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

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

This Version: December 2016

Abstract

Domestic trade costs imply more restricted access to consumption varieties in smaller and less connected cities. By eliminating the fixed cost of firm entry and reducing the effect of distance on trade costs, e-Commerce might disproportionately improve these cities' access to varieties, and reduce the associated real income inequality across cities. The implication of this hypothesis is that residents in small and remote cities purchase more intensively online. Using unique data from China's leading e-Commerce platform, we show that online expenditure share is negatively correlated with population and market potential. We then build a multi-region general-equilibrium model to quantify the welfare gains from e-Commerce. We find the welfare gains from e-Commerce to be 1.6 percent. Furthermore, e-Commerce reduces the elasticity of real income with respect to population by 1.9% and the elasticity with respect to market potential by 4.1%.

Key Words: Intra-national trade cost, spatial consumption inequality, gains from e-Commerce
JEL Code: R13 F15 R12

*We thank Ali Research for providing access to the data for this research. For helpful comments we thank Roger Betancourt, Erik Brynjolfsson, Jonathan Dingel, Garrett Johnson, Nuno Limao, Asha Sundaram, and participants at the 2015 Midwest Trade Meeting, 2015 Urban Economics Association Meeting, 2016 NBER China Meeting, 2016 Econometric Society North America Summer Meeting, 2016 Econometric Society China Meeting, 2016 NBER Summer Institute, 2016 Econometric Society Asia Meeting, 2016 World Bank Annual Conference on Development Economics, 2016 NEUDC, the 9th Annual Conference on China's Economic Development and the U.S.-China Relationship, and seminars at Shanghai University of International Business and Economics, Shanghai Jiao Tong University and Federal Communications Commission. All errors are our own.

†Jingting Fan, University of Maryland; Lixin Tang, Shanghai University of Finance and Economics; Weiming Zhu, IESE Business School, University of Navarra; Ben Zou, Michigan State University. Contacts: Fan, fan@econ.umd.edu; Tang, tang.lixin@mail.shufe.edu.cn; Zhu, wzhu@iese.edu; Zou, benzou@msu.edu.

1 Introduction

Firms face significant trade costs in reaching consumers in other geographic locations: setting up storefronts in destination cities requires upfront investment; shipping goods from production location to destination cities is also costly. By impeding the movement of goods and services between locations, these trade frictions play an important role in shaping the spatial distribution of income and welfare within a country.[1]

While domestic trade frictions reduce welfare everywhere, their impacts on small and remote cities might be larger due to two reasons. First, because of the fixed costs associated with setting up a physical store and entering a market, smaller cities are less attractive to firms from other cities. This reduces the varieties of goods in small cities. Second, the large cost of shipping goods to remote destinations increases the price of outside goods in those cities and reduces firms' profit from entry, putting the remote cities at a disadvantage in terms of accessing consumption goods. Consistent with this view, recent studies (see, for example, Handbury and Weinstein, 2014; Feenstra et al., 2016) document that consumers in small cities have access to smaller varieties of goods, at higher prices.[2]

This paper studies the extent to which the rise of e-Commerce facilitates inter-city trade and reduces spatial inequality in consumption between cities. In recent years, e-Commerce has become an integral part of the modern economy and fundamentally changed the landscape of inter-city trade.[3] According to *emarketer.com*, a market research company for 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. E-Commerce differs from traditional trade conducted offline on two aspects. First, e-Commerce eliminates the need to set up brick-and-mortar stores and a distributional network, allowing firms to reach consumers in cities that otherwise would not be served. Second, e-Commerce can significantly reduce the penalty of distance on trade volumes, possibly by decreasing the cost of information.[4] With online shopping platforms, consumers can easily obtain information on numerous goods from a distant location. Consequently, trade volume on the online marketplace is much less affected by geographic distance than is trade through the traditional offline channel, as documented by Lendle et al. (forthcoming) and Hortaçsu et al. (2009). For the reasons discussed earlier, these two features of e-Commerce make it especially beneficial to residents in small and remote cities.

This paper focuses on China, which offers a suitable setting for studying the role of e-Commerce in reducing domestic frictions for inter-city trade and alleviating spatial inequality in consumption.[5] On the one hand, economic activity, including retailing, are very unevenly distributed in

[1]See, for example, Ramondo et al. (2016) for a quantification of how domestic trade frictions affect the aggregate welfare of a country.

[2]In many cases, small cities are also remote.

[3]For the purpose of this paper, we use the term "e-Commerce" to refer to commercial transactions conducted electronically between firms and consumers. A broader definition of "e-Commerce" would include economic transactions among firms, or any transaction of goods and services via computer networks, which is not the focus of this paper.

[4]Anderson and van Wincoop (2004) note that information costs are an important component of trade costs.

[5]In this paper, a "city" refers to a prefecture in China. China is divided into 345 prefectures. A prefecture typically

China. Large retailers are disproportionately concentrated in big cities, leaving many small cities and the vast rural area under-served. On the other hand, the growth of e-Commerce in the past few years has been remarkable. As Figure 1 shows, the total sales in online markets increased by more than 25 times between 2008 and 2014. By 2014, online sales accounted for about 10% of total retail sales. As numerous news articles have noted, online shopping has transformed both the consumption experience and the economic landscape in China.[6] Yet despite the sheer size and immense growth potential of e-Commerce, its impacts on inter-city trade and welfare of residents from different cities have not been systematically examined in the literature. This paper aims to fill this gap through the lens of domestic trade frictions.

We study the spatial distribution of online sales using unique data from Taobao, a subsidiary of Alibaba Inc. and the dominant online platform that accounts for more than 80 percent of online retail sales in China. Since consumers in smaller and more remote markets have more limited access to varieties of goods via brick-and-mortar stores, but their access to online varieties is more comparable to consumers in bigger and better-connected markets, it follows that they spend a larger share of their expenditure on online goods. We find empirical evidence consistent with this hypothesis. Using population as the main measure for market size and market potential (Harris, 1954) as the main measure for remoteness, we document a strong negative relationship between the share of online sales in total retail sales (hereafter "online expenditure share") and both population and market potential. In our preferred specification, the elasticities of online expenditure share with respect to population and market potential are -0.12 and -0.91, respectively. Everything else equal, with an average online share of 7.9 percent, the online share for a city in the 1st quintile of population is about 2.1 percentage points higher than that for a city in the 5th quintile; the online share for a city in the 1st quintile of market potential is 7.8 percentage points higher than that for a city in the 5th quintile.

We rule out a few alternative explanations for these negative relationships. The results are unlikely to be driven by differences in access to e-Commerce: they are robust to controlling for a host of supply-side shifters including access to transportation and broadband Internet. The results are also unlikely to be driven by heterogeneity in demand: the relationships remain after we control for a rich set of city-level socio-demographic characteristics, such as average income and education, as well as age and gender composition. Furthermore, we show some evidence that the negative correlations hold across the board for *each* consumption category, especially for categories in which Taobao specializes, further ruling out the possibility that our findings simply reflect the differences in consumption compositions between consumers in big and connected cities and those in small and remote ones.

We then develop a multi-region general equilibrium model to evaluate the effect of e-Commerce on spatial inequality in consumer welfare. In the model, a firm can use the online platform to sell

consists of several urban centers and the rural areas in between these urban centers.

[6] For example, "Online shopping is growing rapidly in China," www.bbc.com/news/business-14679595; "The Great Leap Online," www.economist.com/node/21540260; "China's E-Commerce Addition Has Serious Market Potential," www.forbes.com/sites/sarahsu/2016/07/16/chinas-growing-e-commerce-addiction. Accessed on 10/16/2016.

to consumers in any city, subject to a per-unit online shipping cost. Additionally, for each destination market, a firm has the option of selling goods offline. In order to serve a city through the offline channel, a firm must pay a fixed cost to set up a store in addition to a per-unit offline shipping cost.[7] To capture the possibility that setting up an offline store could reduce but not completely eliminate online sales from the same firm, we adopt the framework in Tintelnot (2016) to incorporate partial cannibalization of sales within a firm. The model allows the competition between the online and offline channels within a firm to be different from the competition across firms. While a firm will always sell through the online channel, it will set up an offline store only if the additional profit from the offline channels exceeds the fixed cost. As a result, only the more productive firms will set up offline stores. 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.

We estimate the shipping cost parameters and calibrate the model to match salient features of the Chinese economy. We estimate the cost parameters for offline shipping using data on inter-provincial trade flows, and the cost parameters for online shipping using data on inter-provincial e-Commerce flows and city-specific online expenditure shares. Our estimation of shipping costs produces two notable findings. On the one hand, consistent with studies based in the U.S. (e.g, Hortaçsu et al., 2009), we find that the distance elasticity of online shipping cost is about half the magnitude of that for offline shipping cost. The smaller distance elasticity for online shipping cost implies that the ratio between online and offline shipping cost is lower for more remote cities. Consequently, consumers from the remote cities buy disproportionately more from the online platforms. On the other hand, to match the online expenditure share of individual cities, we allow for a destination-specific component in our specification of online shipping cost. We find that the destination-specific component is negatively correlated with city size. This implies that, conditional on distance, the variable cost of online shipping is lower for destinations with larger populations, possibly due to their scale economy in e-Commerce distribution, concentration of better educated residents, and other agglomeration forces in e-Commerce. These forces interact to determine the spatial patterns of e-Commerce in our model.

Our calibrated model implies large differences in access to varieties of goods related to both population size and market potential. We then use the calibrated model to perform counterfactual experiments to study the effects of e-Commerce. We find that the arrival of e-Commerce raises the average share of consumption expenditures on goods produced in other cities, with larger percentage point increases for the smaller and more remote cities. The freer inter-city trade brought about by e-Commerce generates large welfare gains; for an average city, the welfare gains from e-Commerce are about 1.59%. Residents in smaller and more remote cities gain more than those in bigger and better-connected cities: under our benchmark calibration, the average welfare gains

[7] We do not impose any model restriction on the relative magnitudes of the variable shipping costs between the two channels. Nevertheless, our estimation indicates that the variable shipping cost for the offline channel is substantially lower than online shipping on average, except for city pairs very far apart, in which case the offline shipping cost might be higher than the online shipping cost.

are 2.16% for the 1st *population quintile* of cities and 1.21% for the 5th quintile; they are 1.91% for the 1st *market potential* quintile and 1.26% for the 5th quintile. Using the partial elasticities of real income with respect to population and market potential as measures of spatial inequality, we find that the arrival of e-Commerce reduces the inequality associated with population size by 1.9%, and the inequality associated with market potential by 4.1%.

We use the calibrated model to study the effect of further development in e-Commerce by decreasing the online shipping cost in the model. If online expenditure share in China were to become three times its 2013 level, the *cumulative* welfare gains from e-Commerce would be 5.12% on average.[8] Furthermore, e-Commerce would have reduced the inequality associated with population size by 5.2% and the inequality associated with market potential by 8.6%, relative to the economy without e-Commerce. Therefore, further development of e-Commerce in China would increase real income of workers in all cities, and substantially reduce inequality between them.

Our paper contributes to a new and rapidly growing strand of literature on consumption inequality across space. Recent studies show that there are substantial spatial differences in access to varieties. This includes studies on tradeable goods, especially groceries (Broda and Weinstein, 2006; Handbury, 2013; Handbury and Weinstein, 2014; Hottman, 2014) and non-tradeable goods and services such as restaurants and local amenities (Couture, 2015; Diamond, 2015). This paper contributes to this literature in two ways. First, it introduces the role of online shopping to a literature that has so far predominantly focused 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 largely 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. Notable exceptions include Faber (2014b), Atkin et al. (2015), and Feenstra et al. (2016), all of which use detailed category-level price data. The first two papers, which use data from Mexico, do not focus on spatial inequality. Feenstra et al. (2016) shows that big cities in China have more varieties offered at lower prices than small cities. In calibration, 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 such an exercise.

Our paper is related to the large literature that studies domestic trade frictions and their implications for the spatial distribution of income and welfare. Redding and Turner (2014) provide an excellent review of the relevant literature. The main focus of the literature is on transportation infrastructure, (see,for example Donaldson (2010), Faber (2014a) and Asturias et al. (2016). In

[8]We decrease the online shipping cost proportionally for all city pairs without changing the distance elasticity or the destination-specific components. We note that further development of e-Commerce may change the distance elasticity or destination-specific component. These alternative ways of decreasing online shipping cost would lead to different welfare changes.

contrast, the present paper quantifies the welfare effects of a reduction in domestic trade frictions brought about by a new distribution channel, namely the e-Commerce channel.

Our model exploits the similarity between online sales and exports. Some earlier studies have shown this similarity. Lendle et al. (forthcoming) look at international purchases on eBay and find that the effects of distance are smaller than the estimates from the off-line gravity estimation; Chen et al. (2015) use Taobao data for the electronics product category to study the role of distance for online transactions in China; Lendle et al. (2013) document 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]

Existing studies have identified specific mechanisms through which e-Commerce benefits consumers.[10] Most of these studies do not have a spatial dimension. An exception is Forman et al. (2009), who use data from top-selling books from Amazon.com and find 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) do not explicitly model the entry decision and the channel choice by firm and therefore cannot quantify the general equilibrium welfare effects, which might be important given the growing size of e-Commerce. 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.

The rest of the paper is organized as follows: Section 2 introduces the retailing industry and online shopping in China and describes the data used in the paper. Section 3 introduces the empirical approach, establishes the basic empirical relationship between online shopping and market size, and discusses the welfare implications using a simple consumption model. Section 4 presents the spatial general equilibrium model of online shopping. Section 5 calibrates the model and quantifies the welfare gains from E-commerce. Section 6 concludes.

2 Retail Industry and e-Commerce in China

The growth of e-Commerce in China has been spectacular. As Figure 1 shows, between 2008 and 2014, the total value of online retail sales grew 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 is the

[9]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) examine 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.

[10]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).

dominant online retail platform in China, accounting for 82 percent of the total online retail sales in 2014.[11]

The rise of e-Commerce in China has an important geographic dimension. Although bigger and better-connected cities on the East Coast have the highest per capita online retail consumption, smaller and more remote cities spend a higher proportion of their income online. According to a survey performed by the McKinsey Global Institute, conditional on shopping online, 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, respectively, while residents in smaller and less connected tier-3 and tier-4 cities spend 21 percent and 27 percent, respectively.[12]

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 that provide limited numbers of varieties—the top 100 retail chains in China collectively accounted for only 9% of retail sales in 2012. Furthermore, existing large retailers are traditionally concentrated in large and rich cities. For example, when foreign retail chains such as Walmart and Carrefour entered China, their first groups of stores were predominately located in the biggest metropolitan areas. This is in sharp contrast to the big-box 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. For example, the success of Walmart can be attributed to its growth in mid-sized cities.

Two sources of costs involved in traditional offline retailing contribute to its spatial distribution. First, there is a fixed cost to be paid in order to sell to a new market: it can be the cost involved in setting up a brick-and-mortar store or the cost involved in contracting with local retailers and distributors. These upfront costs can only be justified by a sufficiently-large local demand. Therefore, everything else equal, a traditional retailer is less likely to find it profitable to enter a small market. Second, the shipping costs make it more expensive to serve a remote market. With China's substantial regional inequality, the small and remote cities also tend to be much poorer, making traditional retailers even less willing to serve these markets with a competitive price.

E-Commerce is a more economical way for many producers to reach customers in small and remote markets. First, e-Commerce essentially eliminates the fixed cost involved in serving a new market. After signing up with a large e-Commerce platform such as Taobao for a small fee, there is no additional charge for a producer to sell to any market. Therefore, a producer may find it profitable to sell a small number of items to consumers in markets that it would not have found profitable to enter via traditional retailing. Shopping online is therefore particularly attractive for residents in small and remote 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.[13] Second, e-

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

[12] `https://goo.gl/fKuPcI` , exhibit 2.

[13] Ali Research and McKinsey Global Institute. `https://goo.gl/TcBAli`, Page 27.

Commerce can significantly reduce information costs associated with trade. With online shopping platforms, consumers can easily obtain information on numerous goods from a distant location. As a result, e-Commerce appears to be less hindered by distance. In fact, using transaction data from other online platforms, Lendle et al. (forthcoming) and Hortaçsu et al. (2009) find that trade volume on the online marketplace is much less affected by geographic distance than trade through the traditional offline channel. When we quantify our formal model in Section 5, we allow the elasticity of shipping cost with regard to distance to differ by online and offline. Using inter-provincial trade flows for both channels, we also find that the distance elasticity for e-Commerce is much smaller than that for offline retailing.

Lastly, one might think that the lower availability of credit card in a developing country such as China would hinder the development of e-Commerce, especially for small and remote cities. However, the development of online payment platform has helped overcome this obstacle.[14]

3 Market Size and Online Expenditure Share

3.1 Econometric Model and Measurement

We hypothesize that consumers in small and remote markets are under-served by traditional retailing and that the rise of e-Commerce is particularly beneficial to them. If this is the case, everything else equal, we would expect that consumers in smaller and more remote cities would rely more on online shopping than their counterparts in bigger and better-connected cities. In this section, we test this prediction empirically. We run various versions of the following regression:

$$\ln OnlineShoppingIntensity_i = \beta_0 + \beta_1 MarketSize_i + \beta_2 Connectedness_i + \mathbf{X}_i \cdot \beta_3 + \epsilon_i. \quad (1)$$

The outcome variable measures the relative reliance on e-Commerce for consumers in city i. In the baseline, we measure it by the expenditure on online shopping as a share of city i's total retail sales in 2013. We take the log of the share as the dependent variable. Market size is measured by city i's population in 2010. We use the log of the "market potential" measure to proxy for remoteness. As in Harris (1954), city i's market potential is measured as

$$MP_i = \sum_{i' \neq i} \frac{L_{i'}}{\tau_{ii'}}, \quad (2)$$

where L_i is the population size (in millions) in city i and $\tau_{ii'}$ is the distance (between two centroids, in kilometers) between city i and city i'. Therefore, the market potential is the weighted sum of the inverse of distance to other cities, where the weight is each city's population. As e-Commerce eliminates the entry cost and reduces the impact of distance, we would expect both β_1 and β_2 to

[14] For example, Alipay, a popular payment method developed by Alibaba Inc., facilitates the payment and refund process between sellers and buyers. Consumers can use the Alipay service if they have a basic bank account to transfer funds into their Alipay account.

be negative. We find strong evidence that for both hypotheses. Our results are robust to different functional form assumptions.[15] As we show in Section 3.4, the results are also robust to alternative measures of online shopping intensity, market size, and remoteness.

To isolate our preferred interpretation of the mechanisms from alternative explanations, we control a set of city-level characteristics in $\mathbf{X}_i$. $\mathbf{X}_i$ includes supply and demand shifters that also affect online shopping. The supply shifters include access to highways, railways, and broadband Internet, and physical distance to online sellers. The demand shifters include basic demographic characteristics of the city, such as average income, average educational attainment, share of working age population, and share of employment in the non-agricultural sector (our measure of urban rate). We include an indicator for provincial capital since, as in China's hierarchical system of cities, capital cities often have additional observed and unobserved advantages over other cities. We also include the full set of province dummies.

We estimate Equation 1 using OLS. Since both market size and remoteness are functions of population, one concern with the OLS estimate is that, if population is measured with error, the estimated coefficients suffer from attenuation bias. Population counts used here are from the 2010 Population Census, which is among the most reliable sources of statistics in China. The attenuation bias does not change the sign of the coefficient, so even if it exists, it only biases the estimates toward the null. In alternative specifications not reported here, we construct market size and remoteness measures using population counts from the 2000 census. Under the assumption that the measurement errors in two censuses are uncorrelated, the Two-Stage Least Squares (2SLS) estimate corrects the measurement error. The instrument has a very strong first stage and the 2SLS results are almost identical to the OLS results. Therefore, we report only the OLS results.

3.2 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 in the world.[16] We obtain confidential data for the total sales and purchases by city-category in 2013 from both platforms.[17]

We supplement the sales data from Taobao with data on city-level characteristics. Specifically, from the 2010 census tabulations and the 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, per

[15] The log-log specification is unit free; it simplifies the interpretation of the results: the coefficient can be interpreted as elasticities. When we present the category-level evidence, as different categories have substantially different levels of penetration in the online market, the log-log specification also makes comparison across categories convenient. The results are, nevertheless, robust to specifications in levels (or log-levels).

[16] `https://goo.gl/1Mp5eH`, accessed on 12/15/2015.

[17] 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://goo.gl/kFvNqQ` . The consolidated list of categories is reported in Appendix Table A1.

capita income and consumption, industrial composition, and shares of households with broadband connection and smartphones. We construct the distance to highways and railways for each city from the transportation network database developed by Baum-Snow et al. (2015).[18] We obtain other city-level geographic characteristics, such as the longitude, latitude, and ruggedness from the China Historical GIS database.

We perform our benchmark analyses at the city level. Our baseline sample for the empirical analyses includes 315 cities. To shed more light on the channels behind the correlation between market size, remoteness, and online expenditure share, we also conduct analyses at the city-category level. For these analyses, we construct city-by-category expenditure shares from household surveys conducted by China's National Bureau of Statistics (NBS).[19] The product categories in NBS are broader than those used by Taobao. We match Taobao categories to NBS categories.

Additionally, the quantification part of the paper in Section 5.2 requires estimating online and offline shipping costs. For this purpose, we need trade flow data for both online and offline trade. The trade flow data are at the inter-provincial level. For offline trade, we construct data on inter-provincial trade flows from the 2002 China Regional Input-Output Table. For online trade, we obtain data on inter-provincial e-Commerce sales from Taobao for 17 of its biggest categories. These 17 product categories account for 67% of total Taobao sales in 2013.

Table 1 provides the summary statistics of the key characteristics. In 2013, Taobao accounted for 7.9 percent of total retail sales in an average city. An average Chinese city has a population of about 3.9 million, and the market potential measure for the average city is 1.7. In China, smaller cities also tend to be more remote. The correlation between population and market potential is 0.4. Adults in an average city have 8.8 years of education and have a per capita income of 15 thousand *yuan* (about 2272 USD). For demand shifters, the average urban rate (measured as the share of employment in non-agricultural sectors) is 48%; the share of working age population is 74%; for every 100 women, there are 116 men. We also report the following summary statistics for other factors that affect the supply of online shooping: log distances to the nearest highway and railway, access to online sellers (also measured as the log of "market potential" as in Harris (1954)), and share of households with broadband Internet access. In an average city in 2013, 59% of households had access to broadband Internet.

3.3 Baseline Results

Table 2 shows the baseline results. Column 1 shows the simple correlations. Log online expenditure share is negatively correlated with log population. The coefficient associated with log popula-

[18]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.

[19]The household surveys include the Urban Household Survey and the Rural Fixed-Point Household Survey. These surveys sample households in cities and rural areas for their daily expenditures. Category-level per capita expenditure is available at the city level for only 105 cities.

tion is -0.085, significant at the 5% level. Log online expenditure share is also negatively correlated with log market potential. The coefficient associated with log market potential is -0.229, statistically significant at the 1% level. The signs are as expected. They indicate that consumers in smaller and more remote cities spend a larger share of their expenditure on online shopping.

We hypothesize that residents in smaller cities will spend larger shares of their expenditure online because they are under-served by traditional retailing. Though intuitive, direct evidence on this is scarce. In a recent study, Feenstra et al. (2016) use barcode-level data on a handful of products in China and show that, indeed, smaller cities have access to fewer varieties and face higher prices for the same product. However, since we do not have a measure of the traditional retailing conditions, we cannot directly test this hypothesis in our data. The negative correlations between market size, remoteness, and online expenditure share can also be a reflection of a scenario in which smaller cities have better access to the online shopping market. This may be possible when smaller cities have better access to the Internet, or, although remote from large population centers, are somehow better connected to online sellers or are systematically located in places where online purchases are easily delivered. However, this is unlikely to be the case. If anything, smaller and remote cities are likely to lag in Internet infrastructure and also be located far away from online sellers. Nevertheless, we explicitly control for these alternative supply factors. The remaining columns of Table 2 report the results.

Column 2 includes the log distances to highway and railway. These two variables are highly correlated. When included in the regression together, being close to a highway is associated with higher online expenditure share, while being close to a railway has the opposite effect. Column 3 controls for the broadband Internet penetration rate. As expected, better access to broadband Internet is associated with higher online expenditure share. In both Column 2 and Column 3, the two coefficients of interest are essentially unaffected.

Big cities are often home to large numbers of sellers, both online and offline. While being close to online sellers tends to increase online shopping, being close to offline sellers reduces it. To tease out these two competing forces, we include the log of a measure of market potential to online sellers: $MP_i^{Online} = \sum_{i' \neq i}(OnlineSell_{i'}/\tau_{ii'})$. If the distributions of online and offline sellers across space are the same, then $\ln MP_i^{Online}$ and $\ln MP_i$ will be perfectly correlated; if that is the case, then they cannot be separately estimated. The correlation between the two variables is 0.64. If remote cities are relatively closer to online sellers than to offline sellers, after controlling for $\ln MP_i^{Online}$, the coefficient associated with $\ln MP_i$ will decline in magnitude. According to Column 4, $\ln MP_i^{Online}$ has an expected positive sign while, compared with Column 1, the coefficient associated with $\ln MP_i$ remains negative and more than doubles in magnitude. This is because online sellers are disproportionately concentrated in densely populated coastal area. The concentration of online sellers is even stronger than the concentration of population.[20]

Column 5 includes a dummy indicating whether the city is a provincial capital. In China's

[20]One may argue that MP_i depends only on population and does not take into account income differences across space. If we replace population with GDP in the construction of the market potential measure, we get the same pattern. We also report robustness results using alternative measures in the next subsection.

hierarchical city structure, provincial capitals have access to more resources and better infrastructure than other cities. Residents in provincial capital cities spend more on online shopping, but the inclusion of this dummy variable does not change the coefficients of interest by much. Column 6 adds all the covariates included in Columns 2 through 5 and Column 7 adds a further set of province dummies. Throughout all columns, both log population and log market potential have negative and statistically significant signs. In Column 7, the coefficient associated with log population is -0.14, and that associated with log market potential is -0.99; both are significant at the 1% level.

Another alternative explanation for the negative correlations is that the demand for online shopping might be higher in smaller and more remote cities. One possibility is that residents in smaller cities might be younger, more educated, and more urbanized; the cost of shopping online would be lower for such people than for older, less educated, rural people. It is possible that residents in small and remote cities may find online shopping fashionable.[21] More realistically, the correlations may be due to non-homothetic preference. Residents in smaller and more remote cities are typically also poorer than their larger city counterparts. The correlation between a city's log average income and its log population (log market access) is 0.23 (0.24). Differences in income can lead to differences in consumption structure. In particular, richer households may spend a larger share of their income on services and luxury goods, which are typically not sold on Taobao. This may drive the negative correlations we find.

To rule out alternative explanations due to heterogeneity in demand, we supplement the controls in Table 2, Column 6 with a host of demand shifters such as income levels and demographic characteristics. Column 1 of Table 3 controls for log average per capita income of the city, Column 2 includes log average years of schooling among adults as a proxy for average skill level of city residents, and Column 3 includes log urban rate (measured as the share of non-agricultural employment). Column 4 includes the share of working age population and sex ratio. Because these variables are correlated with covariates already included in the regression, except for log urban rate which is marginally significant, none of these variables adds much explanatory power. Column 5 adds all the demand-side shifters included in Columns 1 through 4. Throughout these columns, the coefficients associated with log population and log market potential remain remarkably stable. In fact, if we further include all the quadratic terms of the demand and supply factors, the coefficients barely budge. Although we cannot control for all the demand shifters, the stability of the coefficients is reassuring.[22]

Column 5 is our preferred specification. It shows that the elasticity of online share with regard to log population is -0.12. An average city in the 1st quintile of the size distribution has a log

[21]However, a McKinsey survey of online shoppers indicates otherwise. It shows that "fashionable" is only the fifth most cited reason for why people buy online, way behind "access to varieties," "lower price," and "convenience in transaction." If anything, consumers in tier-1 (large and well-connected) cities are more likely than those in tier-2 and tier-3 cities to cite "fashionable" as a reason for online shopping.

[22]When we include the full set of demand factors and province fixed effects and add supply factors one-by-one as in Table 2, the coefficients of the two variables of interest also remain remarkably similar. The only exception is the inclusion of online market potential makes the coefficient associated with log market potential larger in magnitude.

population of 13.66; an otherwise similar city in the 5th quintile has a log population 15.87, the online share in the former city is about 2.1 percentage points higher than that in the latter city ($-0.12\times7.9\times(13.66-15.87)$, evaluated at the average online share). The elasticity of online share with regard to log market potential is -0.91. The difference in log market potential between the lowest and largest quintiles is -1.08 (20.52 versus 21.60); other things equal, this translates into a 7.8 percentage point difference in online share.

One remaining concern is that the relationships may not be log linear. Because the two key variables, market size and remoteness, are highly correlated, even if only one truly matters, if it is incorrectly specified as linear, the other variable may pick up the remaining correlation. To rule out this possibility, Columns 6 and 7 include a quadratic term for each key variable. Either quadratic term is statistically significant, rejecting a non-linear specification. In both cases, the coefficient associated with the other key variable does not change at all, suggesting that both market size and remoteness matter.

3.4 Robustness

3.4.1 Alternative Measures

In this subsection we show the robustness results using alternative measures of online expenditure share, market size, and remoteness. Throughout this exercise, we use the specification as in Column 5 of Table 3. Column 1 of Panel A, Table 4 replicates the baseline result.

The outcome variable in the baseline is the log of online expenditure as a share of total retail sales. Total retail sales include sales of both goods and services. As most services are not sold online, the online expenditure share calculated using total retail sales as the denominator may under-value the role of online shopping. As we mentioned earlier, this measure is also not immune to the alternative explanation that the differences in online share are due to differences in consumption structure. Column 2 shows that the results of using an alternative measure that excludes the hospitality industry (mainly hotels and restaurants) are quantitatively similar. Although we cannot purge out all service sales from the total retail sales, the stability of the results is reassuring.

We have been implicitly treating the online share of *sales* as the online share of residents' *expenditure*. However, the total sales in a city may not necessarily be equal to the total expenditure of its residents. Consumers may travel across cities for shopping. We use retail sales as the denominator mainly to be consistent with the nominator: sales realized on Taobao. In Panel A, Column 3 we use total consumption as the denominator in calculating online share.[23] Again, the results remain similar.

We use two alternative measures of market size: log total income and log total consumption. Both are measures for local demand. We also use two alternative measures of remoteness: the log

[23] Total retail sales and total consumption are calculated based on different data sources. The total retail sales are from a representative sample of retailers, while the total consumption is from a representative sample of households. The total consumption in a city is typically smaller than its total retail sales.

market access measure and an alternative measure of log market potential in which we replace population with GDP.[24] Both are measures for access to sellers. Panel B of Table 4 shows the results with different combinations of these measures. The coefficients associated with different measures are not directly comparable as the units and distributions differ, but we obtain the same correlations throughout.

3.4.2 Market Size and Online Expenditure Share by Category

As we mentioned earlier, one alternative explanation for the negative correlations between market size, remoteness and online expenditure share is non-homothetic preference. The baseline regression controls for cities' average income level, as well as a vector of demographic characteristics, to account for the differences in demand. Furthermore, we use alternative measures that are more robust to this alternative explanation and obtain the same results.

In this subsection, we construct category-specific online expenditure share, which explicitly takes into account differences in consumption composition. For this, we exploit our product-category-level sales data from Taobao and match them with city-category-level consumption data.

There are two data issues. First, the online sales data are classified in categories defined by Taobao. We start with 139 product categories and consolidate them into 81 categories. The total consumption data are classified into seven categories defined by the National Bureau of Statistics (NBS).[25] We map each Taobao-defined category into one of the NBS categories. The crosswalk is listed in Appendix Table A1. The second issue involves the coverage of the category-level total consumption data. Category-specific consumption data are only available for a fraction of cities, and for a given city, not all the categories are available. Therefore, for this exercise, we use about 100 cities that report these data.

We run Equation 1 for each NBS category using the same specification as in Column 5 of Table 3. The results are reported in Table 5. The first thing to note is that the online expenditure share differs substantially across categories. The categories with the largest online expenditure shares are clothing (46% sold online) and household appliances and services (about 32% sold online). Categories that have the smallest online expenditure share are residence and related goods and services (less than 0.01%) and food (less than 1%). Not surprisingly, the two most important categories for Taobao are clothing and household appliances and equipment, accounting for 44% and 29% of total Taobao sales, respectively. Although average online expenditure shares differ by category, the log-log specification renders the coefficient to be neutral to the relative size of each category and makes the comparison easier.

The estimates are much less precise than the baseline results in which we look at the aggregate online share. The imprecision in estimates may be due to the much smaller sample size and noisier measures. The measurement error may come from multiple sources. First, there is likely to be

[24]Market access is given by $MA_i^{Online} = \sum_{i' \neq i}(L_{i'}/\tau_{ii'}^{\theta})$, where we use $\theta = 8.22$ as in Donaldson and Hornbeck (2016). The optimal θ may be different in our setting, but the results are not sensitive to the choice of θ.

[25]These categories are food; clothing; residence-related goods and services; household appliances and services; health care and medical services; transportation and communications; recreation, education and culture.

much noise in more disaggregated consumption data that are obtained from household surveys. Second, to the extent that industries are clustered, our market potential variable, which is based on population or GDP, may mis-measure the access to a specific category. Nevertheless, we find that the coefficients associated with both log population and log market potential are overall negative. The log market potential measure only has a positive sign for the food and household appliances categories. In particular, for clothing, the most important category for Taobao, both coefficients are negative and statistically significant.

3.5 Welfare Implications of the Empirical Results

Our empirical analysis finds that residents in smaller and more remote cities rely more on online shopping. How does the different online expenditure share translate into differential gains from e-Commerce for residents in different cities? In Appendix II, we develop a simple model to answer this question. We derive the gains from e-Commerce, defined as the percentage change in welfare as a city moves from a case without e-Commerce to one with e-Commerce to be

$$\text{Gains from e-Commerce} = (1-\lambda_i)^{\frac{1}{1-\sigma}} - 1, \tag{3}$$

where λ_i is the share of expenditures spent on the online good and $\sigma > 1$ is the elasticity of substitution between the online and local goods. Therefore, welfare gains from e-Commerce are an increasing function of online expenditure share. 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 3 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 $\sigma = 5$, which is in the range of the values commonly used in the literature, the average gains from e-Commerce are $(1-0.08)^{-\frac{1}{3}} - 1 = 2.1\%$. 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 $(1-0.123)^{-\frac{1}{4}} - 1 = 3.3\%$. This value is more than 50% higher than that of a city with the average online expenditure share. Therefore, there is large spatial dispersion across cities in the gains from e-Commerce. Crucially, together with our empirical results above, the welfare formula shows that residents in smaller and more remote cities enjoy larger gains from e-Commerce.

4 The Quantitative Framework

The simple model in Appendix II focuses on the consumption decisions of residents and abstracts from decisions by firms. In this section, we build a multi-region general equilibrium model of online shopping by incorporating the decisions by firms. The model is similar in spirit to that of Helpman et al. (2004) in which firms may pay a fixed cost of FDI to take advantage of a lower variable cost. To capture the possibility that a producer that sets up an offline store could reduce but not completely eliminate online sales, we adopt the framework in Tintelnot (2016) to allow for partial cannibalization of sales within a firm.

4.1 Preference

There are N regions, each corresponding to a prefecture city in the data. We index the regions using $i \in \{1, 2, \cdots, N\}$ or $j \in \{1, 2, \cdots, N\}$. City i is endowed with population L_i. A worker in city i supplies one unit of labor inelastically and receives a wage of w_i, which is determined endogenously.

The preference of the representative consumer in region j is given by

$$U_j = (u_j^T)^{\beta_j^T} (u_j^{NT})^{\beta_j^{NT}},$$

where u_j^T and u_j^{NT} are subutility functions defined below and β_j^T and β_j^{NT} are preference parameters. We assume that β_j^T and β_j^{NT} are region-specific to allow for preference differences across regions.

In each sector $h \in \{T, NT\}$, a good is indexed by a firm ω and a variety v. Each firm in each sector produces a unit measure of varieties. The subutility function for sector $h \in \{T, NT\}$ of the representative consumer in region j is

$$u_j^h = \Big(\int_{\Omega_j^h} \int_0^1 q_j(w, v)^{(\sigma^h - 1)/\sigma^h} dv dw \Big)^{\sigma^h/(\sigma^h - 1)}, h \in \{T, NT\},$$

where Ω_j^h is the set of all sector h firms selling in city j. We assume the same elasticity of substitution σ^h between products both inside and outside the firm.[26]

4.2 Firms in the Tradeable Sector

We describe the firms in the tradeable sector before moving to the nontradeable sector. There is an exogenous mass M_i^T of tradeable firms, which are collectively owned by the workers in region i. A firm is characterized by its city of origin i, and its productivity parameter ϕ. Each firm receives a productivity draw ϕ from a common cdf $F_i(\phi)$ with support of $[\underline{\phi}_i, \infty)$. We assume that $F_i(\phi)$ follows a Pareto distribution, $F_i(\phi) = 1 - (\frac{\phi}{\underline{\phi}_i})^{-\alpha}$, where α is the shape parameter and $\underline{\phi}_i$ is the technology level of region i.

Each firm is infinitesimal and takes aggregate quantities as given. A firm acts as a monopolist and faces the CES demand function for each of its varieties. Quantity demanded in region j for

[26] As should be clear shortly, the model setup for the nontradeable sector is isomorphic to the case in which each firm in the nontradeable sector produces only a single variety of goods. The current assumption maintains the symmetry between the two sectors and simplifies the notation.

variety v supplied by firm w at price $p_j(w,v)$ is given by

$$q_j^T(w,v) = \frac{\beta_j^T Y_j p_j(w,v)^{-\sigma^T}}{(P_j^T)^{1-\sigma^T}}, \tag{4}$$

where Y_j is the aggregate income in region j and P_j^T is the aggregate price index for the tradeable sector in region j given by

$$P_j^T = \Big[\int_{\Omega_j^T} p_j(w)^{1-\sigma^T} \Big]^{1/(1-\sigma^T)}.$$

The price index for a firm w, $p_j(w)$ is a CES aggregate over its unit measure of varieties

$$p_j(w) = \Big(\int_0^1 p_j(w,v)^{1-\sigma^T} dv \Big)^{1/(1-\sigma^T)}.$$

Lastly, the expenditure on goods produced by firm w in region j is $s_j(w) = \beta_j^T Y_j \frac{p_j(w)^{1-\sigma^T}}{(P_j^T)^{1-\sigma^T}}$.

4.2.1 Channels and Cost of Distributions

There are two potential channels $m \in \{E, P\}$, corresponding to "E-Commerce (online)" and "Physical (offline)," respectively, through which a firm can sell each variety to consumers in a given market. For each market, a firm may have one channel of distribution (henceforth referred to as an ONline-only firm and indexed by "ON") or two channels of distribution (henceforth referred to as a Two-Channel firm and indexed by "TC").

We assume that each firm is able to sell online to all regions without incurring any fixed cost, subject an iceberg cost. To sell one unit of goods to consumers in city j through the online channel, a firm has to ship τ_{ij}^E units from city i. For each market *outside the home market*, a firm has the additional option of opening a physical store by paying a fixed f_P units of labor in the destination. The iceberg cost of serving market j from market i through the offline channel is given by τ_{ij}^P. We interpret the iceberg shipping costs τ_{ij}^P and τ_{ij}^E in our model as the costs incurred in getting one unit of goods to consumers in addition to the cost of production, including the transportation cost, information cost and distribution cost.[27]

For each variety, the representative consumer in each city has a pair of random taste shocks (v^E, v^P). Each physical unit of this variety produced by a firm translates into v^P *quality* units through the offline store, and v^E *quality* units in the online marketplace. Therefore, the marginal cost of bringing one *quality* unit to the consumers in city j is $\frac{w_i \tau_{ij}^E}{\phi v^E}$ for the online channel, and $\frac{w_i \tau_{ij}^P}{\phi v^P}$ for the offline channel, respectively. Since a two-channel firm will pick the lower-cost channel to serve a variety to a given city, the cost for the variety will be $\min(\frac{w_i \tau_{ij}^P}{\phi v^P}, \frac{w_i \tau_{ij}^E}{\phi v^E})$ for the two-channel firm. By contrast, the cost for an online-only firm is $\frac{w_i \tau_{ij}^E}{\phi v^E}$.

Following the vast literature pioneered by Eaton and Kortum (2002) and adapted by Tintelnot

[27] See Anderson and van Wincoop (2004).

(2016) to model sales cannibalization within a firm, we assume the taste shocks are independent across varieties and between channels, with the following distribution:

$$Pr(v_j^m \leq x) = \exp(-x^{-\theta^T}). \tag{5}$$

To serve region j using a given channel m, the cumulative density function (cdf) of marginal cost across varieties for a firm is then

$$Pr(\frac{\tau_{ij}^m w_j}{\phi v^m} \leq c) = 1 - \exp(-(\frac{\tau_{ij}^m w_i}{\phi})^{-\theta^T} c^{\theta^T}),$$

for $m \in \{P, E\}$.

4.2.2 Two-Channel Firms

We first derive the profit function of a two-channel firm from region i selling to region j. Since the firm will pick the lower-cost channel to serve a variety, the cost for the variety will be $\min(\frac{w_i \tau_{ij}^P}{\phi v^P}, \frac{w_i \tau_{ij}^E}{\phi v^E})$. With the distributional assumption in Equation 5, for a two-channel firm, the cdf of marginal cost across varieties is

$$G_j^{TC}(c) = 1 - \exp\Big(-\sum_{m\in\{P,E\}} (\frac{\tau_{ij}^m w_i}{\phi})^{-\theta^T} c^{\theta^T}\Big).$$

As is well known, with CES utility and monopolistic competition, the firm charges a constant markup of $\sigma^T/(\sigma^T-1)$ over marginal cost for each of its products. The firm-level price index is

$$p_{ij}^{TC}(\phi) = \frac{\kappa^{T\frac{1}{1-\sigma^T}} w_i}{\phi}\Big(\sum_{m\in\{P,E\}} (\tau_{ij}^m)^{-\theta^T}\Big)^{-\frac{1}{\theta^T}}, \tag{6}$$

where $\kappa^T = \Gamma(\frac{\theta^T+1-\sigma^T}{\theta^T})(\frac{\sigma^T}{\sigma^T-1})^{1-\sigma^T}$ is a constant. Total sales of a firm from region i with productivity ϕ in market j is given by

$$s_{ij}^{TC}(\phi) = \frac{\beta_j^T Y_j \kappa^T w_i^{1-\sigma^T}}{(P_j^T)^{1-\sigma^T}}\Big(\sum_{m\in\{P,E\}} (\tau_{ij}^m)^{-\theta^T}\Big)^{-\frac{1-\sigma^T}{\theta^T}} \phi^{\sigma^T-1},$$

and firm profit in market j is a constant fraction of sales revenue with $\pi_{ij}^{TC}(\phi) = \frac{1}{\sigma^T} s_{ij}^{TC}(\phi)$.

The share of a two-channel firm's online sales in total sales, ρ, is given by $\rho = \frac{(\tau_{ij}^E)^{-\theta^T}}{\sum_m (\tau_{ij}^m)^{-\theta^T}}$. Therefore, θ^T determines the relative sales by the two channels within a two-channel firm. In Appendix III, we derive the elasticity of e-Commerce sales of a two-channel firm with respect to the offline shipping cost to be $\theta^T - \sigma^T$.

4.2.3 Online-Only Firms

If a firm does not have a physical store in market j, it serves that market through the online channel for its unit continuum of varieties. The marginal cost for a variety is $\frac{w_i \tau_{ij}^E}{\phi v^E}$. The cdf of the marginal cost to serve region j is

$$G_j^{ON}(c) = 1 - \exp\Big(- (\frac{\tau_{ij}^E w_i}{\phi})^{-\theta^T} c^{\theta^T} \Big),$$

and the price level of the firm is given by

$$p_{ij}^{ON}(\phi) = \frac{\kappa^{T\frac{1}{1-\sigma^T}} w_i \tau_{ij}^E}{\phi}. \tag{7}$$

Total sales of a firm from region i with productivity ϕ in market j is

$$s_{ij}^{ON}(\phi) = \frac{\beta_j^T Y_j \kappa^T w_i^{1-\sigma^T}}{(P_j^T)^{1-\sigma^T}} \big(\tau_{ij}^E\big)^{1-\sigma^T} \phi^{\sigma^T - 1},$$

and firm profit in market j is given by $\pi_{ij}^{ON}(\phi) = \frac{1}{\sigma^T} s_{ij}^{ON}(\phi)$.

4.2.4 Offline Entry

For any firm in the tradeable sector, setting up a physical store in any region would decrease the online sales while increasing the total sales in the region by the firm. The firm would set up a physical store in region j only if the marginal profit from having the store exceeds the fixed cost, $\pi_{ij}^{TC}(\phi) - \pi_{ij}^{ON}(\phi) > w_j f^P$. A firm from region i sets up a physical store in region j if and only if $\phi > \phi_{ij}^*$ where

$$\phi_{ij}^* = \Big(\frac{\kappa^T \beta_j^T Y_j}{\sigma^T w_j f^P}\Big)^{\frac{1}{1-\sigma^T}} \frac{w_i}{P_j^T} \Big[\Big(\sum_{m \in \{P,E\}} (\tau_{ij}^m)^{-\theta^T} \Big)^{-\frac{1-\sigma^T}{\theta^T}} - (\tau_{ij}^E)^{1-\sigma^T} \Big]^{\frac{1}{1-\sigma^T}} \tag{8}$$

for any $j \neq i$. [28]

4.3 Firms in the Non-Tradeable Sector

There is an exogenous mass M_i^{NT} of nontradeable firms in region j. For simplicity, we assume that the nontradeable sector is the same as the tradeable sector, except that $\tau_{ij}^P = \infty$ for $i \neq j$ and $\tau_{ij}^E = \infty$ for any (i,j) pair. As in the tradeable sector, a firm is characterized by a productivity level ϕ drawn from the same Pareto distribution $F_i(\phi)$ as firms in the tradeable sector. As before, there is a taste shock with dispersion parameter θ^{NT}. Each firm acts as a monopolist and faces the CES demand function for its variety. Quantity demanded in region j for a variety v of firm w is given

[28] Since we assume that each firm can sell offline to the home region without incurring f_P, we have $\phi_{ii}^* = \underline{\phi}_i$.

by

$$q_j^{NT}(w,v) = \frac{\beta_j^{NT} Y_j p(w)^{-\sigma^{NT}}}{(P_j^{NT})^{1-\sigma^{NT}}}, \tag{9}$$

where the price index of nontradeable goods in region j, P_j^{NT}, is defined as

$$P_j^{NT} = \Big[\int_{\Omega_j^{NT}} p_j(w)^{1-\sigma^{NT}} \Big]^{1/(1-\sigma^{NT})}.$$

The various functions of a firm in the non-tradeable sector with productivity ϕ, including price index $p_j^{NT}(\phi)$, total sales $s_j^{NT}(\phi)$, firm profit $\pi_j^{NT}(\phi)$, are defined analogously to their counterparts for the tradeable sector.

4.4 Aggregate Household Income

We assume that all the domestic firms in both sectors are owned by a representative household. The aggregate income in region j is then given by

$$Y_i = w_i L_i + M_i^T \sum_{j=1}^{N} \Big[\int_{\underline{\phi}}^{\phi_{ij}^*} \pi_{ij}^E(\phi) dF_i(\phi) + \int_{\phi_{ij}^*}^{\infty} (\pi_{ij}^{PE}(\phi) - f^P w_j) dF_i(\phi) \Big] + M_i^{NT} \int_{\underline{\phi}^{NT}}^{\infty} \pi_j^{NT}(\phi) dF_i(\phi). \tag{10}$$

4.5 The Equilibrium

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

1. Equation 4 and Equation 9 are the solution to the consumer's optimization problem.
2. Firms' choices of price and sales channel are optimal. Optimal prices are given by Equations 6 and 7 while channel choice is characterized by Equation 8.
3. The equilibrium price in both sectors of any city is consistent with the distribution of wages and firms' channel choice. The price index for the tradeable sector in region j is

$$\begin{aligned} (P_j^T)^{1-\sigma^T} &= \int_{\Omega_j^T} (p_j(w))^{1-\sigma^T} dv \\ &= \sum_{i=1}^{N} M_i^T \Big[\int_{\underline{\phi}}^{\phi_{ij}^*} (p_{ij}^{ON}(\phi))^{1-\sigma^T} dF_i(\phi) + \int_{\phi_{ij}^*}^{\infty} (p_{ij}^{TC}(\phi))^{1-\sigma^T} dF_i(\phi) \Big]. \end{aligned} \tag{11}$$

   The price index for for the nontradeable sector in region j is given by

$$(P_j^{NT})^{1-\sigma^{NT}} = M_j^{NT} \int^{\infty} (p_j^{NT}(w))^{1-\sigma^{NT}} dw. \tag{12}$$

4. Labor market clears in each region. 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 $L_i = L_i^T + L_i^{NT}$. Total labor demand from the tradeable sector in city i, L_i^T, is in turn given by

$$L_i^T = \underbrace{M_i^T \sum_{j=1}^{N} \Big(\int_{\underline{\phi}}^{\phi_{ij}^*} \frac{\sigma^T - 1}{\sigma^T} \frac{s_{ij}^{ON}(\phi)}{w_i} dF_i(\phi) + \int_{\phi_{ij}^*}^{\infty} \frac{\sigma^T - 1}{\sigma^T} \frac{s_{ij}^{TC}(\phi)}{w_i} dF_i(\phi) \Big)}_{\text{labor used as variable production cost}}$$
$$+ \underbrace{f^P \sum_{j=1}^{N} M_j^T \int_{\phi_{ji}^*}^{\infty} dF_i(\phi)}_{\text{labor used in setting up physical stores}} .$$

Total labor demand from the nontradeable sector in city i, L_i^{NT}, is

$$L_i^{NT} = \underbrace{M_i^{NT} \int_{\underline{\phi}}^{\infty} \frac{\sigma^{NT} - 1}{\sigma^{NT}} \frac{s_i^{NT}(\phi)}{w_i} dF_i(\phi)}_{\text{labor used as variable production cost}} .$$

5. Y_i satisfies Equation 10.

4.6 Measure of Welfare

The real income of a worker in city i, W_i, is given by

$$W_i = \frac{w_i}{(P_i^T)^{\beta_i^T} (P_i^{NT})^{\beta_i^{NT}}},$$

where the sector-specific price indices P_i^T and P_i^{NT} are defined in Equations 11 and 12. The percentage gains from e-Commerce for workers in city i are then given by

$$\text{Gains from e-Commerce} = \frac{W_i^{\text{E}}}{W_i^{\text{pre-E}}} - 1 \approx \ln W_i^{\text{E}} - \ln W_i^{\text{pre-E}}. \tag{13}$$

where W_i^{E} and $W_i^{\text{pre-E}}$ are real income and the superscripts "E" and "pre-E" refer to model economies with and without E-Commerce, respectively. Equation 13 gives us the key outcome variable of interest for our quantification exercise in Section 5.

4.7 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 to reach 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 relationships with regional or local retailers. The costs incurred during this process are 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. These two assumptions together imply fixed markups and rule out any pro-competitive effects of e-Commerce—when facing 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, 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.

5 Quantification

In order to study the effects of e-Commerce on welfare and inequality in the model, we calibrate the model to match the salient features of the Chinese economy. Each region in the model corresponds to a prefecture city in the data. Our calibrated model includes a total of 325 prefecture cities.

5.1 Parameters Assigned Externally

We pick the Cobb-Douglass preference parameter β_j^{NT} to match the share of services in total household expenditure from the NBS surveys data and then set $\beta_j^{T} = 1 - \beta_j^{NT}$.[29] We set the elasticity of substitution σ to 5 for both the tradeable sector and the non-tradeable sector, which is in the middle range of values commonly used in the literature. For example, Melitz and Redding (2015) use a value of 4 while Tintelnot (2016) employs a value of 6. Bernard et al. (2003) obtain an estimate of 3.79 using US manufacturing data.

We set the measure of firms in each region to be proportional to population, as in Tintelnot (2016). The shape parameter of the productivity distribution α governs the domestic firm sales

[29] Recall that we allow β_j^{NT} and β_j^{T} to represent preference differences across prefecture cities. Since we only have the service expenditure data at the provincial level, we follow the procedure outlined in Section 3.4.2 to impute service expenditure at the prefecture level.

distribution. To match a shape parameter of 1.3 for firm sales, which is in the range of empirical estimates, we set the Pareto shape of productivity α to be 5.2.[30].

A key parameter of our model is the dispersion of taste shocks θ_T. As Appendix III shows, $\theta^T - \sigma^T$ governs the extent of sales cannibalization between channels within a firm. In principle, θ^T can be estimated using online sales data and offline sales data at the firm level. Unfortunately, our data do not permit such estimation. In Tintelnot (2016), the strength of cannibalization between export platforms is captured by $\theta - \sigma = 1$. With these caveats in mind, we set $\theta^h = 6$ for both sectors so that $\theta^h - \sigma^h = 1$.

5.2 Parameters Determined in Equilibrium

In addition to the parameters discussed above, we have four sets of parameters to be estimated or calibrated, including the technology parameter for city i, $\underline{\phi}_i$; offline shipping costs τ_{ij}^P; online shipping costs τ_{ij}^E; and the fixed cost of offline entry f_p. We set values for these parameters using a two-step procedure. Towards this end, we employ data on the Chinese economy including inter-provincial offline trade flows, inter-provincial online e-Commerce flow, nominal income by city, and online expenditure share by city.

In the first step of our calibration procedure, we set the online shipping cost to infinity. We parametrize the offline costs τ_{ij}^P as $\ln \tau_{ij}^P = \delta \ln D_{ij}$, where D_{ij} is the bilateral distance between city i and city j. We choose this parsimonious specification since we only have data at the inter-provincial level while the model is constructed at the city level.[31] For a given value of δ, we pick the fixed cost of offline entry f_p so that the share of firms with offline sales to any other province is 78%, as in the 2004 World Bank Investment Climate Survey Data for China. We pick the technology parameter for city i, $\underline{\phi}_i$, to match the nominal income of each city in the model to the data from 2005 mini-census data. Finally, we look for the coefficient δ to minimize the distance between inter-provincial trade flow in China in 2002 and the model counterpart. The first step of our calibration procedure produces our model economy before the emergence of e-Commerce.

In the second step, we take the offline trade costs τ_{ij}^P, fixed cost f_p, and the measure of firms from the first step, and the calibrated online shipping cost τ_{ij}^E. We parametrize the e-Commerce shipping cost τ_{ij}^E as $\ln \tau_{ij}^E = \gamma_0 + \gamma_1 \ln D_{ij} + \epsilon_j$. The constant term γ_0 captures the cost of online transactions relative to offline transactions while ϵ_j is a destination-specific component of online shipping cost, with the sum normalized to 0. For given values of γ_0 and γ_1, we pick ϵ_j to match the online expenditure share for each city. Lastly, we look for γ_0 and γ_1 to minimize the distance of inter-provincial e-Commerce flows and city-level online expenditure share between the model and the 2013 data. This step produces our model economy with e-Commerce.

[30]($\alpha = (\sigma - 1) \times 1.3 = 5.2$). Our conclusions are robust to alternative values for α between $(\sigma - 1) \times 1.1$ and $(\sigma - 1) \times 1.5$.

[31]An alternative specification of τ_{ij}^P is $\ln(\tau_{ij}^P) = \delta_0 + \delta \ln(D_{ij})$. Since we have normalized τ_{ii}^P to 1 for all i, δ_0 captures the level of shipping cost between two different cities relative to trading within a city. However, our data at the inter-provincial level does not allow us to reliably identify δ_0.

To summarize, we match both the pre e-Commerce nominal income and the online expenditure share to the data, for all 325 cities. In addition, our shipping cost parameters minimize the distance in inter-provincial trade flows and e-Commerce flows between the model and data. We calculate the standard errors for our estimates using a bootstrap procedure.[32]

Table 6 presents our estimates of shipping cost parameters. We estimate the distance elasticity of online shipping cost to be 0.2378, with a standard error of 0.0039.[33] In contrast, we estimate the distance elasticity of online shipping cost γ_1 to be 0.1194, about half the magnitude of the estimate for offline shipping, with a standard error of 0.0124.[34] The difference in distance elasticity is consistent with Lendle et al. (2013) who find that, in the international context, the negative effect of distance is 65% smaller for trade flows on eBay than for total trade flows. According to our estimates, the implied bilateral offline shipping cost ranges from 1.584 to 7.377, with an average of 5.394. On the other hand, the iceberg cost for online shipping, including that within a city, ranges from 1.965 to 12.79, with an average of 7.337. Figure 2 plots the log of shipping costs τ_{ij}^P and τ_{ij}^E against the log of bilateral distance, over the range of bilateral distance in our data. Since we have set $D_{ii} = 1$ km for all cities in our estimation, the value for $\ln(D_{ij}) = 0$ in Figure 2 corresponds to the shipping cost within a city. Given that offline trade still accounts for the bulk of total trade in the data, it is not surprising that the online shipping cost is higher than the offline cost for most of the bilateral city pairs in Figure 2. On the other hand, since the online shipping cost increases with distance more gradually than the offline shipping cost, the online shipping cost eventually becomes comparable to or even lower than the offline shipping cost when two cities are very far apart. The lower distance elasticity for the online channel implies that the ratio between online shipping cost and offline shipping cost is lower for more remote cities. Consequently, consumers from more remote cities would buy disproportionately more from the online marketplace.

Since welfare gains from e-Commerce are most closely related to the online expenditure share, in order to calculate the welfare gains for the city, it is crucial to match the online expenditure share of each city with the data. Towards this end, we include a destination-city-specific component ϵ_j in our specification of the online shipping cost. This component ϵ_j captures the disadvantage of online shipping to a city.[35] In Table A2, we analyze the relationship between the recovered ϵ_j and observed city characteristics. In Column 1, we regress ϵ_j on the log of population in a univariate regression. The coefficient is -0.0950 and significant at 1%, indicating that there is substantial scale economy for online shipping. We add broadband penetration rate and urban rate as additional

[32] We repeat the joint estimation procedure 500 times. For each repetition, we resample (with replacement) the inter-provincial pairs. We calculate the objective function according to these resampled interprovincial pairs and minimize the resulting objective function accordingly.

[33] This number needs to be combined with the trade elasticity in order to produce the distance elasticity of trade flows, which is the coefficient on the log of bilateral distance in a typical gravity equation estimation. For a trade elasticity of 4, our estimate of 0.2378 corresponds to a distance elasticity for trade flow of 0.9512, which is very close to the estimates in international trade.

[34] The standard error for γ_1 is about three times as large as that for δ. This is explained by the fact that, for each repetition of our bootstrap procedure, the estimate of γ_1 is based on the estimate of δ in our two-step estimation procedure.That is, the dispersion in estimates of δ contributes to the dispersion in estimates of γ_1.

[35] Since our specification of offline shipping cost $\ln(\tau_{ij}^P) = \delta \ln(D_{ij})$ does not allow for a destination-city-specific component, ϵ_j also absorbs destination-city-specific differences in offline shipping cost not captured by bilateral differences.

covariates to the regression in Columns 2 and 3, respectively. We add a dummy variable for being a provincial capital and provincial fixed-effects, respectively, in Columns 4 and 5. The coefficient on the log of population remains negative and significant at 1% throughout Table A2. According to Column 5, all else equal, the iceberg online shipping cost is lower for a destination city with a larger population, better access to the Internet, and higher urban share of employment.

Our calibrated ϵ_j implies that large cities have significant cost advantage in buying goods online, which in turn reduces the gap in online expenditure share between large and small cities. Put differently, if online shipping cost were purely determined by bilateral distance such that $\ln(\tau_{ij}^E) = \gamma_0 + \gamma_1 \ln(D_{ij})$, the gap in online expenditure share between large and small cities would be larger than what is observed in the data. One explanation is that small cities are still severely under-served by courier services.

5.3 Spatial Consumption Inequality before e-Commerce

In this subsection, we analyze the spatial inequality through the lens of the model. We focus on the tradeable sector in our discussion of the online expenditure share, import ratio, and offline varieties, but we take into account the non-tradeable sector in our calculation of real income.

Panel A of Table 7 summarizes the key variables in the economy with e-Commerce by city quintile of population while Panel B does so by city quintile of market potential. When the online shipping cost is infinite, online expenditure share is zero for all cities. We report the import ratio of each city, defined as total expenditure on goods produced in other cities as a share of total expenditure produced in other cities. As Panel A shows, the import ratio ranges from an average of 21.29% for the 20% of cities with the largest populations, to an average of 33.81% for the 20% with the smallest populations. The negative correlation between import ratio and city size is intuitive: smaller cities have a more limited number of varieties and therefore tend to import more (from other cities) as a share of their total expenditure.

There are substantial differences in real income associated with city population. As Table 7 shows, the average real income for cities in the 5th population quintile is 77.4% higher than the average for the 1st population quintile. The population elasticity of real income in our calibrated economy is 0.219. In our calibration procedure, we have matched the nominal income of each city to the data. Therefore, the population elasticity of nominal income in the model is exactly the same as in the data, at 0.065. The difference in population elasticity between nominal and real income is explained by the variation in ideal price index across cities.

As Figure 4 shows, the total number of firms selling through the offline channel at each location, including both local firms and firms from other cities, differs substantially across cities along both the population dimension (Figure 4a) and the market potential dimension (Figure 4b). In our calibration, cities in the largest quintile have access to three times as many offline stores than cities in the smallest quintile. The differential access to varieties of goods across cities contributes to the variation in ideal price index. The population elasticity of ideal price index in the model is -0.154. For comparison, Feenstra et al. (2016) uses the Homescan Data from Nielsen China and find that the corresponding elasticity for the price index is between -0.02 and -0.05 for a small number of product categories. Aside from the fact that the analysis in Feenstra et al. (2016) employ very different methodology and focuses on only four product categories, another explanation for the difference is that their analysis is limited to selected large Chinese cities.

Panel B examines the spatial inequality by city quintile of market potential. There is a positive correlation between import ratio and market potential. This is intuitive since well-connected cities tend to buy more from other cities. We find that there are also substantial differences in real income associated with market potential. The real income of an average city in the 5th quintile of market potential is 33.5% higher than an average city in the 1st quintile.

5.4 The Effects of e-Commerce on Inter-city Trade and Welfare

Table 8 summarizes the changes in the various outcome variables due to the arrival of e-Commerce. Panel A of Table 8 examines these changes along the population dimension. In our calibration procedure, we match the online expenditure share of each city to the data. As a result, online expenditure share is negatively correlated with population. The arrival of e-Commerce has large effects on the patterns of inter-city trade. Consider the cities in smallest population quintile. The average import ratio for these cities increases from 33.81% to 39.92%, of which 9.88% is conducted through the online channel and 30.04% through the offline channel. As Table 8 shows, while the average import ratio for these cities (in the smallest population quintile) increases by 6.11 percentage points, *offline* expenditure on goods produced in other cities, as a share of total tradeable expenditure, decreases by 3.77 percentage points. With the reduction of online shipping cost, some firms may find it more profitable to close their offline stores in certain destinations to save on the fixed cost and to serve these destinations exclusively through the online channel instead. Figure 4 shows that e-Commerce indeed crowds out offline stores in the model. Therefore, *online* inter-city trade crowds out *offline* inter-city trade in our model. More importantly, the effect of e-Commerce on inter-city trade varies across cities of different sizes. The increase in import ratio ranges from 3.8% to 6.1%, with higher increases for cities with smaller populations. Panel B of Table 8 shows that similar patterns are present along the market potential dimension.

The gains from e-Commerce for a city in our sample is 1.59%. To put this welfare number into perspective, consider the gains from international trade for the United States. Arkolakis et al. (2012), using their sufficient-statistics formula, find that the gains from international trade for the US are on the order of 1%. Therefore, the gains from e-Commerce for Chinese cities are substantial.

As Figure 3a shows, the welfare gains of each city are highly correlated with the share of total expenditure spent through the online channel. Nevertheless, the correlation between online expenditure share and gains from e-Commerce is far from perfect. Figures 3b and 3c show that the gains from e-Commerce decrease with city population and market potential, respectively. Table 8 presents the welfare gains from e-Commerce by population quintile and by market potential quintile of cities. Along the population dimension, the increase in real income due to e-Commerce ranges from an average of 1.23% for the 5th quintile to an average of 2.18% for the 1st quintile.

Along the market potential dimension, the increase in real income due to e-Commerce ranges from an average of 1.26% for the 5st quintile to an average of 1.91% for the 1st quintile.

Table 9 examines the roles of population and market potential jointly using linear regressions. The dependent variables in Columns 1 and 2 are log real income in the economies without and with e-Commerce, respectively. As Column 1 shows, without e-Commerce, the coefficients on log population and log market potential are 0.207 and 0.0425, respectively, indicating spatial income inequality along these two dimensions. The arrival of e-Commerce reduces inequality along both dimensions. In Column 2, the coefficients on log population and log market potential are reduced to 0.203 and 0.0408, respectively. The dependent variable in Column 3 is welfare gains from e-Commerce, which are equal to the difference between the dependent variables in Columns 1 and 2. According to Column 3, a 10% decrease in population is associated with greater welfare gains of 0.0399 percentage points, while a 10% decrease in market potential is associated with greater welfare gains of 0.0172 percentage points. Taking the ratios of the coefficients from Column 1 and Column 3 of Table 9, we can obtain two measures of the effect of e-Commerce on spatial inequality. According to the quantification exercise, the arrival of e-Commerce reduces the inequality associated with population size by 1.9% and the inequality associated with market potential by 4.1%.[36]

In Column 4 of Table 9, we add the interaction of log population and log market potential as another right-hand-side variable. The coefficient on the interaction is positive and statistically significant. Loosely speaking, this indicates that population and market potential are substitutes in influencing the welfare gains from e-Commerce.

5.5 Alternative Specifications of Online Shipping Costs

In our calibrated model, the online shipping cost differs from the offline shipping cost in two aspects: online shipping cost has a destination-specific component and has a smaller distance elasticity. To explore how our results are affected by these attributes of online shipping cost, we conduct two additional exercises with alternative specifications of online shipping cost in Appendix V: eliminating the destination-specific components and setting the online shipping cost to be proportional to offline shipping cost. We find that the slope between log online share and log population would be too negative relative to the data if we eliminate the destination-specific component in online shipping cost. Moreover, simply setting the online shipping cost to be proportional to offline shipping cost can produce the negative relationship between online expenditure share and population to some extent, but would under-predict the negative relationship between online expenditure share and market potential. Therefore, both features are important for the patterns of online expenditure share and welfare gains from e-Commerce in our model.

[36]Specifically, $\frac{-0.00399}{0.207} = -1.9\%$ and $\frac{-0.00172}{0.0425} = -4.1\%$.

5.6 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 subsection, we examine the potential for e-Commerce to further improve consumer welfare and reduce spatial consumption inequality. We lower the e-Commerce shipping cost τ_{ij}^{E} to triple the online expenditure share. The total revenue of e-Commerce in China is projected to increase by as much as 225% between 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 eight-year period (Dobbs et al., 2013). Therefore, the experiment considered in this section are well within reach in the near future. We decrease the e-Commerce cost wedges according to $\tau_{ij}^{E\prime} = k\tau_{ij}^{E}$, where k is a common factor. We pick k so that the overall online expenditure share for the economy is about 24.0%, three times its 2013 value.

We compare the outcomes under this experiment to the calibrated economy without e-Commerce to obtain the *cumulative* effects of e-Commerce. Table A4 summarizes the results of the experiment. Figures 5a and 5b plot the cumulative welfare gains against population and against market potential, respectively. As a result of the further reduction in online shipping cost, the cumulative percentage increase in real income for an average city is quite large, at 5.1%. Importantly, as Figure 5 shows, the cumulative welfare gains from e-Commerce are negatively correlated with log population and with log market potential. The slopes for the economy with higher levels of e-Commerce in Figure 5 are much more negative than for the benchmark e-Commerce economy. Column 5 of Table 9 relates the *cumulative* welfare gains from e-Commerce to population and market potential using a linear regression. According to the estimated coefficients, by the time the average online expenditure share in China grows to three times its 2013 level, e-Commerce will have reduced the inequality associated with population size by 5.2% and the inequality associated with market potential by 8.6%.[37] Therefore, further development of e-Commerce in China would be able to substantially reduce the inequality in real income between cities, while increasing the average real income.

A caveat is in order. In the counter-factual experiment above, we decrease the online shipping cost proportionally for all city pairs without changing the distance elasticity or the destination-specific components. However, the distance elasticity or the destination-specific components of online shipping cost may in fact change as the e-Commerce technology evolves. These alternative ways of decreasing online shipping cost could lead to different welfare changes.

6 Concluding Remarks

This paper studies the welfare effects of e-Commerce on consumers across cities. Using data from a major e-Commerce platform in China, we first show a robust empirical pattern that residents in smaller and more remote cities have larger online expenditure shares. We then build a general

[37]Specifically, $\frac{-0.0107}{0.207} = -5.2\%$ and $\frac{-0.00365}{0.0425} = -8.6\%$

equilibrium model and quantify the welfare effects of e-Commerce. We find the welfare gains from e-Commerce to be 1.6 percent. Furthermore, the arrival of e-Commerce reduces the elasticity of real income with respect to population by 1.9% and the elasticity with respect to market potential by 4.1%. With its rapid growth, e-Commerce promises to further reduce this inequality in the coming years.

We made a few simplifying assumptions in the paper. In particular, limited by data, 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. Incorporating this within-city dimension is an important direction for future research.

References

Anderson, James E. and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, September 2004, *42* (3), 691–751.

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.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos Magdaleno, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," 2016.

Atkin, David, 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.

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

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.

Chen, Liang, Garrett A Johnson, and Yao Luo, “Great and Small Walls of China: Distance & Chinese E-Commerce,” *Available at SSRN 2668023*, 2015.

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.

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.

Donaldson, Dave, “Railroads of the Raj: Estimating the impact of transportation infrastructure,” *National Bureau of Economic Research Working Paper*, 2010.

_ and Richard Hornbeck, “Railroads and American Economic Growth: A “Market Access” Approach,” *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

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 integration, market size, and industrialization: evidence from China’s National Trunk Highway System,” *The Review of Economic Studies*, 2014, p. rdu010.

_ , “Trade Liberalization, the Price of Quality, and Inequality: Evidence from Mexican Store Prices,” *Unpublished Manuscript*, 2014.

Feenstra, Robert, Mingzhi Xu, Alexis Antoniades, and John Romalis, “What is the Price of Tea in China? Towards the Relative Cost of Living in Chinese and U.S. Cities,” *Working Paper*, 2016, *55.*

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.

Harris, Chauncy D., "The, Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 1954, *44* (4), 315–348.

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.

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

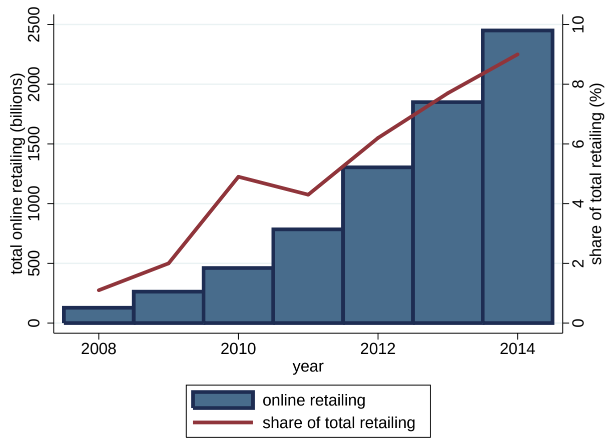
Ramondo, Natalia, Andres Rodriguez-Clare, and Milagro Saborio-Rodriguez, "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, October 2016, *106* (10), 3159–84.

Redding, Stephen J and Matthew A Turner, "Transportation costs and the spatial organization of economic activity," Technical Report, National Bureau of Economic Research 2014.

Tintelnot, Felix, "Global production with export platforms," Technical Report, National Bureau of Economic Research 2016.

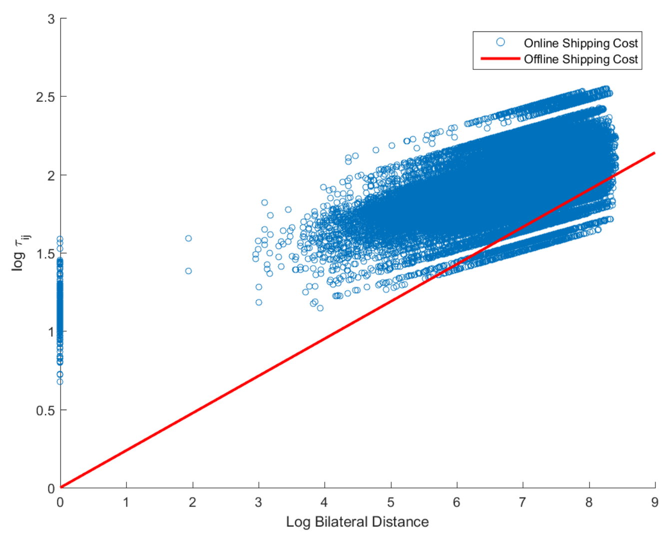
Tran, Allen, "The Aggregate Impact Of Online Retail," *Unpublished Manuscript*, 2014.

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 online retail sales as a share of total retail sales. Source: CNNIC (2015).

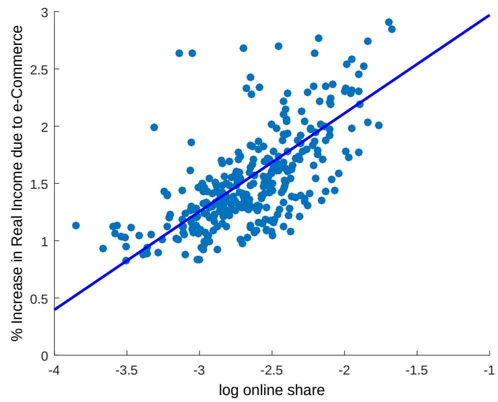
Figure 2: Estimated Offline and Online Shipping Costs



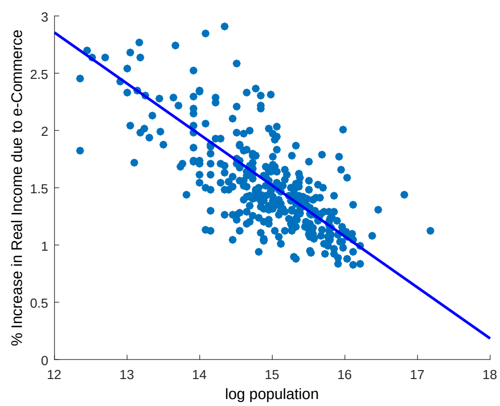
Notes: estimated values by the authors.

Figure 3: Market Size, Online Expenditure Share, and Welfare Gains

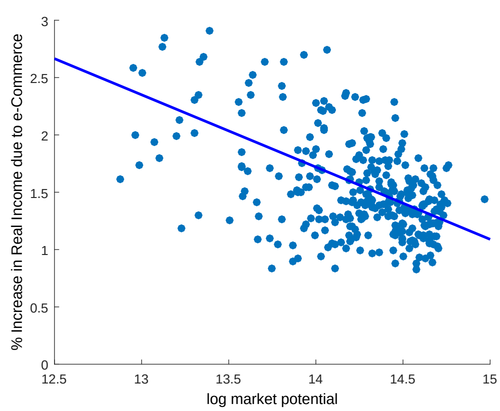
(a) Online Share and Welfare Gains



(b) Population and Welfare Gains



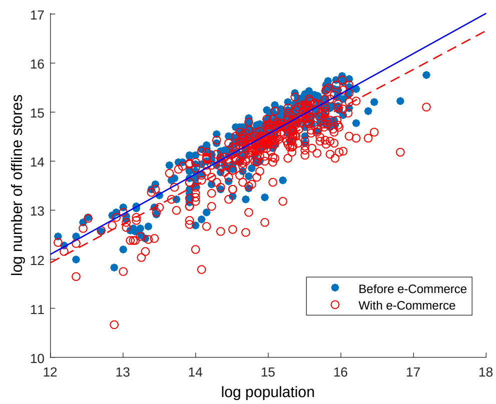
(c) Market Potential and Welfare Gains



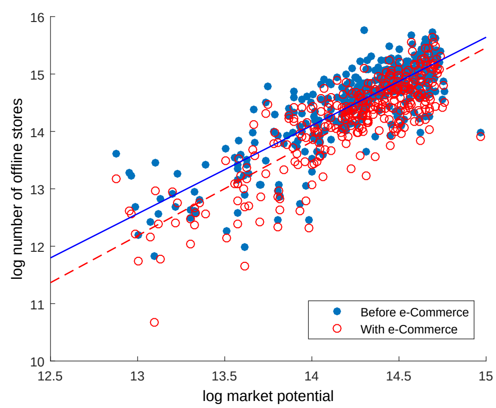
Notes: calculation based on model simulations by the authors.

Figure 4: The Effects of e-Commerce on Offline Stores

(a) By Population



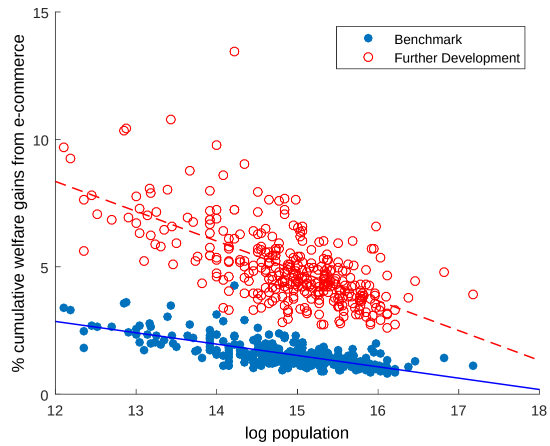
(b) By Market Potential



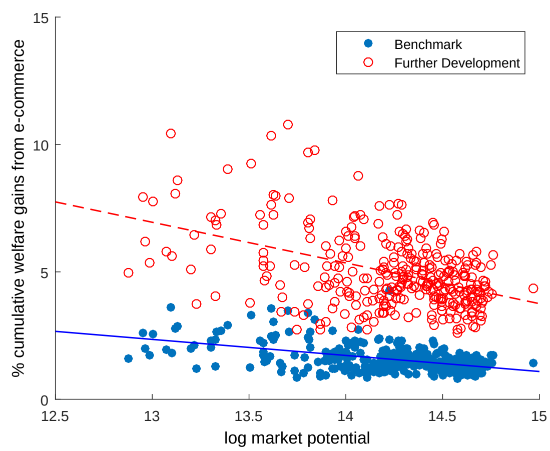
Notes: calculation based on model simulations by the authors.

Figure 5: Cumulative Welfare Gains with Further Development of e-Commerce

(a) By Population



(b) By Market Potential



Notes: calculation based on model simulations by the authors.

Table 1: Summary Statistics

	N	mean	st. dev.
Online share and market size			
online sales as share of total retail sales	315	.079	.034
population (million)	315	3.902	2.958
market potential	315	1.678	.551
Demand and supply factors			
average income (thousand *yuan*)	315	14.992	5.936
average years of schooling	315	8.802	.963
urban share	315	.48	.203
share of working age residents	315	.741	.043
sex ratio	315	1.158	.094
log distance to road	315	6.512	3.774
log distance to railroad	315	6.414	3.897
share of household with access to Internet	315	.593	.632
log online market potential	315	.643	.549
provincial capital = 1	315	.083	.276
Alternative measures			
online sales as a share of consumption	305	.088	.041
online sales as a share of non-hospitality retail sales	312	.083	.047
total income (billion *yuan*)	315	63.742	71.177
total consumption (billion *yuan*)	305	45.113	55.655
log market access	315	35.838	4.166
log market potential (to GDP)	315	31.863	.427

Notes: Each observation is a city.

Table 2: Online Expenditure Share and Market Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dep var:	online sales as share of total retail sales						
ln population	-0.085**	-0.093***	-0.086***	-0.101***	-0.120***	-0.152***	-0.142***
	(0.033)	(0.033)	(0.032)	(0.033)	(0.034)	(0.033)	(0.035)
ln market potential	-0.229***	-0.254***	-0.259***	-0.691***	-0.195***	-0.616***	-0.989***
	(0.058)	(0.058)	(0.059)	(0.092)	(0.058)	(0.095)	(0.226)
ln distance to highway		-0.021***				-0.008	-0.006
		(0.006)				(0.006)	(0.005)
ln distance to rail		0.026***				0.021***	0.009*
		(0.006)				(0.006)	(0.005)
ln broadband penetration rate			0.077***			0.013	0.083*
			(0.029)			(0.028)	(0.045)
ln online market potential				0.416***		0.387***	0.630***
				(0.069)		(0.072)	(0.170)
= 1 if provincial capital					0.269***	0.328***	0.175**
					(0.075)	(0.081)	(0.072)
constant	3.480***	4.107***	4.210***	13.252***	3.277***	12.339***	20.968***
	(1.054)	(1.058)	(1.077)	(1.895)	(1.058)	(1.928)	(4.802)
province FE							X
N	315	315	315	315	315	315	315
R^2	0.124	0.180	0.143	0.218	0.155	0.292	0.637

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Adding Demand Shifters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dep var:	online sales as share of total retail sales						
ln population	-0.141***	-0.139***	-0.134***	-0.144***	-0.116***	-0.116***	-0.527
	(0.035)	(0.036)	(0.034)	(0.036)	(0.037)	(0.037)	(0.571)
ln market potential	-0.996***	-0.986***	-0.955***	-0.943***	-0.907***	-1.517	-0.910***
	(0.238)	(0.225)	(0.228)	(0.241)	(0.248)	(7.633)	(0.248)
ln per capita income	-0.016				-0.212	-0.212	-0.228*
	(0.099)				(0.137)	(0.136)	(0.138)
ln average years of schooling		-0.138			-0.556	-0.554	-0.518
		(0.300)			(0.421)	(0.417)	(0.435)
ln urban rate			0.124*		0.256***	0.257***	0.251***
			(0.068)		(0.087)	(0.085)	(0.088)
share of working age				0.798	1.473	1.469	1.479
				(0.834)	(0.919)	(0.913)	(0.916)
sex ratio				0.025	0.087	0.083	0.066
				(0.272)	(0.262)	(0.265)	(0.263)
ln market potential squared							0.014
							(0.020)
ln population squared						0.015	
						(0.182)	
province FE	X	X	X	X	X	X	X
N	315	315	315	315	315	315	315
R^2	0.637	0.637	0.642	0.639	0.653	0.654	0.653

Note: The dependent variable is the log share of online sales over aggregate retail sales. Supply-side controls are the same as in Column 7 of Table 2. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Alternative Measures

Panel A: alter dep var	(1)	(2)	(3)					
dep var: online sales share of:	retail sales	non-hsptl sales	consump					
ln population	-0.116***	-0.119***	-0.099***					
	(0.037)	(0.038)	(0.031)					
ln market potential	-0.907***	-0.943***	-0.885***					
	(0.248)	(0.269)	(0.218)					
N	315	312	305					
R^2	0.653	0.649	0.769					
Panel B: alter indep var	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
dep var: online sales share of:	retail sales	retail sales	retail sales	retail sales	retail sales	retail sales	retail sales	retail sales
Alt. measures of market size								
ln population	-0.127***	-0.114***						
	(0.040)	(0.036)						
ln total income			-0.116***	-0.127***	-0.114***			
			(0.037)	(0.040)	(0.036)			
ln total consumption						-0.090***	-0.103***	-0.090***
						(0.034)	(0.037)	(0.034)
Alt. measures of remoteness								
ln market potential			-0.907***			-0.458***		
			(0.248)			(0.159)		
ln market access	-0.012**			-0.012**			-0.016***	
	(0.006)			(0.006)			(0.005)	
ln market potential (GDP)		-1.277***			-1.277***			-1.316***
		(0.342)			(0.342)			(0.350)
N	315.	315	315	315	315	305	305	305
R^2	0.629	0.656	0.653	0.629	0.656	0.656	0.634	0.659

Note: Additional covariates include a full set of province fixed effects and supply shifters and demand shifters as in Column 5 of Table 3. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: City by Category

	(1) food	(2) cloth	(3) residence	(4) hhd app	(5) health	(6) trans	(7) recreation
ln population	-0.0771	-0.195***	-0.222	-0.106	-0.131	-0.0189	-0.0183
	(0.0673)	(0.0596)	(0.217)	(0.0799)	(0.0930)	(0.0677)	(0.0830)
ln market potential	0.495	-1.171**	-1.911	0.567	-0.0929	-0.0557	-0.195
	(0.510)	(0.521)	(2.018)	(0.569)	(0.638)	(0.713)	(0.605)
supply shifters	X	X	X	X	X	X	X
demand shifters	X	X	X	X	X	X	X
province FE	X	X	X	X	X	X	X
N	105	105	105	93	84	84	84
R^2	0.823	0.794	0.517	0.812	0.705	0.761	0.786
% in category consump	0.88	45.91	0.003	31.94	4.04	4.24	6.15
% in ttl Taobao Sales	4.980	43.97	0.0100	28.90	5.670	7.940	8.530

Note: The dependent variable is ln share of online sales over consumption by NBS category. Supply and demand shifters are the same as Table 3, Column 5. The 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. For each category, we report the percent of Taobao sales in total consumption for this category (second to bottom row), and the Taobao sales in this category as a share of total Taobao sales (bottom row). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Estimation of Shipping Cost Parameters

	Estimate	Standard Error
Offline Shipping Cost:		
$\ln \tau_{ij}^{P} = \delta \ln D_{ij}$		
δ	0.2378	0.0039
Online Shipping Cost:		
$\ln \tau_{ij}^{E} = \gamma_0 + \gamma_1 \ln D_{ij} + \epsilon_j$		
γ_0	1.1432	0.0846
γ_1	0.1194	0.0124

Note: Estimation based on the full-fledged model by the authors.

Table 7: Spatial Consumption Inequality Before e-Commerce

Panel A	By Quintile of Population				
Quintile	1st	2nd	3rd	4th	5th
Online Share	0.00%	0.00%	0.00%	0.00%	0.00%
Import Ratio	33.81%	30.76%	29.02%	27.25%	21.29%
Online	0.00%	0.00%	0.00%	0.00%	0.00%
Offline	33.81%	30.76%	29.02%	27.25%	21.29%
Real Income	1.000	1.121	1.138	1.284	1.774
Panel B	By Quintile of Market Potential				
Quintile	1st	2nd	3rd	4th	5th
Online Share	0.00%	0.00%	0.00%	0.00%	0.00%
Import Ratio	23.24%	27.72%	26.63%	29.85%	34.69%
Online	0.00%	0.00%	0.00%	0.00%	0.00%
Offline	23.24%	27.72%	26.63%	29.85%	34.69%
Real Income	1.000	1.088	1.374	1.429	1.335

Note: calculation based on model simulations. “Online share” refers to expenditure spent through the e-Commerce channel as a share of total expenditure in the tradeable sector. “Import Ratio” refers to expenditure on goods produced in other cities as a share of total expenditure in the tradeable sector. The next two lines break the “import ratio” into the online and offline channels. Both the online share and the import ratio are calculated based on the tradeable sector only.

Table 8: The Effect of e-Commerce on Inter-City Trade and Welfare

Panel A	By Quintile of Population				
Quintile	1st	2nd	3rd	4th	5th
Online Share	9.94%	7.95%	7.44%	6.92%	7.10%
Change in Import Ratio	6.11%	4.59%	4.31%	3.87%	3.84%
Online	9.88%	7.88%	7.35%	6.82%	6.88%
Offline	-3.77%	-3.28%	-3.04%	-2.95%	-3.03%
Change in Real Income	2.16%	1.57%	1.47%	1.32%	1.21%
Panel B	By Quintile of Market Potential				
Quintile	1st	2nd	3rd	4th	5th
Online Share	9.39%	7.69%	8.97%	7.14%	6.16%
Change in Import Ratio	6.42%	4.62%	4.73%	3.80%	3.15%
Online	9.31%	7.62%	8.80%	7.00%	6.08%
Offline	-2.89%	-3.00%	-4.07%	-3.20%	-2.92%
Change in Real Income	1.91%	1.51%	1.54%	1.37%	1.26%

Note: calculation based on model simulations. “Online share” refers to expenditure spent through the e-Commerce channel as a share of total expenditure in the tradeable sector. “Import Ratio” refers to expenditure on goods produced in other cities as a share of total expenditure in the tradeable sector. The next two lines break the “import ratio” into the online and offline channels. Both the online share and the import ratio are calculated based on the tradeable sector only. Changes in import ratios and real income are calculated relative to the economy without e-Commerce. Changes in import ratios are reported in terms of percentage point differences while changes in real income are reported in terms of percentage differences.

Table 9: Regression Analysis using Welfare Gains from e-Commerce

	(1)	(2)	(3)	(4)	(5)
	Log Real Income		Benchmark Welfare Gains		Cum. Gains with More E
	No E	E			
ln population	0.207***	0.203***	-0.00399***	-0.0290***	-0.0107***
	(0.0359)	(0.0358)	(0.000330)	(0.00854)	(0.000996)
ln market potential	0.0425	0.0408	-0.00172***	-0.0275***	-0.00365*
	(0.0669)	(0.0668)	(0.000633)	(0.00895)	(0.00200)
(ln population)				0.00178***	
×(ln market potential)				(0.000604)	
Constant	-1.275	-1.175	0.0995***	0.463***	0.261***
	(0.871)	(0.870)	(0.00798)	(0.126)	(0.0245)
N	325	325	325	325	325
r2	0.178	0.172	0.533	0.546	0.424

Note: dependent variables are calculated from model simulations by the authors. Columns 1 and 2 use log real income from the economies without e-Commerce and with e-Commerce as dependent variable, respectively; Columns 3 and 4 use welfare gains (difference in log real income) from e-Commerce; Column 5 uses cumulative welfare gains under further development of e-Commerce. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix I Category Crosswalk

Table A1: Taobao Category and NBS Category Crosswalk

category name	%	category name	%
1. Food	0.3	**4. Hhd Articles, Appliances and Services**	
snacks, nuts and specialty food	0.0	jewelry	0.6
dining, catering and takeouts	0.1	kitchen appliances	1.3
staple food, dry food, oil and seasoning	0.7	MP3s, iPods and voice recorders	0.2
pet food	0.0	consumer electronics	0.1
perishable food and prepared food	0.2	home appliances	2.7
tea, alcoholic drinks and beverages	0.3	home furniture	3.0
formula/baby food		desktop computers and servers	0.5
2. Clothing and Accessories	1.1	cameras	1.1
fashion accessories	0.3	tablet computers	0.7
lighters, Swiss knives and optical glasses	0.8	laptop computers	1.0
outdoor, hiking and camping gear	1.6	**5. Health Care and Medical Service**	
underwear and pajamas	1.9	beauty products and cosmetics	8.9
children's clothing	1.9	baby care and gear	1.3
bags and luggage	0.5	personal hygiene and care	1.2
sportswear, sports bags and accessories	11.5	health and medical equipment	0.4
women's clothing	2.6	nutrition supplements	1.3
women's shoes	4.3	maternity care	0.6
men's clothing	1.0	**6. Transport and Communications**	
sports shoes	1.2	transportation tickets	0.0
men's shoes	0.5	prepaid cellphone cards	0.2
watches		web equipment	0.2
3. Residence	0.2	car accessories and parts	1.2
basic building materials	0.0	motorcycle parts and accessories	0.1
renovation contractor service	0.0	new and used car parts	0.0
real estate services		cellphones	2.3
4. Hhd Articles, Appliances and Services	0.3	**7. Recreation, Education, and Culture**	
floral and gardening	0.9	credits for computer games	0.0
festive and party supplies, gifts and crafts	0.9	credits and gift cards for websites	0.1
hardware and tools	0.6	e-dictionary, e-reader and office stationary	0.3
cleaning supplies, paper towels and deodorizer	2.8	web services	0.6
consumer electronics parts	0.8	visitor pass, trips and travel services	0.2
cleaning tools, organization and storage	0.9	toys, models and comic books	0.8
office appliances and supplies	0.7	Internet games	0.1
electrical and electronic products and parts	1.1	education and training	0.0
kitchen ware	0.6	books, magazines and newspapers	0.3
antique, stamps, coins, paintings and collectibles	0.5	sports, yoga, and bodybuilding supplies	1.1
home decoration	2.5	leisure and entertainment	0.0
beddings, linens and curtains	0.4	photography service	0.0
office furniture and equipment	2.8	music, movies and TV shows	0.0
home improvement supplies	0.8	musical instruments and parts	0.2
video and audio equipment	1.1	flash drive, digital storage and harddrives	0.2
computer hardware, accessories and monitors		hotels and lodging	0.1

Note: online sale in each Taobao category as a percent share in its corresponding NBS category reported.

Appendix II A Simple Model to Evaluate Welfare Gains from e-Commerce

In this section we develop a simple model based on revealed preference to shed light on the welfare implications of our empirical results in Section 3. In Section 4 we present a quantitative framework with more realistic assumptions.

Suppose consumers in each city i can purchase two goods, one sold online and the other sold in the local offline market.[38] 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

$$U_i = \left((q_i^P)^{\frac{\sigma-1}{\sigma}} + a \cdot (q_i^E)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{14}$$

where q_i^P and q_i^E are the consumption of goods from the local (physical, denoted by superscript "P") market and the online (e-Commerce, denoted by superscript "E") market, respectively; $\sigma > 1$ is the elasticity of substitution between the online and local goods; a governs the strength of taste for the online good. Since we have shown that the correlations between online share and city characteristics are not driven by differential preference for online shopping, we assume the preference parameter a 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. On the other hand, since the welfare formula we derive below is independent of a, the assumption does not affect our measure of gains from e-Commerce.

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

$$\frac{a \cdot (q_i^E)^{-\frac{1}{\sigma}}}{(q_i^P)^{-\frac{1}{\sigma}}} = \frac{p_i^E}{p_i^P}.$$

Using λ_i to denote the share of their expenditures spent on the online good, it is straightforward to show that

$$\frac{\lambda_i}{1-\lambda_i} = a^\sigma \left(\frac{p_i^E}{p_i^P}\right)^{1-\sigma}.$$

If consumers in city i spend a larger share of expenditure on the online good than consumers in city j, it follows that $\frac{p_i^E}{p_i^P} < \frac{p_j^E}{p_j^P}$. That is, the relative price of the offline good in city i is higher. 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 more expensive and thus less attractive in that city.[39]

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

[39] Of course, without a micro foundation on why consumers choose to purchase goods from online or local markets,

To evaluate the gains from e-Commerce, we hold w_i and p_i^P constant and evaluate the welfare of consumers in city i while changing the value of p_i^E. Let P_i be the price index for city i, then $P_i^{1-\sigma} = (p_i^P)^{1-\sigma} + a^\sigma (p_i^E)^{1-\sigma}$, and the welfare of a consumer in city i is

$$u_i = \frac{w_i}{P_i} = \frac{w_i}{((p_i^P)^{1-\sigma} + a^\sigma (p_i^E)^{1-\sigma})^{\frac{1}{1-\sigma}}}.$$

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

$$\text{Gains from e-Commerce} = (1-\lambda_i)^{\frac{1}{1-\sigma}} - 1. \tag{15}$$

The welfare gain is an increasing function of the online expenditure share, λ_i. 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.

Appendix III Derivation of e-Commerce Demand Elasticity with Respect to Offline Price

Let $Q_{ij}^{TC,E}(\phi)$ denote the e-Commerce sales quantity of a two-channel firm. Then

$$\begin{aligned} Q_{ij}^{TC,E}(\phi) =& \frac{s_{ij}^{PE}(\phi)}{p_{ij}(\phi)} \cdot \rho \\ =& \frac{\beta_j^T Y_j \kappa^{\frac{-\sigma}{1-\sigma}} w_i^{-\sigma}}{(P_j^T)^{1-\sigma}} \Big(\sum_{m \in \{P,E\}} (\tau_{ij}^m)^{-\theta} \Big)^{\frac{\sigma}{\theta}} \phi^\sigma \cdot \frac{(\tau_{ij}^E)^{-\theta}}{\sum_m (\tau_{ij}^m)^{-\theta}}. \end{aligned} \tag{16}$$

Consider the case in which e-Commerce is less developed so that τ_{ij}^P is small compared to τ_{ij}^E such that $\sum_{m\in\{P,E\}} (\tau_{ij}^m)^{-\theta} \approx (\tau_{ij}^P)^{-\theta}$. Then

$$\begin{aligned} Q_{ij}^{TC,E}(\phi) \approx& \frac{\beta_j^T Y_j \kappa^{\frac{-\sigma}{1-\sigma}} w_i^{-\sigma}}{(P_j^T)^{1-\sigma}} \big((\tau_{ij}^P)^{-\sigma} \big) \phi^\sigma \cdot \frac{(\tau_{ij}^E)^{-\theta}}{(\tau_{ij}^P)^{-\theta}} \\ \approx& \frac{\beta_j^T Y_j \kappa^{\frac{-\sigma}{1-\sigma}} w_i^{-\sigma} \phi^\sigma}{(P_j^T)^{1-\sigma}} \cdot (\tau_{ij}^P)^{\theta-\sigma} (\tau_{ij}^E)^{-\theta}. \end{aligned} \tag{17}$$

this simple model cannot be used to evaluate what the difference in the relative price 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 good 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.

Finally, we can take the log on both sides and show that

$$\frac{\partial \ln Q_{ij}^{TC,E}(\phi)}{\partial \ln \tau_{ij}^{P}} = \theta - \sigma. \tag{18}$$

Appendix IV The Determinants of Destination-Specific Online Shipping Cost ϵ_j

Table A2: Destination-Specific Online Shipping Cost ϵ_j and Population

	(1)	(2)	(3)	(4)	(5)
ln pop 2010	-0.0950***	-0.0850***	-0.0851***	-0.0802***	-0.0940***
	(0.00884)	(0.00906)	(0.00833)	(0.00842)	(0.0103)
ln broadband penetration rate		-0.0307***	-0.01000	-0.00884	-0.0318**
		(0.00872)	(0.00905)	(0.00904)	(0.0137)
ln urban rate			-0.0881***	-0.0728***	-0.0708**
			(0.0233)	(0.0244)	(0.0283)
=1 if provincial capital				-0.0569**	-0.00524
				(0.0232)	(0.0214)
Constant	2.555***	2.378***	2.328***	2.271***	2.243***
	(0.132)	(0.136)	(0.127)	(0.126)	(0.160)
Province FE					X
N	325	314	314	314	314
r2	0.354	0.341	0.382	0.394	0.719

Note: The dependent variable is the calibrated destination-specific component ϵ_j from our specification of online shipping cost $\ln \tau_{ij}^{E} = \gamma_0 + \gamma_1 \ln D_{ij} + \epsilon_j$. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix V Exercises with Alternative Specifications of Online Shipping Costs

We take the estimated online shipping cost ($\ln \tau_{ij}^{E} = \gamma_0 + \gamma_1 \ln D_{ij} + \epsilon_j$) in Table 6 as our benchmark and conduct two different exercises with alternative specifications of online shipping costs. In each exercise, we modify the specification of online shipping cost and then compare the modified e-Commerce economy to the corresponding economy without e-Commerce. The results of these two exercises are summarized in Tables A3.

In Exercise 1, we take the estimate of distance elasticity $\gamma_1 = 0.1194$ from Table 6 and set the destination-specific component $\epsilon_j = 0$ for all cities. We then pick γ_0 to match the overall online expenditure share of the economy. As Panel A of Table A3 shows, online expenditure share continues to decrease with city size, from an average of 20.16% for the 1st quintile to 6.35% for the 5th quintile. As Figure A1a shows, the negative slope between log online expenditure share and log population is -0.481, which is much more negative than for the counterpart value for the benchmark calibration, which is -0.131. To obtain the welfare gains from e-Commerce for this economy, we then increase the online shipping cost to infinity for all bilateral city pairs. Figure A2a plots the welfare gains from e-Commerce under Exercise 1, against log population. Not surprisingly, the larger differences in online expenditure share across population size translates into greater welfare gains from e-Commerce. The real income for cities in the smallest quintile increases by 3.15% due to e-Commerce, which is more than three times higher than the average of 0.96% for the largest cities. Therefore, e-Commerce has a larger impact on spatial income inequality associated with population under Exercise 1 than in the benchmark case. If we interpret the patterns of destination-specific component of online shipping cost as small cities lacking infrastructure, then this experiment suggests an important step to further reduce consumption inequality is to enhance the infrastructure for online shipping in small cities. Figure A1b plots log online share from Exercise 1 against log market potential. We find that the slope between log online expenditure share and log market potential under Exercise 1 is comparable to that for the benchmark calibration. Therefore, the destination-specific components of online shipping cost are not as crucial in matching the negative correlation between online share and market potential.

In Exercise 2, we set $\epsilon_j = 0$ and $\gamma_1 = \delta$, and adjust γ_0 to match the overall online expenditure share. This experiment effectively sets the online shipping cost between two cities to be proportional to the offline shipping cost. As Figure A1c shows, online expenditure share remains negatively correlated with population. The population elasticity of online share is -0.185, which is close to the value of -0.131 in the data. Therefore, our model is able to generate the negative relationship between online share and population without appealing to a different distance elasticity and destination-specific component. However, the online shipping cost under Exercise 2 would be inconsistent with inter-provincial e-Commerce flows. Furthermore, without the destination-specific component, Exercise 2 does not match the online expenditure share by city. As a result, the negative relationship between online share and population is much tighter than in the data. Lastly, as Figure A1d shows, Exercise 2 is unable to produce the negative relationship between log online share and log market potential. As before, we then increase the online shipping cost to infinity for all bilateral city pairs to obtain the welfare gains from e-Commerce for this economy. The welfare gains from e-Commerce are presented in Figures A2c and A2d and follow the same patterns of log online share.

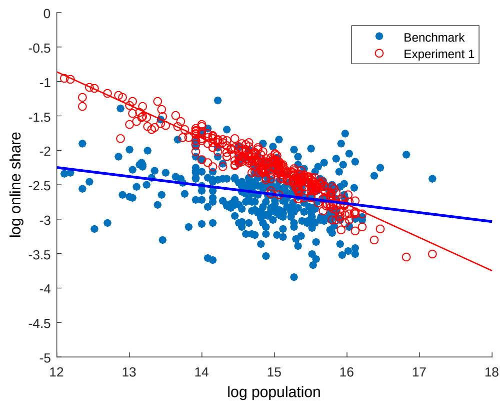
Table A3: The Effect of E-Commerce Under Alternative Shipping Costs

Exercise 1: No Destination-Specific Component					
Panel A	by Quintile of Population				
Quintile	1st	2nd	3rd	4th	5th
Online Share	20.12%	12.61%	10.64%	8.62%	6.33%
Change in Import Ratio	9.37%	5.89%	5.16%	4.21%	3.30%
Online	20.01%	12.49%	10.52%	8.50%	6.19%
Offline	-10.64%	-6.60%	-5.36%	-4.29%	-2.89%
Change in Real Income	3.09%	1.98%	1.67%	1.38%	0.96%
Panel B	by Quintile of Market Potential				
Quintile	1st	2nd	3rd	4th	5th
Online Share	16.84%	12.61%	9.75%	9.56%	9.56%
Change in Import Ratio	9.04%	6.12%	4.70%	4.23%	3.85%
Online	16.71%	12.49%	9.62%	9.44%	9.45%
Offline	-7.68%	-6.37%	-4.92%	-5.21%	-5.60%
Change in Real Income	2.56%	1.96%	1.42%	1.37%	1.46%
Exercise 2: Proportional to Offline Shipping Cost					
Panel C	by Quintile of Population				
Quintile	1st	2nd	3rd	4th	5th
Online Share	11.52%	9.32%	8.66%	8.10%	7.45%
Change in Import Ratio	1.94%	1.10%	0.93%	0.72%	0.54%
Online	7.58%	5.15%	4.36%	3.69%	2.66%
Offline	-5.64%	-4.05%	-3.43%	-2.97%	-2.12%
Change in Real Income	1.54%	1.30%	1.23%	1.17%	1.05%
Panel D	by Quintile of Market Potential				
Quintile	1st	2nd	3rd	4th	5th
Online Share	10.15%	9.24%	8.47%	8.54%	8.64%
Change in Import Ratio	1.75%	1.18%	0.85%	0.77%	0.67%
Online	5.55%	4.89%	4.03%	4.29%	4.68%
Offline	-3.80%	-3.71%	-3.18%	-3.52%	-4.01%
Change in Real Income	1.35%	1.29%	1.16%	1.17%	1.22%

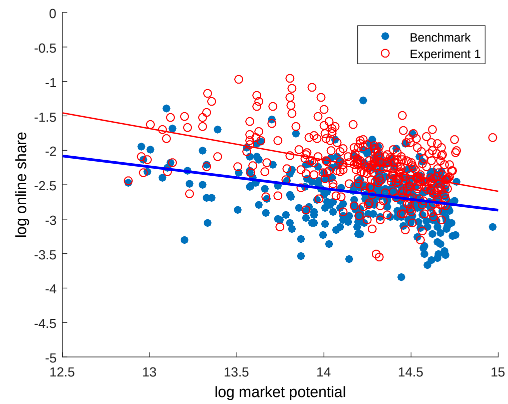
Note: calculation based on model simulations. In Exercise 1, we take the estimated online shipping cost in Table 6 as our benchmark and then set the destination-specific component ϵ_j to 0 and adjust γ_0 to match the overall online expenditure share. In Exercise 2, we set the online shipping cost to be proportional to the offline shipping cost and match the overall online expenditure share. Changes in import ratios and real income are calculated relative to the economy without e-Commerce. Changes in import ratios are reported in terms of percentage point differences while changes in real income are reported in terms of percentage differences.

Figure A1: Online Share with Alternative Online Shipping Costs

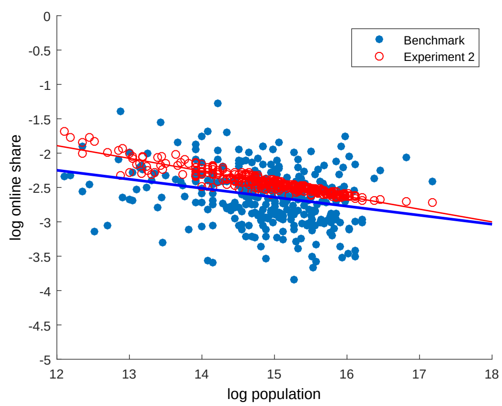
(a) Exercise 1, by Population



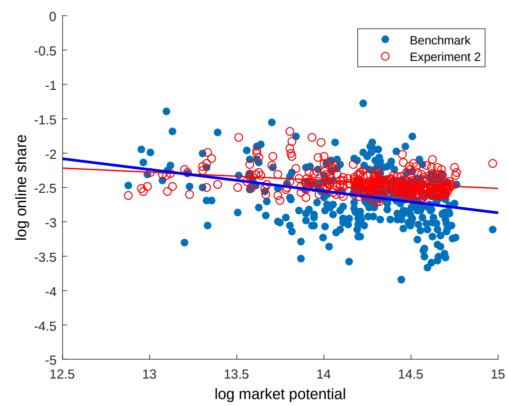
(b) Exercise 1, by Market Potential



(c) Exercise 2, by Population



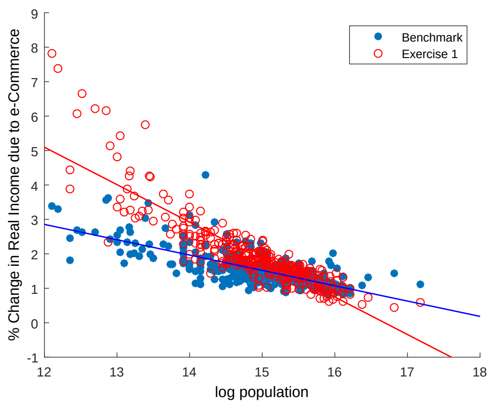
(d) Exercise 2, by Market Potential



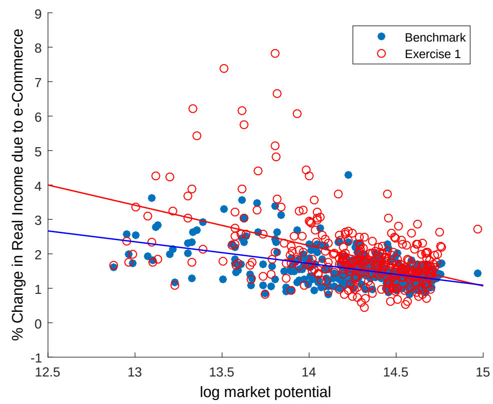
Notes: calculation based on model simulations by the authors. We take the estimated online shipping cost in Table 6 as our benchmark and conduct two different experiments. We then compare the economy under each experiment to the economy without e-Commerce. In Exercise 1, we set the destination-specific component ϵ_j to 0 and adjust γ_0 to match the overall online expenditure share. In Exercise 2, we set $\epsilon_j = 0$ and $\gamma_1 = \delta$, and adjust γ_0 to match the overall online expenditure share.

Figure A2: Gains from e-Commerce with Alternative Online Shipping Costs

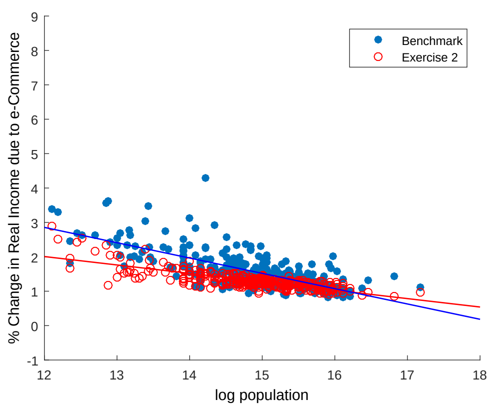
(a) Exercise 1, by Population



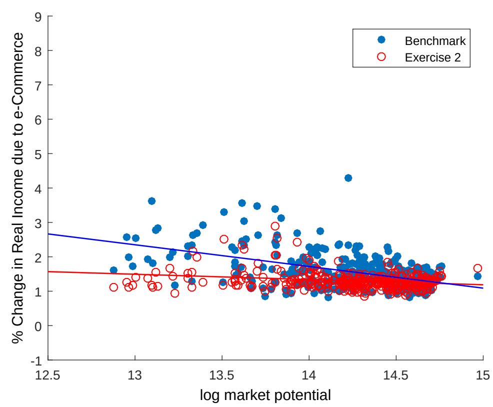
(b) Exercise 1, by Market Potential



(c) Exercise 2, by Population



(d) Exercise 2, by Market Potential



Notes: calculation based on model simulations by the authors. We take the estimated online shipping cost in Table 6 as our benchmark and conduct two different experiments. We then compare the economy under each experiment to the economy without e-Commerce. In Exercise 1, we set the destination-specific component ϵ_j to 0 and adjust γ_0 to match the overall online expenditure share. In Exercise 2, we set $\epsilon_j = 0$ and $\gamma_1 = \delta$, and adjust γ_0 to match the overall online expenditure share.

Table A4: Cumulative Effects with Further Development of E-Commerce

Panel A	By Quintile of Population				
Quintile	1st	2nd	3rd	4th	5th
Online Share	28.63%	24.83%	23.72%	22.46%	21.97%
Change in Import Ratio	17.71%	14.48%	13.85%	12.85%	12.73%
Online	28.41%	24.53%	23.36%	21.99%	20.96%
Offline	-10.70%	-10.05%	-9.51%	-9.14%	-8.22%
Change in Real Income	6.60%	4.99%	4.73%	4.37%	4.07%
Panel B	By Quintile of Market Potential				
Quintile	1st	2nd	3rd	4th	5th
Online Share	26.95%	24.00%	27.07%	22.88%	20.71%
Change in Import Ratio	18.58%	14.47%	15.31%	12.56%	10.70%
Online	26.60%	23.68%	26.33%	22.30%	20.34%
Offline	-8.02%	-9.21%	-11.01%	-9.74%	-9.64%
Change in Real Income	5.83%	4.75%	5.08%	4.52%	4.14%

Note: calculation based on model simulations."Online share" refers to expenditure spent through the e-Commerce channel as a share of total expenditure in the tradeable sector. "Import Ratio" refers to expenditure on goods produced in other cities, as a share of total expenditure in the tradeable sector. The next two lines break the "import ratio" into the online and offline channels. Both the online share and the import ratio are calculated based on the tradeable sector only. Changes in import ratios and real income are calculated relative to the economy without e-Commerce. Changes in import ratios are reported in terms of percentage point differences while changes in real income are reported in terms of percentage differences.