} + \\ & \underbrace{\int_{a_{\text{off}}^{ji}}^{\bar{a}} \frac{1}{\sigma_h} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ji} e^{ji})^{1-\sigma_h} dG(a)}_{\text{Expected Profit from online sales}} \Big) \end{aligned} \tag{11}$$

Lastly, the labor market clearing condition specifies that the sum of labor demand across sectors is equal to the exogenous labor supply in that city, as given by

$$\sum_h L_{jh} = L_j. \tag{12}$$

Total labor demand from sector h in city j, L_{jh}, is in turn given by

$$\begin{aligned} L_{jh} =& M_{jh} f_e^h \\ &+ \sum_{i \in \{1,2,\cdots,N\}} M_{jh} \Big(\underbrace{\int_0^{a_{\text{off}}^{ji}} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ji})^{-\sigma_h} a \tau^{ji} dG(a)}_{\text{Variable labor demand from serving markets offline}} + \underbrace{\int_{a_{\text{off}}^{ji}}^{\overline{a}} A_{ih} (\frac{\sigma_h}{\sigma_h - 1} a w_j \tau^{ji} e^{ji})^{-\sigma_h} a \tau^{ji} e dG(a)}_{\text{Variable labor demand from serving market online}} \Big) \\ &+ \underbrace{f_{\text{off}} \sum_{i \in \{1,2,\cdots,N\}} M_{ih} (1 - G(a_o^{ij}))}_{\text{Fixed labor demand from firms serving the home city}}, \end{aligned} \tag{13}$$

where M_{jh} is the mass of entrants from sector h in city j.

5.5 The Equilibrium

The equilibrium of this economy is defined as a set of prices w_i, P_{ih}; quantities, M_{ih} ; and decision rules α_{off}^{ji}, such that that given exogenous parameters, the following conditions are satisfied:

1. Firms' choice of distributional channel is optimal (Equation 10).
2. The equilibrium price in both sectors of any city is consistent with the distribution of wages and firms' channel choice.[24]
3. Zero profit condition in both sectors of any city (Equation 11).
4. Labor market clears in each city (Equation 12).

5.6 Discussion on Model Assumptions

We make several simplifying assumptions in the model. Here we discuss how these assumptions might affect the interpretation of the model and the results. First, our model abstracts from retailers. In reality, however, firms do not always set up their own storefronts when they enter a city offline. Instead, they might rely on retailers in reaching their customers. Our assumption is motivated by the limited importance of big retail chains in China. Due to the underdevelopment of big retail chains, if firms do not enter offline themselves, they usually have to establish relationship with regional or local retailers. The costs incurred during this process is captured in the model as the fixed entry cost, and we can interpret an offline subsidiary of a firm as the combination of the firm and a local retailer.

Second, in the model, we assume that consumers have CES preference, and firms compete monopolistically. This simplification allows us to focus on the effects of e-Commerce by providing more varieties of consumption goods to consumers, which, as discussed in Section 2.1, appears to be an important channel in China. However, these two assumptions together also imply fixed markups and rule out any pro-competitive effects of e-Commerce—when facing the competition from online sellers, local sellers might respond by reducing their product markups. This channel is likely to be empirically important, given the recent findings by Hottman (2014). We abstract from it primarily because we are unable to separate the pro-competitive effects and the gains from varieties given the data constraint. Recent studies in international trade, such as Arkolakis et al. (2015) and Edmond et al. (2015), suggest that, while the commonly-used alternative models that generate pro-competitive effects will imply different *compositions* of the gains from trade, they will not predict differently on the *magnitude* of the gains from trade,

[24] Specifically, the sector-specific price index in each city can be expressed as a linear function of the mass of entrants in each city and their cutoffs for firm entry through offline stores, as given by

$$
\begin{aligned}
(P_{ih})^{1-\sigma_h} &= \int_{\Omega_{ih}} (p(v))^{1-\sigma_h} dv \\
&= \sum_{j\in\{1,2,\cdots,N\}} M_{jh}\kappa_h (w_j \tau^{ji})^{1-\sigma_h} \Big({a_{\text{off}}^{ji}}^{1-\sigma_h+k_h} + {e^{ji}}^{1-\sigma_h} (\bar{a}^{1-\sigma_h+k_h} - {a_{\text{off}}^{ji}}^{1-\sigma_h+k_h}) \Big),
\end{aligned} \tag{14}
$$

where $\kappa_h = \frac{1}{-\sigma_h+k_h+1} (\frac{\sigma_h}{\sigma_h-1})^{1-\sigma_h} \frac{k_h}{\bar{a}^{k_h}}$ is a *sector-specific* constant.

compared to models without pro-competitive effects. Therefore we conjecture that this simplification is unlikely to substantially change the magnitudes of the gains from e-Commerce in this paper.

6 The Effects of e-Commerce on Welfare and Inequality

6.1 Calibration

We calibrate the model to the Chinese economy to study the effects of e-Commerce on welfare. We feed the data on population and bilateral characteristics (such as the bilateral distance and whether the two cities belong to the same region) of all 337 cities in our sample into the model and calibrate the other parameters to match the salient features of the Chinese economy.

The parameter β_{NT} governs the share of non-tradable goods in total expenditure in the economy. We set $\beta_{NT} = 0.35$ to match the share of the service sector in the economy.[25] We set the elasticity of substitutions $\sigma_h = 4$ for both the tradable sector and the non-tradable sector, which is consistent with the estimates by Jensen et al. (2003) using US manufacturing data and is also the value used by Melitz and Redding (2015). We set the Pareto shape parameter $k_h = 4.27$ following Melitz and Redding (2015). The scale parameter for the productivity distribution $\bar{a}$ corresponds to a choice of unit in which to measure productivity and we set $\bar{a} = 10$ for both sectors.

We use a joint calibration procedure to calibrate the rest of parameters, including the fixed cost of entry f_E, the fixed cost of offline retail f_{off}, variable trade costs τ_{ij} and e-Commerce trade costs δ_{ij}, to match other important moments.[26]

We set $f_E = 9.3$ so that the average employment size of a firm is 40, which is the average size of manufacturing firms in the 2004 Chinese economic census. We set $f_{\text{off}} = 0.023$ so that about 75% of Chinese firms in the tradable sector sells offline to another province *when the cost for e-Commerce in the economy is set to infinite*. We choose the 75% figure as the calibration target for the case *without* e-Commerce because the source of the data, the World Bank Investment Survey, is from 2004. We parameterize bilateral trade cost τ^{ij} with the log-linear functional form $\tau^{ij} = \bar{\tau} e^{\gamma_X X_{ij}}$, following a large literature in international trade.[27] We take the γ_X coefficients from Fan (2015) to pin down the relative effects of various bilateral characteristics on bilateral trade costs.[28] We then set the scale parameter $\bar{\tau}$ such that the intra-provincial trade accounts for 70% of total sales in the model.[29]

[25]We make the following adjustment in our calculations to the source data from the 2013 China Statistical Yearbook, which contains detailed components of service value added. To be consistent with the model, we reclassify shipping cost as belonging to the tradable sector, and then assign value-added in retail service proportionally to these reclassified tradable and non-tradable sectors. We then calculate the nontradable share based on the adjusted classification.

[26]In the discussion that follows, it is understood that the model moments are affected by the parameters jointly, but some moments are more closely associated with certain parameters.

[27]The vector of bilateral characteristics X_{ij} includes whether the two cities belong to the same province, whether the two cities belong to the same region, and bilateral distance and its interactions with the dummy variables above.

[28]Fan (2015) estimates a model with inter-city trade. There is no data on inter-city trade flows in China, so the author aggregates city-level trade flows to the provincial level and estimates the relevant parameters by using non-linear least square to the 2002 inter-provincial input-output table.

[29]There is no readily available data source for intra-provincial trade share in recent years. In the public available 2002 *inter-provincial* input-output table, for a typical province, the intra-provincial trade accounts for about 90% of total trade. In the 2007 and the most recent *inter-regional* input-output table (the data is only down to the regional level, which there are eight in

To determine the values of the e-Commerce cost δ^{ji}, we match the aggregate patterns of online sales. We assume that δ^{ji} follows the function form $\delta^{ji} = \tau^{ji}\bar{e}L_i^{\zeta}$, where $\zeta < 0$. Therefore, the cost wedge of e-Commerce e^{ji} is given by $e^{ji} = \bar{e}L_i^{\zeta}$, meaning that the cost wedge of e-Commerce relative to offline retail is lower for destinations with larger population. This assumption is motivated by the increasing return to scale in transportation and distribution. For example, with more packages going to cities with larger population, delivery trucks serving these routes are more likely to make multiple full-load trips a day, ensuring fast delivery. Given that big cities are usually also the air hubs and rail rubs in China, the difference in time and cost between shipping to big and small cities can be even larger.[30]

We set $\bar{e}$ and ζ to match the average online expenditure share of 7.9% and a population elasticity of online expenditure share of -0.116 among the 337 cities. The -0.116 figure is in the lower range of our estimates in the empirical section. We choose this as calibration target to be conservative, and the use of a larger elasticity will produce larger impacts of e-Commerce on spatial inequality. The calibrated e-Commerce cost e^{ij} ranges from 1.15 to 2.97, with an average of 1.69. The population elasticity of e-Commerce cost ζ is set at -0.159. We call the resulting economy "the benchmark economy with e-Commerce." Table 6 summarizes our calibration strategy.

6.2 Spatial Consumption Inequality before e-Commerce

In this subsection, we analyze the spatial inequality through the lens of the model. We focus on the tradable sector in our discussion of the online expenditure share, number of entrants, offline varieties and expenditure share on goods from other provinces, but we take into account the non-tradable sector in our calculation of real income. We examine these variable by quintile of city population.

The first panel of Table 7 summarizes the key variables before the emergence of e-Commerce. The number of entrants is closely linked to city population, with an elasticity of 1.040 before e-Commerce. The average number of entrants is 9.7 times greater for the cities in 5th population quintile than for the cities in the 1st quintile. A larger city also attracts more firms from other cities to set up offline stores. However, due to trade costs, the population elasticity of offline varieties from other cities is smaller than the population elasticity of entrants. As shown in Table 7, the total number of offline varieties, which include local varieties and offline varieties imported from other cities, differs substantially across cities. The population elasticity of total offline varieties is 0.770.

Table 7 also summarizes the expenditure share spent on varieties imported, online or offline, from other provinces. This expenditure share is strongly and negatively correlated with population, with the average ranging from 28.5% for the 5th population quintile to 53.0% for the 1st population quintile. While smaller cities have lower local entry and lower number of offline stores from other provinces, proportionally speaking, the disadvantage in the number of varieties from other provinces is smaller.

total in China), intra-regional trade share decrease by about 7-12% compared to 2002. Applying this decrease uniformly to all provinces within a region, and assuming that the growth in intra-provincial trade during 2007-2012 is the same as that during 2007-2012, we obtain that for a typical province in 2012, the intra-provincial trade is about 70%. We also perform robustness by using 80% as the targets. The results are similar.

[30]If we restrict $\zeta = 0$, the model would predicts larger cities purchase too little from the online platform compared to the data.

As a result, workers in smaller cities spend a larger share of their expenditure on varieties from other provinces.

We can now examine the real income of a worker. The real income of a worker in city i, W_i, is defined as

$$W_i = \frac{w_i}{\prod_h P_{ih}^{\beta_h}}, \tag{15}$$

where the city-sector-specific price index P_{ih} is governed by not only the number of varieties, but also the marginal cost, including trade costs, of each variety. According to Table 7, the real income of an average city in the 5th quintile is about twice of that of an average city in the 1st quintile. The population elasticity of the (overall) ideal price index and real income is -0.209 and 0.287, respectively. In summary, our model generates large differences in access to varieties of goods and in real income, and these differences are closely related to the population size of a city.

These results are obtained under the benchmark calibration, which targets the overall level of e-Commerce and the elasticity of online expenditure share with respect to population. However, we leave out important information by not matching the online expenditure share of individual cities. One question is to what extent the population elasticities of price index and of real income are driven by the scale effect inherent in the model, rather than by the data moments regarding online expenditure shares.[31]

To address this question, we pursue an alternative calibration strategy. Under this calibration strategy, conditional on access to the online seller (δ^{ij} as well as the distribution of online sellers across cities), we attribute the observed difference in online expenditure shares to a city-specific *offline* import cost, and thereby match the online expenditures shares for all cities. The broad interpretation of this city-specific offline import cost is any determinants of the availability and attractiveness of local offline goods, such as local distribution infrastructure or business atmosphere. In addition, we match the nominal income of each city in the model with the data by adding city-specific labor productivity to the model. An advantage of this alternative approach is that, since we match the online expenditure shares city by city, our measured inequality does not stem mechanically from the scale effect in our model which features endogenous entry. Instead, the approach aims to extract information contained in the online purchase behavior. The details of this alternative calibration strategy are presented in Appendix II. We find that the population elasticity of price index is -0.109 under the alternative calibration. This is smaller in absolute value than the elasticity of -0.209 from the benchmark case. We also find that the real income inequality across cities with different sizes, as measured by the population elasticity of real income, are almost twice of that for nominal income. Therefore, focusing only on the difference in nominal income across cities will likely understate the significance of inequality in China.[32]

In the following exercises, we choose to have the more parsimonious calibration, which does not match the online expenditure shares by city, as our benchmark calibration for the transparency of the

[31]Scale effect refers to that, in new trade/economic geography models with endogenous entry, such as the one presented here, larger cities will have more varieties and hence higher real income. The baseline calibration strategy relies on using the model to pin down the offline varieties, and therefore have a built-in force that generates spatial inequality.

[32]The population elasticity is 0.228 for real income, and 0.119 for nominal income. In Appendix II, we further compare these numbers to relevant estimates in the literature.

results. However, the main conclusion of this paper that e-Commerce reduces real income inequality across cities, is robust to this alternative calibration strategy.

6.3 The Welfare Gains from e-Commerce

The second panel of Table 7 summarizes the economy with e-Commerce while the third panel reports the changes in the various outcome variables. The online expenditure share ranges from an average of 6.70% for the 5th population quintile to an average of 9.11% for the first quintile, with an overall average of 7.85%.[33] The log of online expenditure share of each city in the model is also plotted against the log of population in Figure 2.

The emergence of e-Commerce, as induced by the reduction of e-Commerce cost δ^{ij} from infinity to its current level, increases firm entry in all cities by an average of 5.98%, with larger percentage increase for the smaller cities. On the other hand, as shown in Figure 3a, the emergence of a new retail channel crowds out offline stores. While about 73.5% of firms sell offline to another province in the economy before e-Commerce, 32.5% of the firms do so in an economy with e-Commerce. As shown in Table 7 and in Figure 3b, the larger cities experience the largest percentage drop in the number of offline varieties. These changes affect the composition of consumption for workers. The expenditure share on varieties from other provinces increases for most cities due to e-Commerce, with an average increase of 1.47 percentage points.

We are now ready to examine the welfare effects of e-Commerce in this section. The percentage gains from e-Commerce for workers in city i is given by

$$\text{Gains from e-Commerce} = \frac{W_i^{\text{E}}}{W_i^{\text{pre-E}}} - 1 \tag{16}$$

where W_i^{E} and $W_i^{\text{pre-E}}$ are real income and the superscripts "E" and "pre-E" denote the model economies with and without E-Commerce, respectively.

Figure 4 plots the welfare gains of each city against the log of population. The gains from e-Commerce range from an average of 1.21% for the 5th population quintile to an average of 2.15% for the first quintile, with an overall average of 1.62%. To put these welfare number into perspective, consider the gains from international trade for the United States. Arkolakis et al. (2012) find that the gains from international trade for the US on the order of 1% using their sufficient-statistics formula. Therefore, the gains from e-Commerce for Chinese cities are substantial. More importantly, Figure 4 shows that the gains from e-Commerce decrease with city population. The gains from e-Commerce decrease by -0.522 percentage points for every log point increase in population. In this sense, e-Commerce can reduce inequality in welfare across Chinese cities.

This reduction, however, is small, compared to the spatial inequality that are present in China—with the emergence of e-Commerce, the population elasticity of the ideal price index increases modestly from

[33] Recall that the average of online expenditure share and the population elasticity of online expenditure share are targeted moments in the calibration exercise.

-0.209 to -0.205, a drop of 2%. This is largely due to the high inequality in the original equilibrium without e-Commerce. In Section 6.4, we show that further development in e-Commerce in China can continue to reduce consumption inequality across cities.

6.3.1 A Comparison with the ACR Formula

It is instructive to compare the predictions of our models to the ACR formula in Arkolakis et al. (2012). With the offline channel in the model, we should expect our results to differ from those predicted by the ACR formula. When e-Commerce emerges, on the one hand, firms may substitute to online sales by decreasing offline sales, offsetting the positive effects of e-Commerce for consumers; on the other hand, the ex-ante expected profit of entry will be larger since firms can now sell to other cities without incurring any fixed cost, and a greater number of entrants would enter the market. This would increase the offline sales, further increasing consumers' welfare. In sum, these two channels imply that the gains from e-Commerce might be larger or smaller than the prediction of the ACR formula.

To make the comparison, the ACR formula of a multi-sector trade model is given by

$$\widehat{W}_j = \prod_h (\widehat{\lambda}_h^{jj})^{\frac{\beta_h}{\varepsilon_h}} \tag{17}$$

where $\widehat{\lambda}^{jj}$ is the share of expenditure on domestic goods and ε_h is the trade elasticity.[34] We use the e-Commerce analogue of Equation 17 to obtain the "ACR Welfare Gains" for the cities in our model. More specifically, we replace the domestic expenditure share in Equation 17 with the expenditure share through the offline channel. In a conventional Melitz (2003) model with Pareto distribution, the trade elasticity is given by the Pareto shape of the productivity distribution. Therefore, we use the Pareto shape parameter of the productivity k_h as our best guess of the relevant e-Commerce elasticity in our model. The resulting "ACR Welfare Gains" are also plotted in Figure 4.

The direct application of the ACR formula gives an average of 1.25% for the welfare gains from e-Commerce. Therefore, the ACR welfare formula provides a rather different approximation of the true welfare gains for our model. Figure 4 shows that the ACR formula under-predicts the gains noticeably for smaller cities (with high online expenditure share). A potential explanation is that the market size effect dominates for the smaller cities—with e-Commerce, the greater potential to sell to other markets increases the number of entrants and total offline sales. This channel effectively makes the online and offline channel complements to each other for consumer welfare. Since the channels are not accounted for by the ACR formula, the welfare gains in our model are systematically different from those from the ACR formula.

6.4 Effects of Future Development in e-Commerce

We have quantified the realized gains from e-Commerce. However, the rapid development of e-Commerce in China is likely to continue at a fast pace in the near future. In this section, we consider two experi-

[34]This formula assumes that share of labor in each sector does not change. See Costinot and Rodríguez-clare (2013)

ments in which we lower the e-Commerce cost wedge e^{ij} so as to double and triple the online expenditure share, respectively.[35]

We decrease the e-Commerce cost wedges according to $e^{ij\prime} - 1 = \phi(e^{ij} - 1)$, where ϕ is a common factor.[36] We pick ϕ so that the average online expenditure share for the 337 cities is 15.5% and 23.55% respectively. We call these two experiments "Experiment 1" and "Experiment 2" respectively. We compare the outcomes under these two experiments to "the benchmark economy with e-Commerce."

Table 8 summarizes the results of these two experiments. Under both experiments, online expenditure share continues to be negatively correlated with city population. The slope between log online expenditure share and log population becomes steeper with lower e-Commerce cost from -0.115 in the present to -0.187 under Experiment 1 and -0.224 under Experiment 2. As before, the reduction in e-Commerce increases firm entry in all cities. Furthermore, the share of firms which sell offline to other provinces and the total number of offline varieties continue to decrease. Again, these changes have implications for the consumption composition of workers. The share of total expenditure spent on varieties from other cities (online and offline included) increases by an average of 1.91 percentage points and 4.00 percentage points for Experiment 1 and Experiment 2, compared to that in the benchmark economy with e-Commerce, respectively.

Figure 5 shows that the percentage increase in real income is negatively correlated with city population for both experiments. The increase in real income for workers in Experiment 1 ranges from an average of 0.94% for the 5th population quintile to an average of 2.05% for the 1st quintile, with an overall average of 1.36%, while the increase in Experiment 2 ranges from an average of 1.25% for the 5th quintile to an average of 4.40% for the 1st quintile, with an overall average of 2.76%.

We can combine the welfare gains from e-Commerce in Section 6 with the numbers in Table 8 to arrive at the overall welfare gains from e-Commerce under Experiment 2. By the time the average online expenditure share in China becomes three times of its 2013 level, the cumulative welfare gains from e-Commerce would be about 6.5% for cities in the smallest quintile of size distribution and 3.0% for cities in the largest quintile. Therefore, further development of e-Commerce in China, in the fashion of Experiment 1 and Experiment 2, would be able to increase real income of workers in all cities, and to substantially reduce the inequality in real income between them.

7 Concluding Remarks

This paper studies the welfare effects of e-Commerce on consumers in different cities. Using data from a major e-Commerce platform in China, we first document a robust negative correlation between market size and share of expenditure spent online. We then use a general equilibrium model to quantify the welfare effects of e-Commerce. We find that, the welfare gains from e-Commerce are 0.94 percentage

[35] The total revenue of e-Commerce in China is projected to increase by as much as 225% from 2012 to 2020. Assuming that total consumption is increasing at annual rate of 7%, the share of total expenditure spent online would increase by 88.9% over the 8 years (Dobbs et al., 2013). Therefore, the experiments considered in this section are well within reach in the near future.

[36] This way of lowering e-Commerce cost ensures that the e-Commerce cost continues to satisfy the restriction that $e^{ij} > 1$.

point higher for cities in the 1st population quintile than those in the 5th quintile. Therefore, e-Commerce reduces the spatial inequality in consumption. With its rapid growth, e-Commerce promises to further reduce this inequality in the coming years.

In this paper, we have focused on the distribution of welfare gains across cities. In reality, online consumption intensity varies greatly over the socio-economic spectrum, and the welfare gains from e-Commerce might have an important within-city dimension. The rise of e-Commerce may also have substantial impacts on other aspects of the local economy, such as the traditional retailing sector, and transportation and logistics sectors. All these are exciting areas for future research.

References

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

_ , _ , Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," *National Bureau of Economic Research Working Paper*, 2015.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," *Unpublished Manuscript*, 2014.

_ , Benjamin Faber, and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," *National Bureau of Economic Research Working Paper*, 2015.

Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang, "Roads, Railroads and Decentralization of Chinese Cities," *National Bureau of Economic Research Working Paper*, 2015.

Brandt, Loren and Carsten A Holz, "Spatial Price Differences in China: Estimates and Implications," *Economic Development and Cultural Change*, 2006, *55* (1), 43–86.

Broda, Christian and David E Weinstein, "Globalization and the Gains From Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Brown, Jeffrey R and Austan Goolsbee, "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 2002, *110* (3), 481–507.

Brynjolfsson, Erik and Michael D Smith, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 2000, *46* (4), 563–585.

_ , Yu Hu, and Michael D Smith, "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 2003, *49* (11), 1580–1596.

Clay, Karen, Ramayya Krishnan, Eric Wolff, and Danny Fernandes, "Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry," *Journal of Industrial Economics*, 2002, *50* (3), 351–367.

CNNIC, *2014 nian zhongguo wangluo gouwu shichang yanjiu baogao* [The 2014 Report on the Market for Online Shopping in China], China Internet Network Information Center, 2015.

Costinot, Arnaud and Andrés Rodríguez-clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in "Handbook of International Economics" 2013.

Couture, Victor, "Valuing the Consumption Benefits of Urban Density," *Unpublished Manuscript*, 2015.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2015.

Dobbs, Richard, Yougang Chen, Gordon Orr, James Manyika, Michael Chui, and Elsie Chang, *China's E-tail Revolution: Online Shopping as a Catalyst for Growth*, The McKinsey Global Institute, 2013.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence from French firms," *Econometrica*, 2011, *79* (5), 1453–1498.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

Faber, Benjamin, "Trade Liberalization, the Price of Quality, and Inequality: Evidence from Mexican Store Prices," *Unpublished Manuscript*, 2014.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *Unpublished Manuscript*, 2015.

Forman, Chris, Anindya Ghose, and Avi Goldfarb, "Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 2009, *55* (1), 47–57.

Glenn, Ellison and Sara Fisher Ellison, "Match Quality, Search, and the Internet Market for Used Books," *Unpublished Manuscript*, 2014.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Önsel Emre, "E-Commerce and the Market Structure of Retail Industries," *Economic Journal*, 2010, *120* (545), 651–682.

Goolsbee, Austan and Peter J Klenow, "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet," *American Economic Review*, 2006, *96* (2), 108–113.

Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living across US Cities," *Unpublished Manuscript*, 2013.

_ and David E Weinstein, "Goods Prices and Availability in Cities," *Review of Economic Studies*, 2014, p. rdu033.

Helpman, Elhanan, Marc Melitz, and Stephen Yeaple, "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 2004, *94.*

Hortaçsu, Ali, F Asís Martínez-Jerez, and Jason Douglas, "The Geography of Trade in Online Transactions: Evidence from eBay and Mercadolibre," *American Economic Journal: Microeconomics*, 2009, pp. 53–74.

Hottman, Colin, "Retail Markups, Misallocation, and Store Variety in the US," *Unpublished Manuscript*, 2014.

Jensen, J Bradford, Jonathan Eaton, Samuel Kortum, and Andrew B Bernard, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Lagakos, David, "Explaining Cross-Country Productivity Differences in Retail Trade," *Journal of Political*

Economy, forthcoming.

Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina, "There Goes Gravity: How eBay Reduces Trade Costs," *Economic Journal*, Forthcoming.

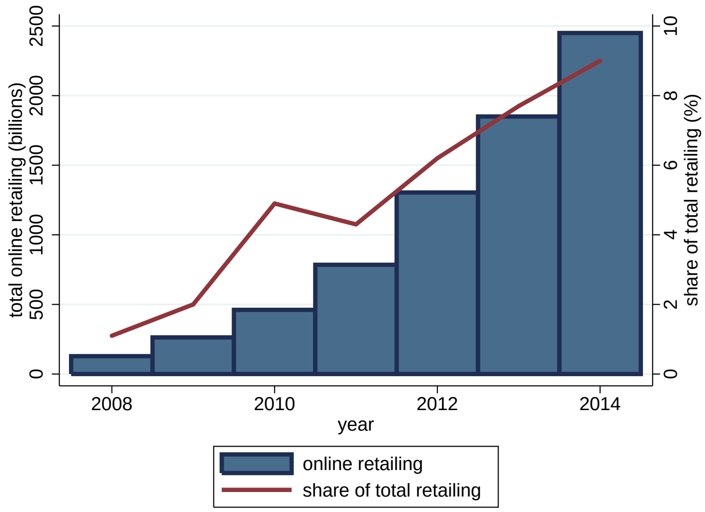
_ , Marcelo Olarrega, Simon Schropp, and Pierre-Louis Vézina, "eBay's Anatomy," *Economics Letters*, 2013, *121* (1), 115–120.

Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

_ and Stephen Redding, "New Trade Models, New Welfare Implications," *American Economic Review*, 2015, *105* (3), 1105–46.

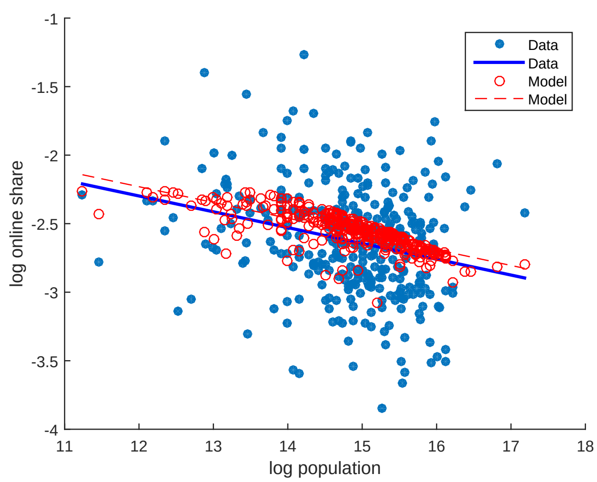
Tran, Allen, "The Aggregate Impact Of Online Retail," *Unpublished Manuscript*, 2014.

Figure 1: Growth of Online Retailing in China



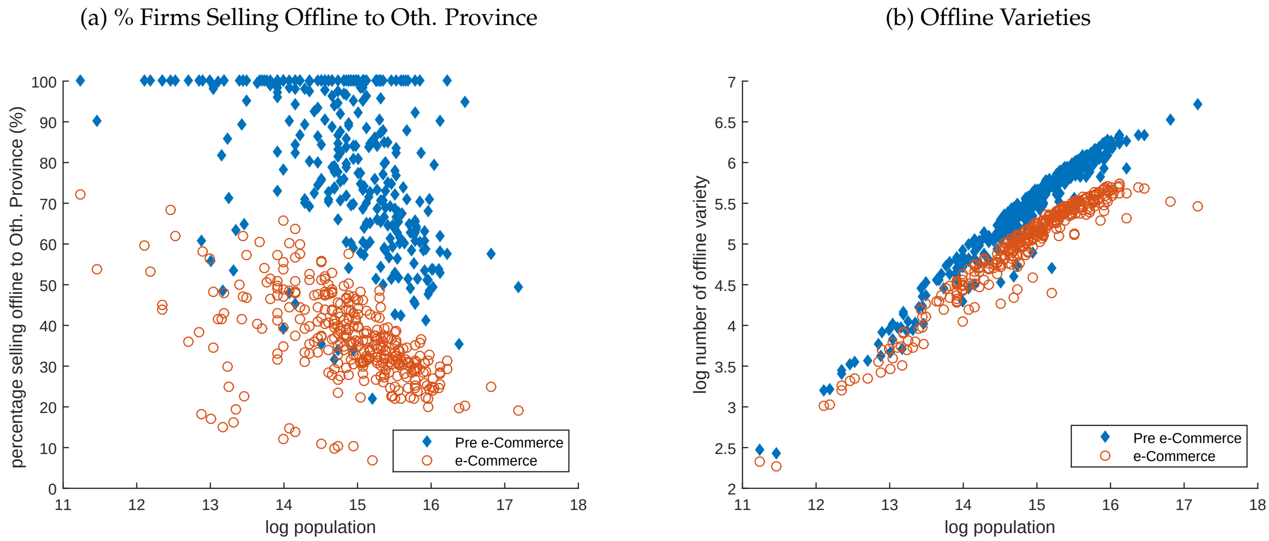
Notes: The bars (left axis) are the total online retail sales (in billion RMB yuan); the line (right-axis) is the share of online retail sales in total retail sales. Source: CNNIC (2015).

Figure 2: Log Online Expenditure Shares and Population: Model and Data



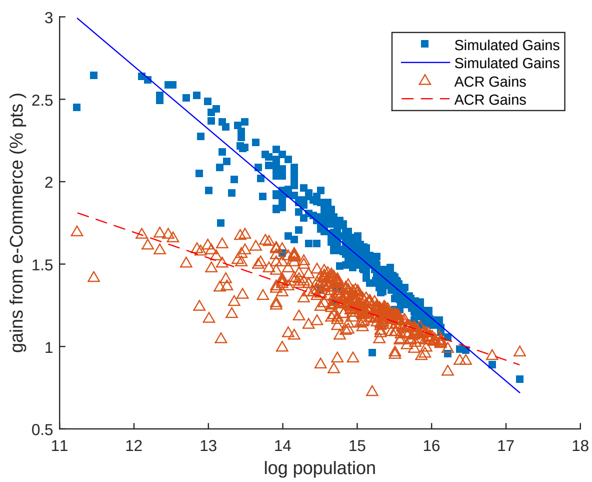
Notes: The solid dots denote data points from Alibaba Inc., while the hollow circles represent calculations from model simulations. The slopes of the fitted lines are -0.116 and -0.115 for the data and the model respectively.

Figure 3: The Effects of e-Commerce on Offline Stores



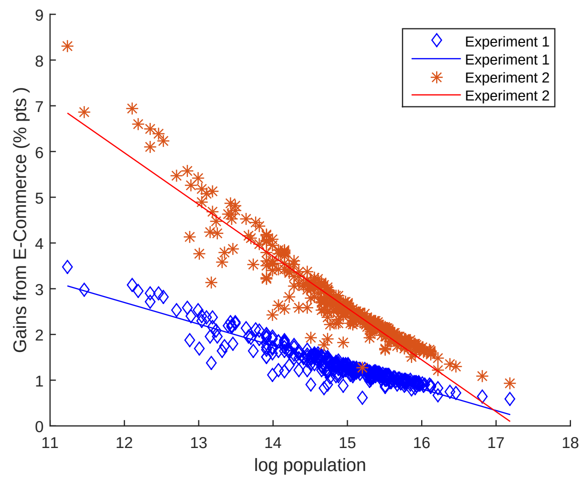
Notes: calculation based on model simulations by the authors.

Figure 4: Welfare Gains from e-Commerce and Population



Notes: calculation based on model simulations by the authors. The simulated gains are computed from the calibrated model while the "ACR gains" are calculated from the formula in Arkolakis et al. (2012). The slopes of the fitted lines are -0.382% and -0.155% for the simulated gains and the ACR gains respectively.

Figure 5: Welfare Gains from Further Development of e-Commerce



Notes: calculation based on model simulations by the authors. We double and triple the average online expenditure share for Experiment 1 and Experiment 2, respectively. All numbers are computed relative to the current level of e-Commerce cost. The slopes of the fitted lines are -0.472% and -1.13% for Experiment 1 and Experiment 2, respectively.

Table 1: Summary Statistics

	N	mean	sd
ln online sales over total retail sales	337	-2.625	.401
ln population	337	14.836	.88
ln average income	337	9.531	.376
ln average years of schooling	337	2.158	.143
ln non-agricultural employment share	337	3.758	.449
ln retail employment share	337	2.028	.451
ln average retailer size	337	2.905	.557
ln distance to highway	327	6.667	3.871
ln distance to railway	327	6.567	3.925
ln share of households with broadband	325	-.856	.739
ln housing price	288	8.54	.384
provincial capital	337	.086	.281
ln online sales over total consumption	327	-2.521	.431
ln online sales over non-hospitality retail	334	-2.588	.423

Notes: Each observation is a city.

Table 2: Market Size and Online Expenditure Share

dep var: log share of online shopping in total retail sales							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln popuplation	-0.116***	-0.135***	-0.133***	-0.135***	-0.102***	-0.120***	-0.119***
	(0.0261)	(0.0260)	(0.0271)	(0.0235)	(0.0267)	(0.0267)	(0.0255)
ln per capita income		0.158**			-0.330***	-0.00461	-0.00526
		(0.0652)			(0.115)	(0.132)	(0.127)
ln average years of schooling			0.307**		-0.307	-0.409	-0.411
			(0.156)		(0.211)	(0.314)	(0.296)
ln share of non-agricultural employment				0.237***	0.530***	0.333***	0.335***
				(0.0496)	(0.0888)	(0.0923)	(0.0902)
province FE						X	X
N	337	337	337	337	337	337	337
R^2	0.0645	0.0847	0.0750	0.133	0.165	0.572	
First stage F-stat							861.8

Notes: Each observation is a city in China in 2013. The dependent variable is log share of online expenditure in total retail sales. The explanatory variable of interest is log population in 2010, which is used as a proxy for market size. Columns 1 to 6 are estimated using OLS. Column 7 is estimated using 2SLS. Log population in 2000 is used as an instrument for log population in 2010. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Additional Covariates

	dep var: log share of online shopping in total retail sales					
	(1)	(2)	(3)	(4)	(5)	(6)
ln population	-0.135***	-0.0940***	-0.200***	-0.141***	-0.196***	-0.254***
	(0.0298)	(0.0289)	(0.0400)	(0.0292)	(0.0415)	(0.0590)
ln distance to highway	-0.00333					0.00279
	(0.00511)					(0.00503)
ln distance to railway	0.0120**					0.00723
	(0.00563)					(0.00593)
ln ratio with broadband internet		0.142***				0.0682
		(0.0492)				(0.0550)
ln housing price (*yuan* per sqm)			0.306***			0.179**
			(0.0713)			(0.0789)
provincial capital				0.135*		-0.115
				(0.0746)		(0.106)
log distance-weighted access to online sellers					0.161***	0.191***
					(0.0460)	(0.0696)
baseline covariates	X	X	X	X	X	X
N	327	325	288	337	333	277
R^2	0.593	0.590	0.624	0.577	0.596	0.662

Notes: Each observation is a city in China in 2013. The dependent variable is log share of online expenditure in total retail sales. The explanatory variable of interest is log population in 2010, which is used as a proxy for market size. Baseline covariates include those in Column 6 of Table 2. All columns are estimated using OLS. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Alternative Measures of Market Size and Online Expenditure Share

	(1)	(2)	(3)	(4)	(5)
dep var	log online sale as share of				
	total retail sales			consumption	non-hospitality retail
ln population	-0.120***			-0.0719***	-0.129***
	(0.0267)			(0.0262)	(0.0284)
ln total income		-0.139***			
		(0.0298)			
ln total consumption			-0.102***		
			(0.0264)		
baseline covariates	X	X	X	X	X
N	337	337	327	327	334
R^2	0.572	0.578	0.566	0.671	0.571

Notes: Each observation is a city in China in 2013. The dependent variable is log share of online expenditure in total retail sales in Columns 1 to 3, log share of online expenditure in total consumption in Column 4, and log share of online expenditure in total retail sales excluding sales by restaurants and hotels. The main explanatory variable is some measure of market size, which is log population in 2010 in Columns 1, 4, and 5, log total income in Column 2, and log total consumption in Column 3. Baseline covariates are the same set of controls as in Column 6 of Table 2. All columns are estimated using OLS. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: UHS Category Regressions

	(1) food	(2) cloth	(3) residence	(4) hhd appliance	(5) health	(6) trans	(7) recreation
ln Population	0.00769	-0.0774***	0.0444	-0.0766***	-0.0218	-0.0536**	-0.00990
	(0.0311)	(0.0272)	(0.102)	(0.0281)	(0.0270)	(0.0245)	(0.0289)
baseline covariates	X	X	X	X	X	X	X
N	337	337	336	337	337	337	337
R^2	0.880	0.798	0.523	0.796	0.861	0.679	0.723
mean dep var	-5.162	-1.414	-10.94	-1.431	-3.162	-3.300	-3.374

Notes: Each observation is a city in China in 2013. Each column is estimated using OLS. The dependent variables are log shares of online expenditure in total expenditure by each category. The categories are defined by Urban Household Survey (UHS). See text for details on how Taobao categories are mapped into UHS categories and how city-category level expenditures are calculated. These categories are: food in Column 1, clothing in Column 2, residence-related goods and services in Column 3, household appliances and services in Column 4, health care and medical services in Column 5, transportation and communications in Column 6, and recreation, education and culture in Column 7. Mean dependent variables (log online shares) for each column are reported in the bottom row. Baseline covariates are the same as those in Column 6 of Table 2. The explanatory variable of interest is log population in 2013, which is used as a proxy for market size. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Parametrization of the Model

Parameters (Symbol)	Value	Source/Target
Non-tradable Share (β_{NT})	0.35	Expenditure Share
Elasticity of Substitution (σ)	4	Jensen et al. (2003) and Melitz and Redding (2015)
Pareto Shape of Productivity (k)	4.27	Melitz and Redding (2015)
Scale Parameter of Productivity ($\bar{a}$)	10	Normalization Constant
Fixed Cost of Entry (f_E)	9.3	Average employment size of firms
Fixed Cost of Offline Sales (f_{off})	0.023	Share of firms selling offline to another province
Offline variable trade costs (τ^{ij})	See text.	Fan (2015);
		Share of total trade flows accounted by intra-provincial trade
Online variable trade cost (δ^{ij})	See text.	Online expenditure share
		Population elasticity of online expenditure share

Table 7: The Effects of E-Commerce on Entry, Offline Varieties and Welfare

Model Economy without e-Commerce			Quintile of City Population		
Variable	1st	2nd	3rd	4th	5th
Online Expenditure Share	0.00%	0.00%	0.00%	0.00%	0.00%
No. of Entrants	0.93	2.44	3.67	5.49	9.98
% Sell Offline to Oth. Prov	91.28%	84.98%	85.75%	75.26%	65.73%
Other Prov. Offline Varieties	68.51	141.58	194.66	263.22	371.74
Total Offline Varieties	85.38	179.85	248.17	334.26	454.49
% Exp. on Oth. Prov Varieties	52.95%	41.26%	37.90%	34.52%	28.46%
Real Income	0.98	1.30	1.47	1.65	1.93
Benchmark Economy with e-Commerce			Quintile of City Population		
Variable	1st	2nd	3rd	4th	5th
Online Expenditure Share	9.11%	8.17%	7.82%	7.38%	6.70%
No. of Entrants	1.00	2.59	3.89	5.78	10.40
% Sell Offline to Oth. Prov	44.71%	39.90%	37.86%	33.38%	28.30%
Other Prov. Offline Varieties	51.67	98.49	128.05	161.11	193.64
Total Offline Varieties	64.81	126.87	166.60	211.93	259.42
% Exp. on Oth. Prov Varieties	54.29%	42.85%	39.47%	35.96%	29.89%
Real Income	1.00	1.33	1.49	1.68	1.95
% Change			Quintile of City Population		
Variable	1st	2nd	3rd	4th	5th
Online Expenditure Share (% pt)	9.11%	8.17%	7.82%	7.38%	6.70%
No. of Entrants	7.87%	6.39%	5.89%	5.38%	4.40%
% Sell Offline to Oth. Prov (% pt)	-46.57%	-45.08%	-47.90%	-41.88%	-37.43%
Other Prov. Offline Varieties	-23.10%	-30.20%	-34.13%	-38.63%	-46.80%
Total Offline Varieties	-22.64%	-29.10%	-32.71%	-36.36%	-42.10%
% Exp. on Oth. Prov Varieties (% pt)	1.33%	1.59%	1.57%	1.43%	1.43%
Real Income	2.15%	1.75%	1.59%	1.41%	1.21%

Notes: calculations based on model simulation. We classify the 337 cities in our sample according to population quintiles (in increasing order). The average number of entrants and the average real income for cities in the 1st population quintile for the "Benchmark Economy with e-Commerce" are both normalized to one. The changes in online expenditure share, share of firms selling offline to other provinces, and share of expenditure spent on varieties from other province are measured in differences in percentage points while the changes in other variables are measured in percentage differences.

Table 8: The Effects of Further e-Commerce Development

Experiment 1: doubling the city average of online expenditure share					
% Changes in Experiment 1	Quintile of City Population				
Variable	1st	2nd	3rd	4th	5th
Online Share (% pt)	10.77%	8.26%	7.47%	6.72%	5.62%
No. of Entrants	3.73%	7.67%	11.55%	17.56%	27.67%
% Sell Offline to Oth. Prov (% pt)	-17.51%	-16.13%	-15.67%	-13.77%	-11.82%
Other Prov. Offline Varieties	-28.46%	-31.75%	-33.66%	-35.71%	-38.89%
Total Offline Varieties	-27.99%	-30.92%	-32.73%	-34.48%	-37.04%
% Exp. on Oth. Prov Varieties (% pt)	2.29%	2.06%	1.90%	1.72%	1.58%
Real Income	2.05%	1.43%	1.27%	1.12%	0.94%
Experiment 2: tripling the city average of online expenditure share					
% Changes in Experiment 1	Quintile of City Population				
Variable	1st	2nd	3rd	4th	5th
Online Expenditure Share (% pt)	22.24%	16.68%	14.99%	13.40%	11.08%
No. of Entrants	7.24%	15.07%	21.43%	30.17%	43.01%
% Sell Offline to Oth. Prov (% pt)	-28.47%	-25.85%	-24.81%	-21.90%	-18.71%
Other Prov. Offline Varieties	-53.89%	-57.03%	-58.82%	-60.70%	-63.51%
Total Offline Varieties	-52.90%	-55.58%	-57.23%	-58.75%	-60.68%
% Exp. on Oth. Prov Varieties (% pt)	5.11%	4.39%	3.97%	3.51%	3.06%
Real Income	4.40%	2.93%	2.55%	2.20%	1.77%

Notes: calculations based on model simulation. We classify the 337 cities in our sample according to population quintiles (in increasing order). The numbers are calculated relative to the benchmark economy with e-Commerce, which we have calibrated to match the 2013 city average of online expenditure share. The changes in online expenditure share, share of firms selling offline to other provinces, and share of expenditure spent on other province varieties are measured in in differences in percentage points while the changes in other variables are measured in percentage differences.

Appendix

Appendix I Regressions by Product Category

Table A: Coefficient by Category

product category	service	coef.	s.e	t-stat
basic building materials		-0.267	0.0427	-6.249
motorcycle, parts & accessories		-0.256	0.0400	-6.412
desktop computers & servers		-0.215	0.0391	-5.493
sportswears, sports bags & accessories		-0.212	0.0287	-7.405
home improvement supplies		-0.210	0.0341	-6.173
beddings, linens & curtains		-0.207	0.0286	-7.219
outdoor,hiking & camping gears		-0.203	0.0317	-6.400
office furnitures		-0.178	0.0336	-5.313
men's shoes		-0.172	0.0271	-6.361
car accessories & parts		-0.169	0.0312	-5.421
men's clothing		-0.168	0.0275	-6.119
children's clothing		-0.162	0.0284	-5.709
hardware & tools		-0.161	0.0315	-5.106
women's shoes		-0.158	0.0288	-5.487
festive & party supplies, gifts & crafts		-0.158	0.0343	-4.592
Electrical & Electronic Products		-0.156	0.0350	-4.450
home decoration		-0.155	0.0338	-4.588
women's clothing		-0.154	0.0279	-5.510
books, magazines & newspapers		-0.151	0.0273	-5.525
home appliances		-0.149	0.0305	-4.869
home furnitures		-0.148	0.0352	-4.198
cleaning tools, organization & storage		-0.144	0.0321	-4.487
musical instruments & parts		-0.143	0.0329	-4.353
underwears & pajamas		-0.141	0.0283	-4.993
sports, yoga & bodybuilding supplies		-0.141	0.0318	-4.441
tablet computer		-0.140	0.0320	-4.371
kitchen appliances		-0.139	0.0297	-4.675
sports shoes		-0.138	0.0275	-5.012
office appliances & supplies		-0.136	0.0334	-4.065
kitchen ware		-0.131	0.0336	-3.905
consumer electronics		-0.131	0.0316	-4.135
cleaning supplies, paper & deodorizer		-0.129	0.0322	-4.011
personal hygience & care		-0.128	0.0282	-4.555
staple food, dry food, oils & seasoning		-0.126	0.0368	-3.422
floral & gardening		-0.124	0.0358	-3.452
cellphone		-0.122	0.0289	-4.224
MP3s, ipods and voice recorders		-0.120	0.0304	-3.945
flash drive, digital storage & harddrives		-0.116	0.0321	-3.603
lighters,swiss knives & glasses		-0.115	0.0341	-3.386
cameras		-0.115	0.0377	-3.051

continued.				
product category	service	coef.	s.e.	t-stat
health & medical equipment		-0.111	0.0373	-2.968
tea, alcohol & beverages		-0.110	0.0343	-3.194
watches		-0.106	0.0334	-3.185
beauty products & cosmetics		-0.104	0.0312	-3.327
e-dictionary, e-reader & office stationary		-0.102	0.0330	-3.099
video & audio equipment		-0.102	0.0324	-3.154
snacks, nuts & speciality food		-0.102	0.0406	-2.508
laptop computer		-0.0985	0.0337	-2.923
fashion accessories		-0.0906	0.0328	-2.761
computer hardware, accessories & monitors		-0.0854	0.0356	-2.397
bags & luggages		-0.0845	0.0319	-2.649
consumer electronics parts		-0.0831	0.0304	-2.736
maternity care		-0.0797	0.0281	-2.837
nutrition supplement		-0.0708	0.0346	-2.045
food & supplies for pets		-0.0533	0.0433	-1.230
music, movies & tv shows		-0.0521	0.0387	-1.344
toys, models & comic books		-0.0498	0.0331	-1.504
web equipment		-0.0495	0.0379	-1.304
baby care & gears		-0.0448	0.0343	-1.304
antique, stamps, coins, paintings & collectibles		-0.0391	0.0404	-0.966
internet games	X	-0.0386	0.0442	-0.872
renovation contractor service	X	-0.0322	0.0855	-0.377
jewelry		-0.0218	0.0448	-0.486
education & training	X	-0.0203	0.0476	-0.426
customization, personal design & DIY	X	-0.00719	0.0393	-0.183
photography service	X	-0.00684	0.0625	-0.109
perishable food & prepared food	X	-0.00530	0.0519	-0.102
web services	X	0.0317	0.0413	0.767
prepaid cards/cellphone service	X	0.0560	0.0467	1.201
formula/baby food		0.0569	0.0451	1.264
credits for computer games	X	0.0698	0.0881	0.792
visitor pass, trips & travel services	X	0.0732	0.0638	1.147
credits & gift cards for websites	X	0.111	0.0870	1.270
dining, catering & takeouts	X	0.249	0.0922	2.697
transportation tickets	X	0.280	0.155	1.803
leisure & entertainment	X	0.296	0.0764	3.878
financial services	X	0.313	0.131	2.393
hotels & lodgings	X	0.367	0.158	2.319
gift cards & shopping cards	X	0.405	0.0951	4.253
real estate services	X	0.413	0.167	2.471
new & used cars		0.801	0.261	3.063

Notes: Each observation is a city. Each row contains an OLS estimation using the same specification as in Column 6 of Table 2, except that the dependent variable is the log of online expenditure in each category as a share of the total retail sales. The estimated coefficients associated with log population, robust standard errors and the t-statistics are reported.

Appendix II Measuring Real Income through the Lens of the Model

The goal of Section 6 is to analyze the effects of e-Commerce with a parsimonious parametrization that matches the most salient features of the data. Specifically, we make use of the city-average of the online expenditure share as well as the population elasticity of online expenditure share to calibrate the e-Commerce cost δ^{ij}. However, as shown in Figure 2, we do not match the online expenditure share of individual cities and therefore leave out important information. As discussed in the literature review, detailed data on the varieties and prices of goods across locations are not readily available, especially for a developing country such as China. In this section, we make use of our model and data on the online expenditure share of *individual* cities to infer the variety-adjusted price index of each city.

Assuming that the households face the same trade-cost-adjusted number of varieties online, different expenditure shares on this same "online bundle" reflects the differences in access to consumption goods in the offline markets and consequently differences in real income. In this section, we first match both the online expenditure share and the nominal income of *individual* cities. We then interpret the extent of differences in price level and real income across Chinese cities through the lens of the model. An important advantage of this approach is that, since we match the online expenditure shares city by city, our measured inequality does not stem mechanically from the scale effect in our model which features endogenous entry. Instead, the approach aims to reveal information contained in the online purchase behavior.

We make one modification to the model described in Section 5. We assume that there are technology differences across cities such that the distribution of effective labor endowment is different from the distribution of city size. Specifically, we assume $L_i = \phi_i \tilde{L}_i$ where $\tilde{L}_i$ corresponds to the data, ϕ_i is location-specific labor productivity, and L_i is now interpreted as "effective labor endowment."[37] We pick the new parameter ϕ_i to match the per-capita nominal income of all 337 cities.

We choose the non-tradable expenditure share parameter β_{NT}, the elasticity of substitutions σ_h, the productivity distribution parameters k_h and $\bar{a}$, bilateral trade cost τ^{ij}, fixed cost of entry f_E and the fixed cost of offline stores f_{off} the same way as in Section 6.

We specify e-Commerce cost as $\delta_{ij} = \tau_{ij}\bar{e}$ and choose $\bar{e}$ to match the average online expenditure share across cities. At this point, the online expenditure share of individual city deviates significantly from the data. We interpret the deviation between model and data for online expenditure share for each city as reflecting idiosyncratic *offline* variable trade cost of selling the differentiated goods *to* that city. To match online expenditure share of individual cities with the data, we keep the level of e-Commerce cost, δ^{ij}, fixed, and fine-tune the bilateral trade costs τ^{ij}. Specifically, we replace τ^{ij} with $k_j\tau^{ij}$ and change k_j in order to match the online expenditure share for city j.

The exact interpretation of k_j is not important for our purpose. The nature of this exercise is that, conditional on the access to online market, determined by online shipping costs and number of varieties

[37]The modification does not affect the equilibrium conditions in Section 5, except that the relevant variables is now measured in "per effective worker" terms rather than in "per capita" terms. To convert the "per effective worker" quantities in the model into the " per capita" quantities in the data, we multiply the relevant model quantities by $\frac{L_i}{\tilde{L}_i} = \phi_i$.

sold online, any deviation in the online expenditure shares is driven by the variation in the attractiveness of local offline varieties. We label this as city-specific import cost, but k_j can capture any destination-city specific factors that affect availability or quality of offline goods sold in city j in excess to those factors already considered in the specification for bilateral trade costs, for example, local distribution infrastructure and business culture.

Real Income Inequality across Cities

Figure Aa plots the log of the ideal price index against the log of population. As shown, there are large differences in the price index, with the most expensive city about 70 log points higher than the least expensive city. Crucially, the price index is strongly and negatively correlated with population, with an elasticity of -0.109.

Figure Ab and Figure Ac plot both the log of nominal income and the log of real income, against the log of population in the model, respectively. In contrast to the results in Table 7, Figure Ac incorporates the information from the online expenditure share of individual cities. Viewed through the lens of the model, dispersion in online expenditure shares translates into differential access to consumption goods in offline markets and hence the measured real income. The population elasticity of real income is 0.228, almost twice the population elasticity of 0.119 for nominal income. Therefore, focusing only on the difference in nominal income across cities will likely understate the significance of inequality in China.[38]

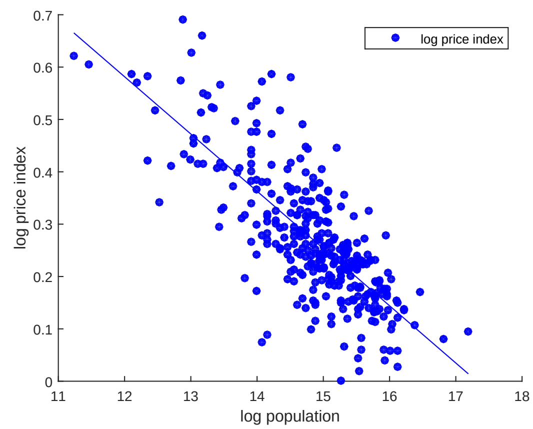
The inequality found in our analysis is higher than the estimates in the literature based on the US. For example, Handbury and Weinstein (2014) finds that the population elasticity of the ideal price index is about -0.01, much smaller than the number we have (-0.109). There are at least two factors that contribute to the difference. First, Handbury and Weinstein (2014) focuses on a sample of the largest MSAs (Metropolitan Statistical Areas) in the US, whereas our calibration exercise includes almost all cities in China, including the smallest ones. Second, the spatial inequality in development is far more pervasive in China than it is in the US.

In summary, our alternative calibration strategy, which incorporates information on online expenditure shares and nominal income of individual cities, finds that differential access to varieties of goods is an important source of real income inequality across cities in China.

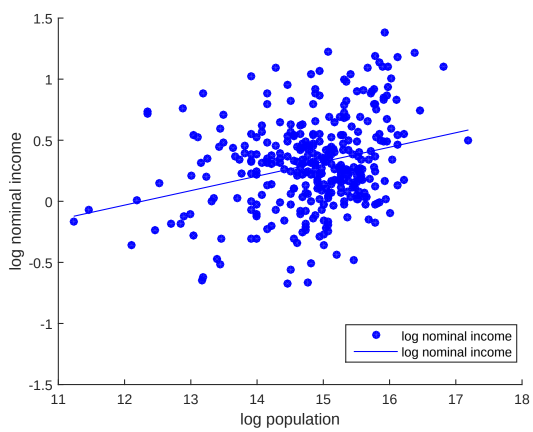
[38]Brandt and Holz (2006) construct a set of provincial-level spatial price deflators in China for the years 1984-2002, and find that provinces with higher income tend to have higher price level. However, they do not account for differences in varieties of goods, which is the focus of this paper and other recent work such as Handbury and Weinstein (2014).

Figure A: Real Income and Nominal Income under the Alternative Calibration

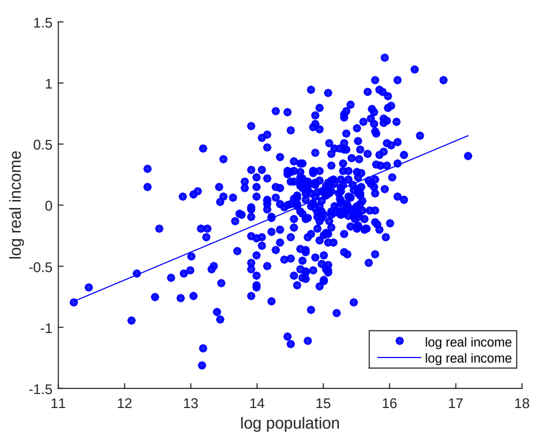
(a) Log Price Index



(b) Log Nominal Income



(c) Log Real Income



Notes: calculation based on model simulations by the authors using an alternative calibration strategy where we match both the online expenditure shares and nominal incomes of all 337 cities. We normalize the lowest price index among all cities to one. The slope of the fitted line is -0.109, 0.119 and 0.228 for price index, nominal income and real income respectively.