NET Institute*

www.NETinst.org

Working Paper #15-14

September 2015

**Great and Small Walls of China:
Distance & Chinese E-Commerce**

Liang Chen
Wuhan University

Garrett Johnson
Simon Business School, University of Rochester

Yao Luo
University of Toronto

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

Great and Small Walls of China: Distance & Chinese E-Commerce*

Liang Chen[†] Garrett Johnson[‡] Yao Luo[§]

September 30, 2015

Abstract

Alibaba's Taobao is China's dominant C2C e-commerce platform. Relying on novel transaction level data from its electronics category, this paper examines the roles of distance as well as seller and buyer experience in Chinese e-commerce. Our estimate of a gravity model shows that the distance elasticity in our data is greater than US C2C e-commerce but smaller than that of South America. In addition, we document greater home bias in China than either of the US or South American Settings. Moreover, by estimating a 'new-new' trade model using seller-level transaction data, this paper studies the link between seller experience and distance. Our results show that more experienced sellers tend to sell more and export their goods farther into the marketplace.

Keywords: e-commerce, China, gravity, distance, domestic trade, Internet marketing

JEL Classification: F10, F14, L86

*We are grateful for the comments from the conference participants at 2015 China India Insights Conference. We thank the NET Institute (www.NETinst.org) for financial support. Any comments are welcome.

[†]Economics and Management School, Wuhan University, Luojia Hill, Wuhan, China, 430072. Email: liang.chen@whu.edu.cn

[‡]Simon Business School, University of Rochester, 3-206 Carol Simon Hall, Box 270100, Rochester, NY 14627. Email: Garrett.Johnson@Simon.Rochester.edu

[§]Department of Economics, University of Toronto, Max Gluskin House 322, 150 St. George Street, Toronto. Email: yao.luo@utoronto.ca

1 Introduction

The Internet dramatically cuts the cost of distance for economic activity. This leads to controversial claims about the “death of distance” (Cairncross, 2001). E-commerce is an appealing domain to investigate the reduced role of distance because, unlike traditional commerce, distant sellers are only a click away for online consumers. The main idea from the existing literature is that while consumers can purchase from sellers hundreds miles away through online platforms, they still reveal a strong preference for co-located sellers or nearby sellers. See, e.g., Hortagsu et al. (2009) on American interstate commerce on eBay and South American commerce on MercadoLibre.

This paper examines the role of distance in Chinese e-commerce. We are interested in how e-commerce platforms differ in a developing country context since participants have varying levels of experience and trust with e-commerce. For instance, we want to know whether consumers prefer to trade within their own locales and whether a seller’s experience involving online shopping can mitigate the importance of distance. In addition, the scale of China’s e-commerce warrants study.

We use data from Alibaba’s Taobao, China’s dominant consumer-to-consumer (C2C) e-commerce platform. Taobao commands 90% of China’s C2C e-commerce. Consumers can buy almost everything they need online, often times at prices much lower than what traditional retailers offer. Taobao is a top 10 global website that represents over 70% of all China’s combined C2C and B2B e-commerce in 2011. Together with Alibaba’s B2B platform, Taobao processed \$170 billion in transactions in 2012. This sum exceeds the domestic output of two thirds of all nations. We collect all transactions from Taobao’s electronics category from the second half of 2011. For each transaction, we observe detailed information about the seller, the buyer and the product itself. In particular, the data reveal both the locations of the buyer and seller as well as their level of experience with the platform.

To examine the role that distance plays in Chinese e-commerce, we first estimate a gravity model of trade by aggregating our data into province-level trade flows. The

gravity model posits that bilateral trade is proportional to the size of each trade partner’s economy and inversely proportional to the distance between the pair. The gravity model is rooted in the trade literature and remains a fixture there due to its elegance and explanatory power. Here, gravity provides a parsimonious means to model a large marketplace yields an elasticity measure for commerce with respect to distance. Our base gravity model employs the modern estimation approach of Anderson and Van Wincoop (2003). This yields an elasticity estimate for how trade falls with distance between trading pairs.

Our analysis provides the differences of e-commerce markets between the US and China. In province-level analysis, we find that Chinese e-commerce exhibits much less dependence on distance than American inter-state commodity trade (Wolf (2000); Hillberry and Hummels (2003)). However, Chinese e-commerce exhibits greater dependence on distance (elasticity=-0.2) than the study by Hortaçsu et al. (2009) of American eBay commerce (elasticity=-0.1). We also find greater “home bias” or preference for own-province trade than on eBay in the U.S..

Our data contain information on the seller and buyer’s experience on the platform. This information provides an opportunity to push the literature further and examine the role that seller and buyer marketplace experience plays in the online marketplace and how experience interacts with distance. To do so, we use our data to estimate a model based on the ‘new-new’ trade literature initiated by Melitz (2003). Recent trade literature emphasizes the importance of the quality of firms and the goods they produce. For instance, Baldwin and Harrigan (2011) modify Helpman et al. (2008) by allowing consumer utility to include preference for quality. Crozet et al. (2012) examines the role of quality in trade on the firm side using French exports of champagne. We extend the model in Crozet et al. (2012) and then estimate our extended model to examine the roles of buyer and seller experiences. Our estimates shows that seller experience raises the probability of selling to a province and the sales value. We also find that more experienced buyers tend to buy more from more experienced sellers.

The ‘new-new’ trade theory explicitly models bilateral trade on both the extensive

and intensive margin. It then can help to solve the zero trade problem-frequency of international trade in one or both directions-discussed in the trade literature. Helpman et al. (2008) illustrate that the zero problem introduces significant bias in gravity model estimates and they propose a Heckman-style correction. Crozet et al. (2012) propose a Tobit approach to solve the zero trade problem. This zero problem also exists in our e-commerce context. In our data, we observe many sellers that do not sell product to some provinces. This zero problem will bias our estimates on the effect of distance. However, prior research on e-commerce rarely takes this problem into account. This paper corrects the zero problem using the Tobit approach in Crozet et al. (2012). Our comparison between OLS and Tobit estimation results show that the zero selection problem bias the estimates of buyer and seller experiences towards zero. After correcting for selection using the Tobit approach, we obtain larger effects of buyer and seller experience on online shopping.

2 Literature Review

The impact of distance on trade has generated considerable research attention in international trade, but less so for online trade. The international trade literature on distance effect makes use of gravity models, which focus on the role of distance and economy size on trade patterns (see Anderson (2010) and Head et al. (2013) for surveys). Anderson and Van Wincoop (2003) estimate a gravity model to compare the border effects between U.S. states and Canadian provinces. Hillberry and Hummels (2003) estimate a gravity model and find that inter-state home bias in the US is explained by wholesaling activities. Helpman et al. (2008) construct a gravity estimation equation founded on the new-new trade model proposed by Melitz (2003). Baldwin and Harrigan (2011) modify Helpman et al. (2008) by allowing consumer utility to include preference for quality. Crozet et al. (2012) examines the role of quality in trade on the firm side using French exports of champagne. In this paper we extend the trade model in Crozet et al. (2012) by incorporating heterogeneity on both the buyer and

seller sides.

For online trade, Hortaçsu et al. (2009) fits a gravity model to American eBay transactions and finds a significantly positive impact of distance though diminished relative to our study of Chinese e-commerce. Hortaçsu et al. (2009) also looks at online trade on a south American e-commerce platform and finds larger impact of distance than both their eBay estimates and our own. Blum and Goldfarb (2006) look at the consumption of digital goods and find that distance matters even in a setting where physical distance is costless. Both Hortaçsu et al. (2009) and Blum and Goldfarb (2006) find that the impact of distance is mediated by cultural goods which are thought to be inherently more local. Cowgill and Dorobantu (2014b) and Cowgill and Dorobantu (2014a) fit gravity models to international and intranational trade using Google’s AdWords conversions to define trade flows; they find evidence for both distance and border effects. Our estimates on the gravity model are consistent with this literature and confirm the effects of distance and home bias.

Other researchers also examine the role of reputation in e-commerce. Reputable sellers are found to sell more products or services at higher prices. Using data from eBay auctions, Melnik and Alm (2002) and Houser and Wooders (2006) find that seller’s reputation has a significant positive effect on price. Using a panel of eBay seller histories, Cabral and Hortacsu (2010) find that, when a seller first receives negative feedback, his weekly sales rate drops from a positive 5% to a negative 8%; subsequent negative feedback ratings arrive 25% more rapidly than the first one and do not have nearly as much impact as the first one. Using the complete selling history of the sellers on the Chinese equivalent of eBay, Eachnet.com, Cai et al. (2014) find that reputation can facilitate trade through buyer’s experience and word-of-mouth. Using a random sample of all sellers on Taobao between March 2010 and April 2011, Fan et al. (2014) find a substantial reward to reputation for established sellers only. As a result, new sellers tend to lower their prices in order to boost their sales and ratings. Using monthly data on sellers of consumer electronics, Yang et al. (2015) study individual sellers’ use of two marketing tools on a Chinese C2C platform: Paid-Search advertising and Hot-

Shop. Our novel approach instead introduces new new trade theory to analyze the role of seller reputation.

Academics have only recently used Taobao data. Most attention to date has focused on Taobao’s reputation system including Xiao et al. (2013), Ju et al. (2013), Ye et al. (2013), and Fan et al. (2014). Chu and Manchanda (2013) use a structural model to study network effects on Taobao. Wu (2014) compares the Taobao ad network’s list price mechanism to Google Ad Sense’s GSP mechanism.

3 Data

3.1 C2C in China

The Chinese online market has expanded dramatically over the last decade. In 2000, China had only about two million internet users and lacked financial and shipping infrastructure to facilitate e-commerce transactions. eBay entered China through the acquisition of EachNet (one third in 2002 and the other two thirds in 2003), a home-grown platform inspired by eBay and founded in 1999. Since the acquisition, eBay changed the operation to align it with its global one. In 2003, taobao.com was founded. In contrast to its main competitor eBay, Taobao adopted a no-fee model: instead of charging commission fees on transactions, its main revenues are from advertising. Users quickly switch to the free platform. Within 3 years, Taobao surpassed eBay and became the dominant C2C platform ever since.

In contract to the maturity of the US retail market, China still lacked development of “bricks and mortar” retailers when e-commerce was introduced. In particular, it is costly to set up physical stores in lower-tier Chinese cities and rural areas. The immaturity of the Chinese offline retail industry makes e-commerce a more efficient way of reaching consumers in the areas that traditional retailers cannot cover. As a result, the Chinese C2C market is much bigger than B2C in China, accounting for 77% of e-tailing in 2011¹ and features a larger number of small sellers. According

¹Source: McKinsey Global Institute 2013 Report “China’s e-tail revolution: Online shopping as a catalyst

to the 2012 Annual E-Tailing Survey of China Internet Network Information Center (hereafter, the 2012 CNNIC survey), e-tailing represents 6.1% of retail sales in China.² Moreover, 55.7% of Chinese internet users are online shoppers.

It has been well documented that national borders have negative effects on trade between geographically separated sellers and buyers. We now describe a few institutional details that possibly make our focus of interstate trade relevant. First, shipping fees are usually structured in a way that leads to provincial border effects. Due to nonuniform development of shipping infrastructure, there are typically more and better shipping services in more developed areas. For instance, sellers located in Yangtze River delta can normally provide cheap shipping prices to the triangle-shaped area of Shanghai, southern Jiangsu and northern Zhejiang. Sellers may charge different shipping fees depending on whether the transaction is within the same province or not. Thus, distance can still matter if short distances mean lower shipping times and/or prices. Second, shorter distances facilitate search/bargaining as sellers and buyers collocated tend to be share similar culture. (The product itself does not seem to be cultural as we focus on the electronics.) For instance, Taobao provides an instant messenger that the seller and the buyer can discuss details of the transaction before ordering. Buyers also often seek better deals by negotiating over discounts or conditions. Similarity in background or culture between seller and buyer may help reaching on an agreement from the discussion.

3.2 Our Sample

Our data are provided by the largest C2C trading platform in China, Taobao. The dataset contains all transactions in the electronics category from seven months between June and December 2011. For each transaction, we observe information about the seller, the buyer and the product itself. The data contain six sub-categories: Accessories, Digital Cameras, Cell Phones, Desktops, Laptops and Tablets.

for growth.”

²Source: http://finance.ifeng.com/a/20150608/13761755_0.shtml

The data serve to study the role of distance and experience in the Chinese e-commerce context. First, online shoppers frequently purchase from the electronics category. In 2014, 37.5% online shoppers bought electronics, which is only second to 75.3% for “Clothing, Footwear and Headgear”. While the touch and feel of clothes in person is still very important, electronics are a lot easier to compare online. This is particularly so for expensive products in our category. Table 1 provides summary statistics by the six product sub-categories. The average price in our sample is 1772 RMB, which is 74% of urban residents’ disposable income per capita per month. Of particular help is also abundant consumer reviews on the internet. Second, as many listed items are unbranded products, trust is important.

Sellers on Taobao list items for sale with a posted price and without a time limit. In contrast, eBay’s sellers can also list items for auction with a fixed time-frame. Compared to auctions, the posted price format is broadly used in online and offline retailing. Consumers may be more at ease with the use of posted prices, especially in a developing country context. Moreover, posted prices are not confounded by the intricacies of auction designs and strategic bidding behaviors.

We observe a host of seller’s characteristics including their demographics, services offered and promotional activity. The basic seller characteristics include their reputation score, age, city, and gender. We have indicators for whether the seller offers various services including: seven-day free returns, expedited shipping, bundle discounts, payment upon delivery, a ten-fold penalty on the value of products found to be fake, 30-day free repair, and a guarantee that the products match their description. We also know if the seller is engaging in various promotional activities: paying to promote their listings, having a premium store, and offering a social referral promotion. On the buyer’s end, we observe fewer characteristics: gender, age, city, and reputation score.

Table 1 shows transaction-level averages of the value of sales, shipping fees as well as seller and buyer reputation and the number of transactions for each of the six product sub-categories after dropping the transactions with missing information on the loca-

Table 1: Summary Descriptions by Products

	Accessory	Camera	Cellphone	Desktop	Laptop	Tablet	Total
Sales Value (RMB)	172.7 (2325.8)	1722.3 (5781.9)	1940.1 (4770.3)	1991.6 (3231.7)	3363.2 (4713.9)	2557.0 (5869.8)	1771.6 (4772.2)
Shipping Fee (RMB)	6.517 (310.5)	7.345 (209.9)	5.388 (163.8)	28.22 (192.7)	23.19 (131.3)	11.47 (386.7)	8.414 (218.6)
Seller Reputation	5.677 (2.257)	5.108 (2.422)	5.183 (2.365)	3.991 (1.966)	3.766 (2.108)	5.107 (2.618)	5.094 (2.393)
Buyer Reputation	4.590 (1.428)	3.491 (1.735)	3.232 (1.708)	3.559 (1.710)	3.407 (1.747)	3.529 (1.768)	3.535 (1.744)
# Transactions	1,481,226	1,628,824	4,626,502	303,507	733,928	553,198	9,327,185

Table 2: Zero Sales by Categories (The numbers are all in thousands.)

	Accessory	Camera	Cellphone	Desktop	Laptop	Tablet	Total
# Sellers	59.2	299.7	589.3	100.4	24.6	180.9	165.9
# Possible Pairs	1835.5	929.1	1826.7	311.3	762.6	560.9	5142.7
# Positive Sales	232.4	111.0	277.6	34.61	76.9	608.4	726.2
% Positive Pairs	12.66%	11.95%	15.20%	11.12%	10.08%	10.85%	14.12%

tions and reputations of buyers or sellers.³ The transaction data are dominated by sales of cellphones (4.5 million transactions), cameras (1.6 million transactions), and accessories (1.5 million transactions). Laptops is the most expensive category with average transaction cost of 3363 Chinese yuan, and the least expensive category is accessories with average cost of 172 Chinese yuan. Table 1 also contains the average seller and buyer reputation by sub-category. We see that sellers and buyers of accessories have the highest average reputation perhaps because experience is more cheaply gained in inexpensive products.

The reputation score plays a central role in the marketplace. The 2014 CNNIC survey shows that the majority of online shoppers rank reputation as the most important factor for their purchase decisions. Taobao uses a similar rating system to eBay.

³We collect all transactions for electronics, which is comprised of more than 13 millions transactions.

For each transaction, the buyer can choose to rate the seller by leaving feedback. The seller earns +1 point for each positive rating, no points for each neutral rating, and -1 point for each negative rating. If the buyer does not submit feedback in 15 days upon completion of the transaction, the seller obtains a positive rating. The reputation score for buyers is accumulated in a similar system.

Thus, the reputation can be used as a good measure for online market experience of both sellers and buyers. The cumulative rating score is then categorized into twenty grades based on threshold values for cumulative successful sales. Taobao divide reputation scores into four categories and each is characterized by a system of symbols: 0-5 (hearts), 6-10 (diamonds), 11-15 (crowns), 16-20 (golden crowns). Our data shows that 94.7% of all the sellers have a reputation grade below 10, which means the sellers each accumulated less than 10,000 successful sales.⁴ Sellers symbolized by hearts and diamonds are usually new startups and small business sellers. On the other hand, sellers symbolized by crowns and golden crowns are established sellers that can be run by companies.

Figure 1 shows the total sales of each province, while Figure 2 shows the total purchase of each province. The blue numbers are the total sales/purchases (in billions), to which the size of the red circle is proportional. Guangdong, Zhejiang, Jiangsu, Beijing and Shanghai are both the top sellers and the top consumers. Accounting for 30.38% of the electronics industry’s revenues in 2011, Guangdong is by far the largest seller on taobao.com.

Table 2 shows number of sellers, number of possible pair of transactions in seller level, and number of pairs with positive sales. These number are calculated from the transactions after dropping the transactions with missing information on buyer or seller characteristics. The number of possibles equals to number of sellers $\times$ number

⁴Fan et al. (2014) use a 25% random sample of all sellers on Taobao between March 2010 and April 2011. In their sample, 97.62% of all the sellers have a reputation grade below 10. This seems to be a natural result as the electronics category consists of more valuable items relative to the whole sample. Therefore, there are less transactions, rendering it harder to accumulate reputation scores (The average price of a transaction is 1772 RMB in our sample, while it is 141 RMB for new sellers and 118 RMB for established sellers in Fan et al. (2014)’s sample.)

of provinces. This is a high proportion of pairs that the sales of them are zeros for each category. The proportion of positive pairs is ranged from 10.08% to 15.20%. The proportion of positive sales pairs for the full sample is 14.12%. Thus, the zero problem also exists in our context. We develop an approach to solve this zero problem.

4 Distance Elasticity at the Province Level

As a first step to examine the role that distance plays in Chinese e-commerce, we estimate a gravity model. In particular, we take the Anderson and Van Wincoop (2003) approach⁵ and use their estimating equation (1):

$$\ln[\text{exports}_{ij}] = \alpha + \gamma \ln[\text{dist}_{ij}] + \theta_i + \theta_j + \beta X_{ij} + \varepsilon_{ij}, \quad (1)$$

where exports from province i to province j are a function of distance, importer and exporter fixed effects, and covariates like own-country or neighboring province indicators. The importer and exporter fixed effects cover both the size of the respective economies as well as each partner's unobserved multilateral resistance to trade. Anderson and Van Wincoop (2003) point out that traditional gravity model estimation that omitted fixed effects would neglect the multi-lateral trade barriers between a country and the outside world and therefore confound this with the bilateral trade barriers between country pairs.

We estimate this model using trade flows between provinces constructed from our transaction data provided in Table 1. We include a same-province indicator to capture the home-province bias. Table 3 shows the province-level estimation results and compare it with the estimates in the literature. Distance has a significant effect on the number of transactions (elasticity: -0.18). This number is in between the distance elasticity from eBay data in the US (-0.07) and the one from MercadoLibre data in

⁵The Anderson and Van Wincoop (2003) gravity model is founded on a model of monopolistic competition in which countries specialize in the production of each good and consumers have CES preferences. Their model boils down to the estimation equation (1).

Latin America (-0.38). Moreover, consumers tend to buy a lot more from the sellers in the same province (own-province coefficient: 1.6). Online shoppers in our sample seem to have much stronger home bias relative to Latin America (own-province coefficient: 1.01). However, we need to be cautious when interpreting our results as we are focusing on the electronics category in this paper.

Table 3: Distance Elasticity at the Province Level

	Taobao	Hortaçsu et al. (2009) eBay	MercadoLibre	Hillberry et al. (2003) US: Inter-state
$\ln[\text{Distance}]$	-0.18*** (0.2)	-0.07*** (0.01)	-0.38*** (0.03)	-1.05 (0.02)
Same-Province/State	1.62*** (0.06)	0.56*** (0.09)	1.01*** (0.16)	0.44 (0.10)
Same-Country			6.07*** (0.08)	
Seller F.E.	Yes	Yes	Yes	Yes
Buyer F.E.	Yes	Yes	Yes	Yes
Observations	841	2181	7175	2304
Adjusted R^2	0.98	0.94	0.69	0.91

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sample Construction: We drop the transactions with missing locations, seller reputation or buyer experience. The total number of transactions reduces from about 13 million to over 9 million as shown in Table 1.

5 The Model

Our data contain information on seller and buyer experience on the C2C online marketplace, measured by their reputation. This information provides an opportunity to push the literature further and to examine the role that seller and buyer experience and how they interact in the marketplace. While recent trade literature adopts the ‘new-new’ trade theory to examine the importance of quality of goods in trade patterns, we believe that marketplace experience should operate in the same way and we have

an opportunity to test this here. To do so, we adapt the standard intra-industry trade model in the ‘new-new’ trade literature by incorporating seller and buyer experience in the marketplace.

In particular, we extend Crozet et al. (2012) to incorporate buyer-side heterogeneity. Our extended model allow us to not only examine the role of seller experience but also the role of buyer experience. The ‘new-new’ trade model explicitly models the extensive and intensive margins of trade. It can then help to deal with the high proportion of zeros in bilateral trade that commonly exists in the data on international trade also in our e-commerce trade data.

5.1 The Model with Two-sided Heterogeneity (extending Crozet et al. (2012))

Consider China is comprised of J provinces, indexed by $j = 1, 2, \dots, J$. Each consumer in China has a subutility from purchasing electronics on the online marketplace. The subutility function for a consumer with l level experience in province j is given by

$$U_{lj} = \left[\int_{\omega \in \Omega_j} [a_j(\omega) b_l(s(\omega)) q_l(\omega)]^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}}, \quad (2)$$

where Ω_j is the set of varieties available in province j and $\omega \in \Omega_j$ denotes a particular variety of electronics provided by a seller. It is also used to indicate the seller hereafter. $q_{lj}(\omega)$ denotes quantity of variety ω purchased by the consumer with l level experience in province j . $a_j(\omega)$ is province j -specific demand parameters capturing the consumer’s preference for variety ω and $s(\omega)$ denotes the experience of seller ω . The $b_l(\cdot)$ is a consumer experience l -specific function that maps the experience of sellers into quantity equivalents in the utility of the consumer. $\sigma > 1$ is the constant elasticity of substitution between different varieties.

The utility function given by (2) differs from it in Crozet et al. (2012) in two aspects. First, we modify Crozet et al. (2012) to replace the quality of products by seller experience. Second, we extend Crozet et al. (2012) to incorporate buyer experience

through the function $b_l(\cdot)$. Other structures have similar interpretations as in Crozet et al. (2012). The province-specific demand parameters on each variety $a_j(\omega)$ can be interpreted as the connections between the seller ω with province j . In the estimation, the term $a_j(\omega)$ is our structural error term, which give us that the sellers with the same reputation sells different amount of products to the same province. The quality function $b_l(\cdot)$ captures the interactions between the seller and buyer experience.

We assume the subutility from purchasing online products enters a Cobb-Douglas full utility. Each consumer then spends a fixed share of his income on electronics purchased online, which implies that he will spend an exogenous amount of money, which is denoted by X_j for province j 's consumers. Solving the consumer's utility maximization problem gives us the expenditure share of seller ω (located at province i), for consumers with l level experience in province j :

$$\frac{x_{lj}(\omega)}{X_{lj}} = \frac{[p_j(\omega) / [a_j(\omega) b_l(s(\omega))]]^{1-\sigma}}{\int_{k \in \Omega_j} [p_j(k) / [a_j(k) b_l(s(k))]]^{1-\sigma} dk} \mathcal{E}_{lj}(\omega), \quad (3)$$

where $x_j(\omega)$ is the money that the consumer in province j spends in variety ω , which is also the exports value that seller ω exports to province i . $p_j(\omega)$ is the price charged to the consumer in province j . $\mathcal{E}_{lj}(\omega)$ is a dummy variable, which is set to one if seller ω sells its products to level l experience buyers in province j .

As in standard ‘new-new’ trade theory models, the online market in every province is characterized by monopolistic competition and free entry. Each seller in the market provide one variety for sale. Each seller entering into the market need to pay a province-specific fixed entry cost, F_j . While e-commerce is cheaper than maintaining a “bricks and mortar” store, F_j not only represents the cost of maintaining an online store but also has the flexibility of allowing province-specific costs such as maintaining sales momentum in a particular province. Although small businesses tend to have a short lifespan, the fixed cost in our context might also accounts for the short-run nature of our 7-month data panel. A seller located at province i incurs an “iceberg” transportation cost, τ_{ij} , where $\tau_{ij} = 1$ for $i = j$ and $\tau_{ij} > 1$ for $i \neq j$. Thus, the profit that the seller

obtains from selling products to l level experience consumers in province j is given by

$$\pi_{lj}(\omega) = [p_j(\omega) - c(s(\omega))\tau_{ij}]q_{lj}(\omega) - F_j\mathcal{E}_{lj}(\omega), \quad (4)$$

where $c(s)$ is the marginal cost and $c'(s) > 0$ captures that more experienced sellers incur higher marginal costs as they provide better products or services. Solving the seller's profit maximizing problem, we obtain the price equation as follows:

$$p_j(\omega) = \frac{\sigma}{\sigma - 1} c(s(\omega))\tau_{ij}. \quad (5)$$

It shows the well-known result that the seller obtains a constant markup, $\sigma/[1 - \sigma]$, for models of monopolistic competition with CES utility function. Note that the price charged in one province is the same for all consumers.

We then combine the maximization choices of consumers and sellers to derive the exports value for the seller ω located at province i from exporting products to province j . Substituting the price equation (5) into the expenditure share of seller ω (3) yields

$$x_{lj}(\omega) = \left[\frac{a_j(\omega)b_l(s(\omega))}{c(s(\omega))} \right]^{\sigma-1} X_{lj} P_{lj}^{\sigma-1} \tau_{ij}^{1-\sigma} \mathcal{E}_{lj}(\omega),$$

where the price index P_{lj} is given by

$$P_{lj} = \left[\int_{k \in \Omega_j} \left[\frac{a_j(k)b_l(s(k))}{c(s(k))\tau_{i(k)j}} \right]^{1-\sigma} dk \right]^{1/[1-\sigma]},$$

in which $\tau_{i(k)j}$ indicates the transportation costs of seller k located at province i selling products to province j .

We now follow Crozet et al. (2012) to define the consumer experience specific attractiveness of a province by $A_{lj} \equiv X_{lj} P_{lj}^{\sigma-1}$, which indicates that if a l -experience consumer spends more income in online products or the price index is higher for the products, the province is more attractive to consumer with l level experience. Abbre-

viating $\alpha_j(\omega) \equiv a_j(\omega)^{\sigma-1}$, the value of exports can be rewritten as

$$x_{lj}(\omega) = \left[\frac{b_l(s(\omega))}{c(s(\omega))} \right]^{\sigma-1} A_{lj} \tau_{ij}^{1-\sigma} \alpha_j(\omega) \mathcal{E}_{lj}(\omega) \quad (6)$$

Moreover, it is well-know that in the Dixit-Stiglitz model of monopolistic competition with CES utility function, the profit to sales ratio has a constant ratio, $1/\sigma$. Using this result, we obtain the profit of seller ω from exporting to province j as:

$$\pi_{lj}(\omega) = x_{lj}(\omega) / \sigma - F_j \mathcal{E}_{lj}(\omega). \quad (7)$$

Substituting the exports value given by (7) into above profit, we obtain the probability of seller ω exporting to consumers with l level experience in province j

$$\Pr(\mathcal{E}_{lj}(\omega) = 1) = \Pr\left(\left[\frac{b_l(s(\omega))}{c(s(\omega))}\right]^{\sigma-1} A_{lj} \tau_{ij}^{1-\sigma} \alpha_j(\omega) > F_j \sigma\right). \quad (8)$$

5.2 Estimating Equations

We parametrically estimate extended model using the seller-level data for two different dependent variables, the probability of exporting and the export value. Let us consider the probability of exporting first. We take logs of exporting probability equation (8), which gives us

$$\Pr(\mathcal{E}_{lj}(\omega) = 1) = \Pr([\sigma - 1] \ln[b_l(s(\omega))/c(s(\omega))] - \ln[\sigma F_j] + \ln A_j + [1 - \sigma] \ln \tau_{ij} + \ln[\alpha_j(\omega)] > 0).$$

Assuming $\ln[\alpha_j(\omega)]$ is distributed normally with variance ψ^2 implies a probit form for estimating the exporting entry equation:

$$\Pr[\mathcal{E}_{lj}(\omega) = 1] = \Phi\left[\psi^{-1}\left(\beta_s \ln[s] + \beta_l \ln[l] + \lambda [\ln[s] \times \ln[l]] - \gamma \ln D_{ij} + EFE_j\right)\right], \quad (9)$$

where Φ represents the standard normal cumulative distribution function. $\beta_s \ln[s] + \beta_l \ln[l] + \lambda [\ln[s] \times \ln[l]]$ are parametric terms capturing $\ln[b_l(s(\omega))/c(s(\omega))]$. D_{ij} is the

log distance between country i and j . $\tau_{ij}^{\sigma-1}$ is set to D_{ij}^γ . EFE_j is the exporting probability fixed effect, capturing $-\ln(\sigma F_j) + \ln A_j$.

We also estimate a specification only considering seller experience,

$$\Pr(\mathcal{E}_j(\omega) = 1) = \Phi\left(\psi^{-1}\left(\beta_s \ln[s] - \gamma \ln D_{ij} + EFE_j\right)\right), \quad (10)$$

which can be derived by assuming function $b(\cdot)$ is independent with the buyer experience.

We now consider the exports value given by (6). Taking log value of (6) gives us the seller-level exports value for each level experience buyer,

$$\ln[x_{lj}(\omega)] = [\sigma - 1] \ln[b_l(s(\omega))/c(s(\omega))] + \ln[A_j] - \gamma \ln D_{ij} + \ln[\alpha_j(\omega)] + \ln \mathcal{E}_j(\omega),$$

which gives us the estimating equation by applying the same parametric specification as exporting probability,

$$\ln[x_{lj}(\omega)] = \beta_s \ln[s] + \beta_l \ln[l] + \lambda [\ln[s] \times \ln[l]] - \gamma \ln D_{ij} + XFE_j, \quad (11)$$

where XFE_j is the province- j fixed effect for exporting value. Similarly, we will also estimate an equation which only considers seller experience,

$$\ln[x_j(\omega)] = \beta_s \ln[s] - \gamma \ln D_{ij} + XFE_j. \quad (12)$$

Consider running regressions using the equations (10) to (12). If we observe each seller sells products to all provinces, then we would have $\mathcal{E}_j(\omega) = 1$ for all j . If we observe many sellers do not sell to some markets, then we would have $\mathcal{E}_j(\omega) = 0$ for many provinces j . However, if we were to run regressions treating $x_j(\omega)$ as zeros, the estimation of β_s will be inconsistent. The intuition is that the probabilities of exporting increases with $\alpha_j(\omega)$. It follows that sellers with more experience will become exporters even with relatively low draws of $\alpha_j(\omega)$, whereas less experienced sellers need high draws

of $\alpha_j(\omega)$ to be observed as exporters. This implies that $\ln[s]$ is negatively correlated with $\alpha_j(\omega)$ in the selected data. The estimates of the effects of $s(\omega)$ will be biased towards zero.

Helpman et al. (2008) apply the Heckman correction to address the sample selection issue. It is well-known that the Heckman approach can be identified off functional form but we can only have confidence in the results if we have an instrumental variable explaining the seller-level decision of exporting to individual province is excludable from equations (11) and (12). However, in our context, it is very difficult to conceive of such an instrumental variable. Thus, to solve the zero problem, we follow Crozet et al. (2012) and apply a Tobit estimation approach.

6 Results

We estimate our extended models using the seller-level data. We use the buyer and seller reputation to proxy for the seller and buyer's experience. We run five different regression models to estimate the role of the seller and buyer's experience: (i) Linear Probability Model (LMP) regression; (ii) Probit Model regression; (iii) OLS regression without zero exports; (iv) OLS regression including zero exports; (v) Tobit Model regression. The OLS has the advantage in handling a large number of fixed effects but it also introduces selection bias due to the zero problem. The Tobit model can correct the selection bias problem.

To focus on the seller's experience, we first estimate the equations considering only the seller's experience, (10) and (12). Table 4 gives the estimates for the five regressions. The estimates for the Probit model is the marginal effects. The estimation results in column (1) and (2) shows that more experienced sellers tend to raise the export probability. They tend to sell more and farther (to more provinces). This suggests that for sellers, they should accumulate more experience through finishing transactions to increase their sales in the future. Especially, it may worthwhile for start sellers to reduce their price at the early stage of their business. This can increase their reputation

Table 4: Baseline Model Estimation

	(1) LMP	(2) Probit	(3) OLS-1	(4) OLS-2	(5) Tobit
ln[Seller Reputation]	0.1420*** (0.0002)	0.1520*** (0.00025)	-0.0070 (0.0057)	1.0510*** (0.0019)	9.1940*** (0.0176)
ln[Distance]	-0.0405*** (0.0001)	-0.0224*** (0.0001)	-0.1040*** (0.00133)	-0.3460*** (0.0010)	-1.2750*** (0.0049)
Constant	0.3300*** (0.0011)	-1.4320*** (0.0052)	7.9630*** (0.0490)	2.7890*** (0.0083)	-15.03*** (0.0533)
Sigma Constant					10.51*** (0.0107)
Buyer F.E.	Yes	Yes	Yes	Yes	Yes
R^2	0.1430	0.1893	0.0110	0.1370	0.0943
Observations	5142683	5142683	726219	5142683	5142683

Standard errors in parentheses

*($p < 0.05$), **($p < 0.01$), ***($p < 0.001$)

Pseudo- R^2 are provided for Probit and Tobit models

Sample Construction: we drop the transactions with missing information on seller characteristics from the data presented in Table 1, which leaves us 7.8 million transactions. We then aggregate the transactions into province level for sellers. We have 165893 unique sellers in our data. We include only the 31 provincial administrative units in mainland. The number of observations in this regression equals to the product of the number of sellers, the number of provinces for buyers and number of buyer reputation, $165893 \times 31 \times 5 = 5,142,683$. There are 726,219 seller-province pair with positive transaction, which is about 14.12% of the total observations.

in the marketplace, which is valuable for them.

Comparing the results in column (3) and (4) in Table 4 shows that the OLS regression without export zeros shrinks the effect of seller reputation on the exports value. It even changes the sign of the effect. This is because that we have a very high proportion zero exports observations. Comparing the results in column (4) and (5) confirms the bias direction of the OLS estimation indeed shrinks the effect of the seller's reputation. This is consistent with our prediction in our model. The estimation of distance in all five models has negative effects on export probability and export value. These results are consistent with the estimates in the trade literature.

We then estimate the extended model in which we consider both buyer and seller experience using (9) and (11). This allows us to investigate the role of buyer experience and how it interacts with distance and seller experience. We believe that the buyer’s marketplace experience plays important role in his (or her) online shopping behaviour. It is well-known that there exists information asymmetry in the online marketplace. Buyers learn to judge and screen the products in the marketplace by accumulating their purchasing experience. It would be interesting to examine whether the role of distance is the same across consumers with different experience. Moreover, it is also worthwhile to understand how the role of the seller’s reputation differs across consumers with different levels of experience.

Table 5 shows the estimates for our extended model. The estimates seller reputation in all five models show that seller reputation raises export probability and value. The estimation results from these two models in column (2) and (5) show that more experienced buyers tend to buy less electronics in the marketplace. This result may reflect that there are lots of flaw products sold online. Once consumers learn more about it, they tend to change their taste and buy less from the online marketplace. The estimates of the interactions between buyer and seller experience in column (2) and (5) are positive, implying that more experienced buyers tend to buy more from more experienced sellers. This indicates that more experienced buyers value the seller’s reputation in the marketplace. The reputation of sellers is valuable for sellers through increasing sales, especially through purchase from the more experienced buyers.

7 Conclusion

This paper examines the role of distance in Chinese e-commerce. Using the transaction level data from a dominant e-commerce website, Alibaba’s Taobao, we fit a classical gravity model by aggregating the data to trade flows between provinces. We find that the distance effect in Chinese e-commerce is larger than in the U.S., but smaller than in Latin America. This may be linked to the maturity of e-commerce market in those

Table 5: Extended Model Estimation

	(1) Linear Prob	(2) Probit	(3) OLS	(4) OLS	(5) Tobit
ln[Seller Reputation]	0.1240*** (0.0002)	0.0337*** (0.0002)	0.2510*** (0.0131)	0.9580*** (0.0019)	9.079*** (0.0442)
ln[Buyer Reputation]	0.0303*** (0.0003)	-0.0521*** (0.0003)	-0.0436 (0.0251)	0.2680*** (0.0026)	-14.38*** (0.0798)
ln[Distance]	-0.0147*** (0.0000)	-0.0040*** (0.0000)	-0.0663*** (0.0009)	-0.1170*** (0.0003)	-1.0650*** (0.0036)
ln[Seller Reputation] × ln[Buyer Reputation]	-0.0442*** (0.0002)	0.0124*** (0.0001)	-0.3010*** (0.0110)	-0.3690*** (0.0014)	3.399*** (0.0363)
Constant	-0.0343*** (0.0006)	-1.305*** (0.0080)	7.913*** (0.0430)	-0.2230*** (0.0044)	-16.35*** (0.0994)
Sigma Constant					12.82*** (0.0100)
Buyer F.E.	Yes	Yes	Yes	Yes	Yes
R^2	0.0760	0.2019	0.0260	0.0730	0.1170
Observations	25713415	25713415	1313822	25713415	25713415

Standard errors in parentheses

*p<0.05, **p<0.01, ***p<0.001

Pseudo- R^2 are provided for Probit and Tobit models

Sample Construction: we drop the transactions with missing information on seller characteristics from the data presented in Table 1, which leaves us 7.8 million transactions. We then aggregate the transactions into province level for sellers. We have 165893 unique sellers in our data. We include only the 31 provincial administrative units in mainland. The number of observations in this regression equals to the product of the number of sellers, the number of provinces for buyers and number of buyer reputation, $165893 \times 31 \times 5 = 25713415$. There are 1,313,822 seller-province pair with positive transaction, which is about 5.11% of the total observations.

areas. To further explore the detailed information on seller and buyer's experience in the marketplace, we estimate a 'new-new' trade model by using the seller-level transaction data. The results show that more experienced sellers sell more and farther in the marketplace. Our Tobit estimation corrects the selection bias problem existing in the data.

We will push our study further in several directions. First, we will investigate how

the seller and buyer's experience interacts with distance. This allows us to understand better the role of distance in e-commerce. Second, our data also contain information on the seller's characteristics including their demographics, services offered and promotional activity. We plan to examine the role of these seller features on transactions and how they interact with distance and experience of sellers and buyers. The study of sellers' characteristics will shed light on the design of the online platform.

References

- Anderson, J. E. (2010). The gravity model. Technical report, National Bureau of Economic Research.
- Anderson, J. E. and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, 93(1):170–192.
- Baldwin, R. and Harrigan, J. (2011). Zeros, quality, and space: Trade theory and trade evidence. *American Economic Journal: Microeconomics*, 3(2):pp. 60–88.
- Blum, B. S. and Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2):384 – 405.
- Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.
- Cai, H., Jin, G. Z., Liu, C., and Zhou, L.-a. (2014). Seller reputation: From word-of-mouth to centralized feedback. *International Journal of Industrial Organization*, 34:51–65.
- Cairncross, F. (2001). *The death of distance: How the communications revolution is changing our lives*. Harvard Business Press.
- Chu, J. and Manchanda, P. (2013). Quantifying cross-network effects in online c2c platforms. Technical report, Working Paper.
- Cowgill, B. and Dorobantu, C. (2014a). The us-canada border effect in online commerce.
- Cowgill, B. and Dorobantu, C. (2014b). Worldwide gravity in online commerce.
- Crozet, M., Head, K., and Mayer, T. (2012). Quality sorting and trade: Firm-level evidence for french wine. *The Review of Economic Studies*, 79(2):609–644.

- Fan, Y., Ju, J., and Xiao, M. (2014). Reputation premium and reputation management: Evidence from the largest e-commerce platform in china.
- Head, K., Mayer, T., et al. (2013). Gravity equations: Workhorse, toolkit, and cookbook. *Handbook of international economics*, 4.
- Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487.
- Hillberry, R. and Hummels, D. (2003). Intranational home bias: some explanations. *Review of Economics and Statistics*, 85(4):1089–1092.
- Hortaçsu, A., Martínez-Jerez, F. A., and Douglas, J. (2009). The geography of trade in online transactions: Evidence from ebay and mercadolibre. *American Economic Journal: Microeconomics*, pages 53–74.
- Houser, D. and Wooders, J. (2006). Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369.
- Ju, H., Su, L., and Xu, P. (2013). Pricing for goodwill: a threshold quantile regression approach.
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.
- Melnik, M. I. and Alm, J. (2002). Does a seller’s ecommerce reputation matter? evidence from ebay auctions. *The Journal of Industrial Economics*, 50(3):337–349.
- Wolf, H. C. (2000). Intranational home bias in trade. *Review of economics and statistics*, 82(4):555–563.
- Wu, C. (2014). Matching markets in online advertising networks: The tao of taobao and the sense of adsense.
- Xiao, J., Wei, L., and Liu, L. (2013). Regulation versus monitoring: Effects and welfare implications of trust management strategies in e-marketplaces.

- Yang, B., Yang, S., and Dutta, S. (2015). Platform service offering to business customers: Strategic considerations in engendering seller use of marketing tools on e-commerce platforms. *working paper*.
- Ye, Q., Xu, M., Kiang, M., Wu, W., and Sun, F. (2013). In-depth analysis of the seller reputation and price premium relationship: A comparison between ebay us and taobao china. *Journal of Electronic Commerce Research*, 14(1):1–10.

Figure 1: Total Sales of Province

Figure 2: Total Purchase of Province