


What Determines the Pattern of China's Cross-Border E-Commerce With the World?


Gaoju Yang, Zhejiang University, China

Yujie Wang, Zhejiang University, China

Futao Lu, Zhejiang University of Finance and Economics, China

 <https://orcid.org/0000-0001-6359-7930>

Linhui Yu, Zhejiang University, China

 <https://orcid.org/0000-0002-6057-1361>

Shuzhong Ma, Zhejiang University, China

ABSTRACT

Cross-border e-commerce has gradually expanded in international trade markets over the past decades. This paper analyzes the determinants of the volume of cross-border e-commerce in a gravity model framework. Moreover, the paper explores the role of internet popularity and finds a significant promotion effect on the volume of cross-border e-commerce. Furthermore, by utilizing cross-border express delivery data, the analysis indicates a significant difference in the impacts of the determinants between the aggregate and consumer levels and provides an important addition to the literature on e-commerce and international trade.

KEYWORDS

Cross-Border E-Commerce, Gravity Model, Internet Popularity

1. INTRODUCTION

Traditionally, international trade literature has focused on offline trade flows, mainly international business-to-business or bulk trade. However, with the help of information and communication technologies, international trade has developed beyond the traditional scope. Hence, other new types of transactions, in a more private setting, have become an area worth following. The World Wide Web is becoming the predominant vehicle for accessing and transmitting information globally, and, in the late 20th century, substantial studies sprouted focusing on the influence of the Internet on trade at the macro-level.

In the trade literature, Freund and Weinhold (2004) find a significant effect of the Internet (measured by growth in web hosts in a country) on the growth of goods exports, which is consistent

DOI: 10.4018/JGIM.20210901.oa4

This article, published as an Open Access article on May 14, 2021 in the gold Open Access journal, Journal of Global Information Management (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

with a theoretical model in which the Internet reduces market-specific fixed trade costs. On overseas e-commerce, Shi (2016) explores the impact of the Internet, indicated by the mutual links data from Chung (2011), on the trade of manufacturing goods, using firm-level customs statistics. Shi constructs detailed firm-volume bilateral trade flows. On exports of services, Freund and Weinhold (2002) show that the development of Internet use in its partner countries has resulted in increased exports of services to the United States. On digital products, Blum and Goldfarb (2006) point out that a gravity model also holds digitally, as physical distance harms the online consumption of taste-dependent digital products, such as music and games, based on a sample of U.S. Internet users. On audiovisual services, Hellmanzik and Schmitz (2015) find that “virtually proximate” countries trade significantly larger amounts of audiovisual services, and virtual proximity has a greater impact on trade in audiovisual services than on total services trade.

Accordingly, the changes in promotion methods have indeed affected consumption habits (Kumar et al, 2006). Since the notable expansions of Amazon and eBay in the 2010s, substantial studies have made efforts to shed light on the influence of online competition on retail business domestically. Naturally, with the development of the Internet and global express network, global free trade is not restrained to digital goods. There are several differences between traditional international trade and new e-commerce. Compared with traditional consignors, e-business suppliers have various options for reducing the distance to their online clients abroad. For example, Gorodnichenko and Talavera (2017) study the pass-through of the exchange rate in online retailers. Cavallo et al. (2014) and DellaVigna and Gentzkow (2017) document the uniform pricing conditions for online and offline retailers separately. Suppliers can also improve the time barriers caused by distance to their clients. They can reduce transit time and delays by offering fast transport modes, such as express delivery, which results in shorter times en route between product order and delivery to the client.

This paper aims to (1) observe the elasticity of other gravity-model determinants by introducing an indicator for Internet development, for instance, the parameter variations of distance, gross domestic product (GDP) per capita, and so forth; and (2) compare the macro and micromechanisms of country heterogeneity for cross-border purchasing behaviors, especially the distribution of consumer choice (i.e., price preference) and the volume of each purchase (i.e., aggregated package value).

We first link our e-commerce study to the traditional trade literature by studying the volume of cross-border e-commerce in a gravity model framework. The gravity model is widely used in the area of international trade to study bilateral trade flows. The model assumes that the volume of trade between any two countries is a function of each country’s GDP and the distance between the countries. The literature finds that GDP has a positive effect on bilateral trade volume, and distance negatively affects volume. Other than goods trade, Kimura and Lee (2006) study services trade in a gravity model framework and find better prediction power. Hellmanzik and Schmitz (2015) investigate the impact of virtual proximity on audiovisual services, also in a gravity model setup. The literature also refers to institutional and psychological distance, such as trade unions, common language, and history (Head and Mayer 2014; Kim et al. 2017; Ndzendze and Monyae 2019). A key difference between the traditional bulk trade and e-commerce is the requirement of the Internet. Low Internet popularity may also potentially enlarge the distance between buyers and sellers in e-commerce. Our estimation finds that the gravity model is well-grounded in our e-commerce data set and provides insights into the determinants of the volume of cross-border e-commerce. Compared with the determinants of traditional bulk trade, we further highlight the impact of Internet popularity.

The paper fills the vacuum in the research on micro-level e-commerce. There is little literature on cross-border online shopping, especially in the individual dimension, since a key impediment is the relative scarcity of data on cross-border e-commerce. We explore patterns and characteristics of overseas online shopping, utilizing data on cross-border express deliveries from China to other countries. As one of the most developed regions of cross-border e-commerce in the world, nearly 30% of China’s trade is cross-border e-commerce, which reached 7.6 trillion *yuan* last year, with a growth rate of 20.6%. Based on the huge amount of underlying transactions, we utilize data from

one of the top three cross-border logistics companies in China. The database contains more than six million records of cross-border deliveries from China in 2016 and 2017 to 189 destination countries. By applying these data, we highlight the factors that affect the aggregate volume of cross-border deliveries. And with detailed information on the value of each package, we further examine cross-border consumer behaviors.

The rest of the paper is organized as follows. Section 2 describes the gravity model as the theoretical foundation. Section 3 presents the data. Empirical results on aggregate-level volume and micro-level distribution are presented in sections 4 and 5, respectively. This paper concludes with section 6.

2. AUGMENTED GRAVITY MODEL

We use a gravity model with multi-country monopolistic competition, following Felbermayr and Toubal (2010), as our theoretical foundation. Hellmanzik and Schmitz (2015) use the same approach to study the virtual proximity effect. Each country i 's populated by a representative individual who derives utility from consuming different varieties of a differentiated good according to a standard constant elasticity of substitution function:

$$U_{it} = \sum_{j=1}^C a_{ijt}^{(\sigma-1/\sigma)} \sum_{z=1}^{n_{jt}} (m_{zijt})^{(\sigma-1/\sigma)} \quad (1)$$

where z denotes the index of a generic variety; n_{jt} is the number of varieties produced in country j at time t , and $\sigma > 1$ is the elasticity of substitution between varieties. The quantity of consumption in country i of variety z from country j is m_{zijt} . Following Combes et al. (2005), we allow for a specific weight, $a_{ijt} \geq 0$, to describe the special preference of the representative consumer in country i for goods from country j .

Maximizing equation (1), subject to an appropriate budget constraint, yields country i 's demand quantity m_{ijt} for a generic variety. Calculating the cost, insurance, and freight value of total imports from country j at time t as $M_{ijt} = n_{jt} p_{ijt} m_{ijt}$, we find:

$$M_{ijt} = \left(\frac{a_{ijt}}{t_{ijt}} \right)^{\sigma-1} \mu_{it} \phi_{jt} \quad (2)$$

The variables $\mu_{it} = E_{it} P_{it}^{\sigma-1}$ and $\phi_{jt} = n_{jt} p_{jt}^{1-\sigma}$ collect terms that depend only on country i 's or j 's characteristics:

$$P_{it} = \left[\sum_{j=1}^C \left(\frac{a_{ijt}}{t_{ijt}} \right)^{\sigma-1} p_{jt}^{1-\sigma} n_{jt} \right]^{1/(1-\sigma)} \quad (3)$$

where P_{it} is the aggregate price index and E_{it} denotes country i 's GDP.

The model captures cultural affinity by Π_{ijt} , including Internet popularity and other traditional measures of cultural proximity. We use $\tilde{\Pi}_{ijt}$ to denote the level of Internet convenience (NET_i), K

to indicate the traditional measure, and κ as a vector of parameters, following Felbermayr and Toubal (2010) and Hellmanzik and Schmitz (2015):

$$\Pi_{ijt} = \tilde{\Pi}_{ijt} + \kappa K_{ijt} \quad (4)$$

The traditional measure of cultural proximity, K includes the stock of migrants (MIG_{ijt}) and common language (LAN_{ij}). To specify country i 's cultural affinity to country j , we assume that country i 's preference for goods from j , a_{ijt} , depends on cultural proximity P_{ijt} in the following way:

$$\ln a_{ijt} = \alpha \Pi_{ijt} \quad (5)$$

where $\alpha > 0$. Similarly, trade costs, t_{ijt} also depend on P_{ijt} . In addition to cultural affinity, we take physical distance, which affects the delivery cost ($\ln DIST_{ij}$), and the existence of a common border (BOR_{ij}) into consideration. Therefore, we obtain the following iceberg costs equation:

$$\ln t_{ijt} = \delta \ln DIST_{ij} - \gamma BOR_{ij} - \pi \Pi_{ijt} \quad (6)$$

All the parameters are expected to be positive. Combining equations (3) to (6), we obtain our main log-linear gravity model specification:

$$\ln M_{ijt} = \beta_1 \ln DIST_{ij} + \beta_2 BOR_{ij} + \beta_3 LAN_{ij} + \beta_4 Net_i + \beta_5 \ln MIG_{ijt} + \sigma Y_{it} + a_{jt} + \varepsilon_{ijt} \quad (7)$$

where Y_{it} are country i 's specific variables, and a_{jt} is the fixed effect. Following the traditional gravity model setup, we take country i 's GDP and population as Y_{it} , which reflects country i 's income level. Then we eventually obtain our baseline log-linear gravity model specification:

$$\ln M_{ij} = \lambda_1 \ln GDP_i + \lambda_2 \ln DIST_{ij} + \lambda_3 Net_i + \lambda_4 \ln POP_i + \lambda_4 BOR_{ij} + \lambda_5 LAN_{ij} + \lambda_6 MIG_{ij} + \varepsilon_{ij} \quad (8)$$

where j represents China in our empirical analysis and MIG_{ij} represents the stock of Chinese migrants in country i .

3. DATA COLLECTION

3.1 Cross-Border Express Delivery Data

In this paper, we build our empirical study with substantial data on cross-border delivery provided by a leading Chinese international logistics enterprise. The data set covers more than 222 days (from

August 16, 2016, to March 26, 2017). Due to the high integrity and amount of information, we focus on the subset of express deliveries instead of regular deliveries. The express deliveries subset contains more than six million cross-border delivery records, with an average of around 30,000 packages delivered from mainland China to 189 countries and regions globally in a single day.

For each day's data pool, we collected the following information on each package: its name (i.e., the sequence number of that package), categories of commodities inside, number of commodities inside, declared price, and gross weight. We also obtained desensitized delivery information for the package senders and consignees, including the address and freight for each transaction. Table 1 gives a detailed overview of the subset of express delivery data employed in this paper. Table 2 reports the top 10 countries in e-commerce volume from the total package numbers and total volume.

Concern about the data is that they may not reflect precisely the panorama of China's cross-border commodity transactions. The data source enterprise is among the top cross-border logistics providers in China, which serves more than 100,000 cross-border e-commerce corporations and entities, including prominent online sellers such as Alibaba, Amazon, and eBay. Furthermore, our analysis at the aggregate level is highly consistent with the results of Wang et al. (2017), which are based on macro-data from the China E-Commerce Research Center. A striking benefit compared with the usual macro trade data is that our subset of express delivery data is much more detailed

Table 1. Summary statistics for the express data set

Variable	Unit	Number	Mean	S.D.	Min	25%	50%	75%	Max
Weight	kg	6320332	0.270	0.280	0.01	0.08	0.17	0.34	1.63
Price	USD	6320332	6.020	13.54	0.04	2	4.36	7.45	50
Quantity		6320332	1.610	5.240	1	1	1	1	20
Value	USD	6320332	7.670	57.02	0.06	2	5	8	65
Freight	USD	6320332	25.23	22.71	0.54	12.53	19.31	31.85	132.3
Country		180							

Note: This is the summary statistics winsorized at 0.5% to 99.5%; 1) 25% 50% 75% represent the value at 25 50 and 75 percentiles; 2) Value means the total reported value of the items in one package; 3) Quantity is the total quantity of items in one package; 4) Price is the price of the items. With more than one different item in one package, we take the price equal to Value/Quantity; 5) Country is the number of destination countries.

Table 2. Top 10 volume countries in package numbers and the total value

Country	Package	Country	Total Value (USD)
Brazil	1723390	Brazil	9023358
United States	904115	United States	6073019
Russian Federation	556855	Russian Federation	5614892
Turkey	494286	Turkey	3388761
India	205586	Belgium	1459725
Belgium	175990	India	1328171
Italy	175296	Italy	1214508
Hungary	170330	France	1175763
Portugal	167602	Portugal	1067653
France	156388	Hungary	1065005

and, consequently, informative, which supports our microscopic analyses of the distributions and behaviors of cross-border e-commerce.

Another concern is that the Internet may distort enterprises' export entering behaviors. More plainly, some ephemeral online shops also retail their products overseas simply because online shopping platforms and websites have attenuated the threshold for exporting to zilch. This distortion can bring out some inferior records that are not conducted by robust cross-border sellers. Therefore, we reviewed the reliability of the global sellers by double-checking their names and addresses provided in our express delivery subset. We found that most vendors (about 95%) have quite a bundle of records labeled with the same commodity but different delivery destinations. This makes sense in supporting the robustness of our data, and also minimizes the potential distortions, such as vicious underpricing and false scalping. Furthermore, considering the relatively high price of international shipment, frequent and intensive noncommercial behaviors are not affordable or reasonable for most households (Rabinovich et al., 2008).

3.2 Gravity Model Data

The standard geographic variables were obtained from the Centre d'Études Prospectives et d'Informations Internationales (CEPII) and the World Bank's World Development Indicators. Distance collected from CEPII captures the bilateral distance between the two countries' capitals. We obtained data for eight other dummy variables evaluating the nexus between mainland China and our destination countries and regions. For instance, we employ indicators for countries and regions that share a common border.

3.2.1 Common Language

The data on common language come from a diverse set of data from CEPII. They measure evidence of linguistic connections, including equality of official and dominant languages (*comlang off*, *comlang ethno*). The data thus go beyond the traditionally used measures of the common language. Melitz and Toubal (2004) find that that common language strongly impacts trade in goods.

3.2.2 Migrants

Data on the stock of migrants were obtained from the World Bank's International Bilateral Migration database to control for demand by major immigrant groups in the host countries. Furthermore, we use a six-dimension index for cultural proximity/distance, based on Hofstede (2013). Data on GDP in 2010 U.S. dollars, population, were retrieved from the World Bank's World Development Indicators.

Since personal overseas shopping behaviors can overcome geographic boundaries with the rise of the Internet, we chose virtual proximity indicators as the "digital" proxy for cultural proximity or purchasing convenience. Accordingly, the main contribution of this paper in data collection is gathering information from various sources on this proxy. This task was complicated because data on bilateral personal purchasing behaviors are underexploited despite the Internet's growing importance.

3.2.3 Internet Popularity

To capture information flow via the Internet, we used the number of netizens (i.e., access to fixed broadband) per 100 people, retrieved from the World Bank. It is a widely accepted indicator of Internet popularity from the scope of infrastructure (Tang et al., 2020). Felbermayr and Toubal (2010) and Hellmanzik and Schmitz (2015) use hyperlink data from 2003 and 2009 as an indicator of Internet popularity, following Chung (2011). We were not able to update the data for recent years after Yahoo terminated the AltaVista service in 2013. Our results still hold, using the hyperlink data from Chung (2011).

Table 3 provides summary statistics for all the gravity model determinants described in this section.

Table 3. Summary statistics of gravity model determinants

Variable	No.	Mean	S.D.	Min	0.250	Mdn	0.750	Max
ln GDP	189	24.14	2.360	19.02	22.64	24.19	25.99	30.55
ln DIS	189	9	0.520	6.700	8.790	9.030	9.380	9.870
Net100	189	14.43	14.06	0.0100	1.770	9.740	26.27	48.35
ln POP	189	15.42	2.280	10.41	14.09	15.78	17.13	21
LAN	189	0.0100	0.100	0	0	0	0	1
BOR	189	0.0700	0.260	0	0	0	0	1
ln MIG	189	8.340	3.590	0	6.440	8.530	10.44	17.48

Note: 1) DIS for distance in km, in log; 2) POP for population, in log; 3) MIG for Chinese migrants' stock, in log; 4) LAN is the common language dummy; 5) BOR is the common border dummy.

4. AGGREGATE-LEVEL ESTIMATION

4.1 Baseline Results

We first run our specifications at the aggregate level for the whole sample period to analyze the impact of potential determinants of cross-border express delivery volume in our log-linear model, equation (8). The log of the summation of the reported package values to the same country is used as the dependent variable. The results are reported in Table 4.

We start with the simplest case of the gravity model, with only log GDP and distance as independent variables, and report the results in Table 4, column 1. We find a positive coefficient of 1.215 for log GDP and a negative coefficient of -0.785 for *log* distance. Hence, the total volume is positively correlated with destination countries' GDP and negatively correlated with the distance between China and the destination country. This finding is consistent with the results for traditional offline bulk trade.

To examine the impact of countries' Internet popularity on cross-border express deliveries, we add the countries' Internet users per 100 people as an independent variable and report the results in Table 4, column 2. A positive coefficient of 0.077 indicates a positive effect of Internet popularity on the volume of e-commerce. This result indicates that Internet development matters for the volume of e-commerce.

In Table 4, columns 3 to 5, we add log population, common language, common border, and log migrants as control variables to examine other potential determinants of the total volume of e-commerce. In the gravity model for traditional bulk trade, common border usually has a significantly positive effect on bilateral trade volume. Our estimation finds a positive but insignificant coefficient for the common border, which is due to the difference in shipment methods. In traditional bulk trade, a common border creates the option for land transportation. Our sample is cross-border express deliveries, which generally use flight as the only shipment method. Hence, the common border effect diminishes. We find a positive and insignificant coefficient for the common language dummy and a negative and insignificant coefficient for the stock of migrants. We will further discuss these two effects on our micro-level estimation.

Another way to define volume is to use the total number of packages sent from China to the destination country. To examine the effect on the number of packages, we replace the dependent variable in equation (8) with the log total number of packages. Table 5 reports the estimation results. The coefficients are consistent with the results in Table 4, suggesting that the gravity model is a robust fit to our data set.

Our estimation at the aggregate level provides insights on the determinants of cross-border express delivery volume. We find similar determinants between traditional bulk trade and cross-border

Table 4. Gravity model estimations of the total value

VARIABLES	ln Total Value				
	(1)	(2)	(3)	(4)	(5)
ln GDP	1.215*** (0.069)	1.114*** (0.061)	1.307*** (0.159)	1.291*** (0.161)	1.313*** (0.170)
ln DIS	-0.785** (0.319)	-0.789*** (0.276)	-0.827*** (0.277)	-0.801*** (0.279)	-0.750** (0.335)
Net100		0.077*** (0.010)	0.060*** (0.016)	0.061*** (0.016)	0.061*** (0.016)
ln POP			-0.226 (0.172)	-0.210 (0.174)	-0.219 (0.176)
LAN				0.955 (1.299)	1.098 (1.345)
BOR					0.161 (0.646)
ln MIG					-0.018 (0.046)
Constant	-14.243*** (3.619)	-12.881*** (3.129)	-13.478*** (3.155)	-13.579*** (3.163)	-14.306*** (3.772)
Observations	189	189	189	189	189
R-squared	0.671	0.756	0.759	0.760	0.760

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

express delivery, that is, GDP and distance. This evidence shows the strong explanatory power of using the gravity model to analyze cross-border express delivery volume. Furthermore, to distinguish the difference between these two types of trade, we highlight the impact of Internet popularity on cross-border express deliveries. This finding suggests an important addition to the literature.

4.2 Robustness Analysis

Zero trade flows are often observed in traditional international trade, which raises concerns about selection bias when estimating in a gravity model framework. In our analysis, this potential problem also exists. Since we only have records for countries with e-commerce delivery activity with China, our previous results for the determinants are only significant in explaining the volume of nonzero cross-border e-commerce. To deal with this concern, we follow Santos Silva and Tenreiro (2006) and use the Poisson quasi-maximum likelihood estimation method (PPML). We expand our data set with countries that do not appear in our original data set that have zero volume of cross-border e-commerce trade.

We report the results for total value and total packages in column 1 of Table 6 and 7, respectively. The coefficients have the same signs as the results in Table 4 and Table 5, with differences in significance levels. The distance effect is negative but insignificant¹ in our PPML regression, which is caused by the different methodologies and zero trade flow added to the model. This may be due to the

Table 5. Gravity model estimations of package number

VARIABLES	ln Package Number				
	(1)	(2)	(3)	(4)	(5)
ln GDP	1.138*** (0.066)	1.039*** (0.058)	1.202*** (0.152)	1.181*** (0.154)	1.244*** (0.161)
ln DIS	-0.755** (0.307)	-0.759*** (0.263)	-0.791*** (0.264)	-0.759*** (0.267)	-0.678** (0.319)
Net100		0.075*** (0.009)	0.062*** (0.015)	0.063*** (0.015)	0.060*** (0.015)
ln POP			-0.191 (0.164)	-0.171 (0.166)	-0.187 (0.168)
LAN				1.189 (1.241)	1.598 (1.278)
BOR					0.212 (0.614)
ln MIG					-0.059 (0.044)
Constant	-14.564*** (3.483)	-13.224*** (2.989)	-13.728*** (3.017)	-13.854*** (3.021)	-15.346*** (3.585)
Observations	189	189	189	189	189
R-squared	0.660	0.752	0.754	0.755	0.758

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

different methodologies and zero trade flow added to the model. The income and Internet popularity effects remain positive and significant, suggesting a robust effect of these two determinants.

We also report the regression results for the three-quarters of sample data separately in Table 7, columns 2 to 4. The results remain consistent with our full sample estimation. The changes in magnitudes may due to the large sample of zero trade flows added to the third quarter of 2016, since there are records for only 158 countries that quarter and more than 40 zero trade flows were added to the sample.

5. MICRO-LEVEL ESTIMATION

One of the biggest advantages of the data set is that it has every single cross-border delivery record from the delivery company over the period, for which we have detailed information on each package. In the previous section, we showed that the determinants in the aggregate-level estimation are similar to those for traditional bulk international trade. Then the natural question becomes: how do these determinants affect individuals' cross-border delivery behavior? From our data set, we are particularly interested in two variables, the value of each package and the price of the item(s) in the package. By analyzing the impact of the value of each package, we can shed light on the link between the determinants and consumers' purchasing power. However, the value is affected by the price and quantity of the item(s)

Table 6. Robustness test in PPML estimation of the total volume

ln Total Volume	(1)	(2)	(3)	(4)
	Full	2016Q3	2016Q4	2017Q1
ln GDP	0.191*** (0.027)	0.302*** (0.041)	0.222*** (0.031)	0.216*** (0.032)
ln DIS	-0.067 (0.054)	-0.091 (0.080)	-0.057 (0.057)	-0.092 (0.065)
Net100	0.005** (0.002)	0.005* (0.004)	0.005* (0.003)	0.006** (0.003)
ln POP	-0.036 (0.031)	-0.072 (0.046)	-0.047 (0.035)	-0.043 (0.036)
LAN	-0.184 (0.368)	-0.056 (0.358)	-0.112 (0.341)	-0.159 (0.375)
BOR	-0.002 (0.102)	0.017 (0.140)	-0.085 (0.123)	0.007 (0.110)
ln MIG	-0.014** (0.006)	-0.024** (0.009)	-0.019*** (0.007)	-0.019*** (0.007)
Constant	-1.430** (0.615)	-3.633*** (0.939)	-2.182*** (0.664)	-1.802** (0.728)
Observations	201	201	201	201
R-squared	0.682	0.652	0.691	0.652

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Column (1) presents the regression results of total volume overall sample periods; Column (2) (3) (4) present the results for three different quarters.

in the package. Therefore, we also analyze the impact on the price of the item(s) in each package. This analysis will help us better understand consumer choice in cross-border e-commerce.

We start with estimation using package value, replacing the dependent variable in equation (8) with the log of the value of the package. We present the results using all six million records in Table 8, column 1. The positive coefficient on log GDP and the negative coefficient on the log population suggest that consumers from countries with a higher level of GDP per capita have greater purchasing power online. The negative coefficient on *log* distance suggests that distance reduces consumers' purchasing power online. In the aggregate-level estimation, we find that countries with higher GDP per capita have a higher total volume of e-commerce, in value, and the number of packages. In this step, we also show that consumers from higher-income countries also tend to buy more at once. However, distance has a negative effect on aggregate-level volume and consumer choice.

In Table 8, column 2, we report the estimation results for the item price. This measures whether consumers want to buy cheaper or more expensive items cross-border. The coefficients have the same signs as the coefficients on package value, except for the stock of migrants. Columns 3 and 4, respectively, report the 25% and 75% quantile regression results on the package value. Columns 5 and 6, respectively, report the 25% and 75% quantile regression results on item price. The difference between the coefficients at 75% and 25% quantiles indicates the impact of the difference between high value/price and low value/price. GDP has a higher coefficient for high-value packages and high-price

Table 7. Robustness check in PPML estimation of total package number

In Package Number	(1)	(2)	(3)	(4)
	Full	2016Q3	2016Q4	2017Q1
ln GDP	0.231*** (0.031)	0.344*** (0.044)	0.269*** (0.035)	0.255*** (0.036)
ln DIS	-0.076 (0.065)	-0.077 (0.089)	-0.073 (0.070)	-0.100 (0.076)
Net100	0.006** (0.003)	0.008** (0.004)	0.007** (0.003)	0.007** (0.003)
ln POP	-0.041 (0.036)	-0.071 (0.050)	-0.049 (0.040)	-0.047 (0.041)
LAN	-0.059 (0.363)	0.168 (0.343)	0.035 (0.336)	-0.048 (0.374)
BOR	0.002 (0.115)	-0.004 (0.157)	-0.072 (0.134)	0.025 (0.124)
ln MIG	-0.025*** (0.007)	-0.038*** (0.011)	-0.031*** (0.008)	-0.029*** (0.009)
Constant	-2.475*** (0.735)	-5.151*** (1.053)	-3.412*** (0.812)	-2.867*** (0.851)
Observations	201	201	201	201
R-squared	0.688	0.684	0.701	0.656

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Column (1) presents the regression results of total volume overall sample periods; Column (2) (3) (4) present the results for three different quarters.

items, which indicates a positive impact of GDP on purchasing power and high-price consumption. Distance has a stronger negative effect on low-value packages and high-price items.

Table 9 presents our first test of country-level heterogeneity with the express delivery sample. We regrouped the data by percentiles of package value with an interval of 10, so that we can see the variations in the distribution of package value for each country-level variable. Table 9 shows the results by “percentile,” which means that the parameters are elasticities of country characteristics on “the number of percentiles,” not “package values.” Therefore, the results show the distribution of package values in view of different country characteristics. In the regression with organized express delivery data, we obtain results that have a sign that is inconsistent with the usual macro-level findings.

When we compare our results to the models of reference, we observe that the effect of GDP on cross-border deliveries is much smaller than in the macro-level data and not as significant. Only the packages with total values near the 80th and 90th percentiles demonstrate positively and significantly the expected GDP effects. For instance, in the case of the 90th percentile, a 10% increase in GDP leads to a 1% increase in “the number of packages at the 90th percentile of the package value.” Considering the distribution pattern of package values, our results show that a country with larger GDP may have a higher proportion of high-value packages compared with lower GDP countries. Conversely, the insignificant coefficient of -0.074 might reveal that a higher GDP diminishes the proportion of low-value packages. To put it briefly, the greater is the GDP, the greater is the number of high-value packages and the smaller is the number of low-value ones. This finding roughly outlines a scissor

Table 8. Micro-level estimation of package value and the item price

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	ln Value	ln P	25% ln V	75% ln V	25% ln P	75% ln P
ln GDP	0.120*** (0.001)	0.100*** (0.001)	0.087*** (0.001)	0.120*** (0.001)	0.060*** (0.002)	0.157*** (0.002)
ln DIS	-0.245*** (0.001)	-0.214*** (0.001)	-0.363*** (0.001)	-0.279*** (0.001)	-0.176*** (0.001)	-0.275*** (0.002)
Net100	-0.005*** (0.000)	-0.004*** (0.000)	-0.016*** (0.000)	0.003*** (0.000)	-0.004*** (0.000)	0.002*** (0.000)
ln POP	-0.154*** (0.001)	-0.123*** (0.002)	-0.151*** (0.001)	-0.126*** (0.001)	-0.065*** (0.002)	-0.169*** (0.002)
LAN	-0.308*** (0.004)	-0.190*** (0.004)	-0.305*** (0.002)	-0.333*** (0.002)	-0.206*** (0.005)	-0.306*** (0.006)
BOR	0.147*** (0.002)	0.183*** (0.002)	-0.282*** (0.001)	0.187*** (0.001)	-0.020*** (0.003)	0.339*** (0.003)
ln MIG	0.002*** (0.000)	-0.006*** (0.000)	0.007*** (0.000)	-0.000 (0.000)	0.005*** (0.000)	-0.004*** (0.000)
Constant	3.183*** (0.014)	2.789*** (0.015)	4.791*** (0.007)	3.484*** (0.007)	1.827*** (0.018)	3.156*** (0.020)
Observations	6,313,949	6,313,949	6,313,949	6,313,949	6,313,949	6,313,949
R-squared	0.021	0.017				

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Column (3), (4) are the quantile regression at 25% and 75% of log package value; Column (5), (6) are the quantile regression at 25% and 75% of log unit price.

disparity in package distribution between high- and low-GDP countries. Migration stock boosts high-value packages in a similar pattern. Apart from the distribution's quasi-scissor disparity, we can observe a U-shaped tendency in the GDP parameters, although they are not significant. That is, the effect of the level of GDP remains neutral for packages of moderate value. Apparently, the quasi-scissor disparity has a flat middle section.

The coefficients of distance in the package value regressions are roughly one-tenth to one-twentieth the magnitudes of the coefficients in the macro-level regressions, implying that the number of packages at the 10th percentile of package value falls 1.5% each time the distance increases 10%. Compared with GDP, the coefficients of distance are coarsely increasing, suggesting that “long distance” can boost the number of low-value packages while staying neutral to the number of high-value packages. Briefly, the greater is the distance, the greater is the number of low-value packages. “Contiguous” borders lift the number of low-value packages in a similar pattern. However, the coefficient on Internet infrastructure is similar to what was found in the macro-level regressions, indicating that Internet popularity can lift the middle percentiles.

The combination of these results seems to indicate that although the common language is effective in mitigating the effect of actual or virtual distance on e-commerce, a “sample bias” persists. Some

Table 9. Distributional estimation of package value in percentile

Package Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	x10	x20	x30	x40	x50	x60	x70	x80	x90
ln GDP	-0.074	-0.039	-0.048	0.009	0.002	0.008	0.034	0.076**	0.106***
	(0.048)	(0.045)	(0.037)	(0.037)	(0.031)	(0.033)	(0.037)	(0.038)	(0.039)
ln DIS	-0.159*	-0.112	-0.162**	-0.121	-0.093	-0.010	0.040	0.056	-0.004
	(0.095)	(0.089)	(0.073)	(0.074)	(0.061)	(0.065)	(0.072)	(0.075)	(0.076)
Net100	0.004	0.000	0.006*	0.007*	0.006*	0.004	0.004	-0.001	-0.002
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
ln POP	0.038	0.014	0.041	0.013	0.012	-0.002	-0.020	-0.060	-0.061
	(0.051)	(0.047)	(0.039)	(0.039)	(0.032)	(0.034)	(0.038)	(0.040)	(0.040)
LAN	-0.039	-0.297	-0.197	-0.180	-0.144	-0.222	-0.324	-0.305	-0.319
	(0.382)	(0.356)	(0.294)	(0.295)	(0.245)	(0.262)	(0.291)	(0.301)	(0.306)
BOR	-0.297	-0.359**	-0.266*	-0.123	-0.140	-0.099	0.013	0.011	0.049
	(0.184)	(0.171)	(0.141)	(0.142)	(0.118)	(0.126)	(0.140)	(0.145)	(0.147)
ln MIG	0.002	0.020	0.021**	0.015	0.008	0.017*	0.025**	0.034***	0.031***
	(0.013)	(0.012)	(0.010)	(0.010)	(0.008)	(0.009)	(0.010)	(0.010)	(0.010)
Constant	2.642**	2.070**	2.597***	1.653**	1.946***	1.444*	0.746	0.496	0.707
	(1.074)	(0.997)	(0.825)	(0.829)	(0.688)	(0.736)	(0.815)	(0.845)	(0.859)
Observations	189	189	189	189	189	189	189	189	189
R-squared	0.047	0.040	0.077	0.118	0.095	0.077	0.122	0.161	0.225

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; x10, x20....x90 represents the 10, 20,90 percentile of item unit price of each country

countries use Chinese daily, but the common language does not bring out enough cultural identity. Therefore, the effect of the stock of migrants eclipses the effect of common language as a determinant.

In Table 10, we regroup our express delivery data by stock, keeping the unit price (SKU price), which reveals the value of each commodity instead of the total price paid per package. The analyses using these data support our previous findings. In the first row, we observe a negative GDP effect that is somewhat lower than that of the macro sample for low-price packages. That is, an increase in GDP lowers the total number of packages at the 20th and 30th percentiles. And higher GDP leads to fewer low-price commodities but more high-price ones. In the second row, the distance effect is mostly negative but insignificant. This effect is attenuated by different measurements of shipping charges. Intuitively, only high-price commodities experience long-distance transport. However, express delivery usually charges transport freight with a fixed rate modified by weight and service. Consequently, long-distance transport goods, especially small ones, do not require high freight charges. We also observe positive and significant effects of migration stock on the 80th and 90th percentiles. A country with more Chinese migrants has a higher proportion of high-price goods, that is, they tend to buy costlier goods from China.

The presented evidence at the micro and distribution levels suggests a crucial difference in the determinants of aggregate-level volume. Furthermore, we can identify the impact on total volume into the micro-level distributional change through the micro-level analysis.

Table 10. Distributional estimation of item price in percentile

Price	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	x10	x20	x30	x40	x50	x60	x70	x80	x90
ln GDP	-0.080	-0.085*	-0.073*	-0.001	0.007	0.056	0.054	0.078*	0.099**
	(0.055)	(0.044)	(0.039)	(0.036)	(0.036)	(0.037)	(0.037)	(0.041)	(0.043)
ln DIS	-0.027	-0.088	-0.039	-0.065	-0.061	-0.056	0.010	0.054	-0.006
	(0.109)	(0.087)	(0.077)	(0.070)	(0.072)	(0.073)	(0.073)	(0.080)	(0.084)
Net100	0.008	0.008*	0.011***	0.010***	0.013***	0.007*	0.005	0.006	0.003
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
ln POP	0.072	0.070	0.094**	0.023	0.015	-0.046	-0.037	-0.050	-0.055
	(0.058)	(0.046)	(0.041)	(0.037)	(0.038)	(0.039)	(0.039)	(0.042)	(0.044)
LAN	0.147	0.121	0.310	0.187	0.164	0.001	-0.152	-0.143	-0.268
	(0.437)	(0.350)	(0.309)	(0.282)	(0.288)	(0.293)	(0.293)	(0.321)	(0.338)
BOR	-0.298	-0.243	-0.167	-0.011	0.003	0.006	0.047	0.145	0.096
	(0.210)	(0.168)	(0.148)	(0.136)	(0.138)	(0.141)	(0.141)	(0.154)	(0.162)
ln MIG	-0.002	0.009	-0.000	0.006	0.004	0.001	0.010	0.023**	0.021*
	(0.015)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)
Constant	0.817	1.780*	1.109	0.953	1.133	1.218	0.707	0.073	0.620
	(1.228)	(0.981)	(0.866)	(0.792)	(0.806)	(0.823)	(0.823)	(0.902)	(0.947)
Observations	188	189	189	189	189	189	189	189	189
R-squared	0.032	0.038	0.070	0.147	0.207	0.152	0.125	0.196	0.181

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; x10, x20....x90 represents the 10, 20,90 percentile of item unit price of each country

6. CONCLUSION

With the booming volume of global cross-border e-commerce, it is gradually expanding as a proportion of the total international trade market. The determinants of e-commerce are of importance for researchers and policymakers (Wang and Wei, 2020).

By utilizing a unique data set on cross-border express deliveries, this paper first studied the determinants of aggregate cross-border volume in a gravity model framework. Our estimations show that the determinants of traditional bulk trade also have significant impacts on the volume of cross-border e-commerce trade. In addition to the traditional determinants in the literature, we also provide evidence of the impact of Internet popularity on the volume of e-commerce trade.

We also investigated the impact of these determinants on micro-level consumer choice. By distinguishing the different effects of package value and item unit price, we find significant differences in the impacts of the determinants of aggregate volume and consumer-level behavior. Our paper contributes to the literature on international trade and e-commerce. Moreover, our findings provide important additions to distinguish the aggregate-level and micro-level differences.

REFERENCES

- Blum, B. S., & Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2), 384–405. doi:10.1016/j.jinteco.2005.10.002
- Cavallo, A., Cruces, G., & Truglia, R. P. (2014). *Inflation Expectations, Learning and Supermarket Prices: Evidence from Field Experiments*. NBER Working Papers.
- Choi, C. (2011). The effect of the Internet on service trade. *Economics Letters*, 109(2), 102–104. doi:10.1016/j.econlet.2010.08.005
- Chung, C. J. (2011). *The geography of global Internet hyperlink networks and cultural content analysis*. University at Buffalo.
- Combes, P.-P., Lafourcade, M., & Mayer, T. (2005). The trade-creating effects of business and social networks: Evidence from France. *Journal of International Economics*, 66(1), 1–29. doi:10.1016/j.jinteco.2004.07.003
- Dellavigna, S., & Gentzkow, M. (2017). *Uniform Pricing in U.S. Retail Chains*. NBER Working Papers.
- Felbermayr, G. J., & Toubal, F. (2010). Cultural proximity and trade. *European Economic Review*, 54(2), 279–293. doi:10.1016/j.eurocorev.2009.06.009
- Freuthnd, C. L., & Weinhold, D. (2004). The effect of the Internet on international trade. *Journal of International Economics*, 62(1), 171–189. doi:10.1016/S0022-1996(03)00059-X
- Gomez-Herrera, E., Martens, B., & Turlea, G. (2014). The drivers and impediments for cross-border e-commerce in the EU. *Information Economics and Policy*, 28(1), 83–96. doi:10.1016/j.infoecopol.2014.05.002
- Gorodnichenko, Y., & Talavera, O. (2017). Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration. *The American Economic Review*, 107(1), 249–282. doi:10.1257/aer.20141127
- Hanson, G., & Xiang, C. (2011). Trade barriers and trade flows with product heterogeneity: An application to US motion picture exports. *Journal of International Economics*, 83(1), 14–26. doi:10.1016/j.jinteco.2010.10.007
- Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (Vol. 4, pp. 131–195). Elsevier.
- Hellmanzik, C., & Schmitz, M. (2015). Virtual proximity and audiovisual services trade. *European Economic Review*, 77, 82–101. doi:10.1016/j.eurocorev.2015.03.014
- Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics*, 123(2), 441–487. doi:10.1162/qjec.2008.123.2.441
- Hortaçsu, A., Martínez-Jerez, F. A., & Douglas, J. (2009). The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre. *American Economic Journal. Microeconomics*, 1(1), 53–74. doi:10.1257/mic.1.1.53
- Kim, T. Y., Dekker, R., & Heij, C. (2017). Cross-Border Electronic Commerce: Distance Effects and Express Delivery in European Union Markets. *International Journal of Electronic Commerce*, 21(2), 184–218. doi:10.1080/10864415.2016.1234283
- Kimura, F., & Lee, H. H. (2006). The Gravity Equation in International Trade in Services. *Review of World Economics*, 142(1), 92–121. doi:10.1007/s10290-006-0058-8
- Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2), 351–363. doi:10.1016/j.jinteco.2014.04.004
- Rabinovich, E., Rungtusanatham, M., & Laseter, T. M. (2008). Physical distribution service performance and Internet retailer margins: The drop-shipping context. *Journal of Operations Management*, 26(6), 767–780. doi:10.1016/j.jom.2008.05.001
- Shi, B. (2016). Internet and International Trade: Empirical Evidence Based on Bilateral and Bidirectional Hyperlinks Data [in Chinese]. *Economic Research Journal*, 5, 172–187.

Silva, J. M. C. S., & Tenreyro, S. (2009). On the existence of the maximum likelihood estimates in Poisson regression. *Economics Letters*, 107(2), 310–312. doi:10.1016/j.econlet.2010.02.020

Tang, Jianmin, Weimin Wang & Zhihao Yu (2020). Globalisation, ICT Investment and Firm Dynamism. *Transnational Corporations Review*, 2 12:2, 193-202, DOI: 10.1080/19186444.2020.1756724

Wang, Y., & Wei, W. (2020). Chinese firms' export dynamics: Experimental but promising. *Transnational Corporations Review*, 12, 1, 10–23. doi:10.1080/19186444.2020.1724483

ENDNOTE

¹ The p-value equals 0.18.