RESEARCH   **Open Access**

Does distance matter in foreign direct investment sub-national location choice? Evidence from China

Yibo Bi[1], Zhiyi Ren[2] and Kun Bao[2*]

* Correspondence: kunbao0223@
163.com
[2]Business School, University of
Nottingham, Nottingham, UK
Full list of author information is
available at the end of the article

Abstract

Multinational enterprises (MNEs) make investment decisions according to the distance factors at a sub-national level. This paper made estimates using the gravity model with provincial foreign direct investment (FDI) data from 2000 to 2012 and employed three concepts of distance. Our empirical results indicate that geographic distance and cultural distance have significant negative effects on FDI flow, whereas economic distance has a significant positive effect. It suggests that FDI prefers to locate in regions that are geographically and culturally close but economically distant from the home country, which further implies that FDI in China is dominated by vertical FDI. Our findings suggest that Chinese provincial governments should place emphasis on attracting FDI from culturally close countries and provide institutional support to encourage and promote horizontal FDI.

Keywords: Multinational enterprises (MNEs), Horizontal FDI, Vertical FDI, Gravity model, Sub-national level, Geographic distance, Cultural distance, Economic distance, China

Introduction

How multinational enterprises (MNEs) make investment decisions has been a dominant topic in the area of international business research over the past few decades. The mainstream of research, focusing on country-specific factors, has influenced the flow, distribution, and entry mode of foreign direct investment (FDI) across the world. It overlooks the effects of relational factors between the source and recipient regions of investment. Recent studies have investigated how the distance between countries impacts the flow of economic activity across borders (Kogut and Singh 1988). However, few studies have investigated the impacts of distance factors at a sub-national level.

A growing number of empirical studies imply that MNEs make equal considerations regarding sub-national locational advantages offered by multi-level local governments when selecting a proper site for investment (Nielsen et al. 2017). Accurate and comprehensive attention paid to sub-national locational characteristics is particularly critical for MNEs that are going to join a large market, due to the tremendous regional disparity in a geographically wide market. That regional disparity includes differences

in language, culture, climate, economic development, and government administration, which are determinative factors in MNEs' location choice.

China is a particularly good case for exploiting the location patterns of FDI at a sub-national level. It has the third-largest territory with the biggest population in the world. Besides the vast territory, China also has a diverse climate and natural environment across various regions. As such, there are differences in economic development, culture and institutions among different regions. Thus, it is not reasonable to measure the distance between the home country of the FDI and China as a whole. In terms of geographic distance, the straight-line distances across China are about 5500 km from the north to the south and about 5200 km from the east to the west. In this case, a particular country may be at distinct distances from different regions in China. For instance, transportation and communications in the eastern region are well-developed, while the middle and western regions have the advantage of being well-endowed with natural resources. Therefore, MNEs face a business environment peculiar to a particular host region in China. The huge regional disparity makes location choice more strategically important for MNEs.

Prior empirical literature on sub-national location choice of FDI in China used aggregated FDI inflow data, which failed to reveal a difference in location choices of various source countries (Cheng 2000). However, some empirical studies found that location choice strategy differs among home countries and is changing over time (Pan 1997). Additionally, according to the eclectic OLI (O-ownership; L-location; I-international) theory, MNEs' ownership advantages highly depend on country-specific characteristics, but crucially determine the destination of foreign investments (Dunning 1980). Thus, MNEs who source from different countries are supposed to have various considerations and standards regarding location choice, since they have distinct ownership advantages. Therefore, we want to examine FDI location choice, considering the characteristics of both home and host regions. Specifically, we focus on the distance and linkage between the home country and host destination.

This study uses novel FDI data, sorted by home country at a regional level, from 2000 to 2012. Distance is introduced as an independent variable, which is widely used in national-level studies but rarely in sub-national location choice studies. This paper contributes to the literature in three distinct aspects. First, although recent studies found the distance between countries to be significant on economic activities across countries, they did not consider the influence of distance on sub-national level investments. We aim to fill this gap by estimating a gravity model with distance proxies between the FDI's home country and host region in China. Second, differing from FDI location choice studies in China, we estimate a gravity model with a novel dependent variable, FDI flow by home country and host region over time, to explore the effects of both home and host regions, rather than just a unilateral effect. Finally, the distance factors are considered from three perspectives. Besides geographic distance and economic distance, cultural distance is also taken into account. The empirical results imply that the gravity model can explain FDI flow in China; FDI increases with market size on both sides of the investment but decreases with geographic distance, and FDI prefers to locate in a culturally close but economically dissimilar region. In addition, economic development in both host and home regions and the research and development (R&D) level in the home country are also contributing factors to FDI flow.

The rest of the paper is organized as follows: The next section reviews relevant theoretical literature and proposes the hypotheses. Section 3 presents the methodology, including the data description, measurement of variables and empirical strategy. Section 4 presents the main results and various robustness tests. Section 5 shows the discussion of the results. Section 6 is the final section with a conclusion.

Theory and hypotheses

Determinants of FDI: the eclectic paradigm

The eclectic paradigm or OLI theory, developed by Dunning (1980), is widely used to explain international production or FDI, which embraces both internalization theory and trade theory. Dunning (1980) proclaimed that FDI is preferred when there are simultaneous ownership advantages (O), location advantages (L), and internalization advantages (I). To be specific, if a firm chooses to become an international enterprise, it must have more ownership advantages than domestic competitors in the host country, and those ownership advantages must significantly outweigh the risks and costs of servicing an unfamiliar market (Hirsch 1976). Besides the ownership advantages, an enterprise that invests abroad has to be able to afford the cost of transport and acquiring information due to geographical distance and cultural differences.

Mathews (2006) extended the eclectic paradigm from the OLI to an alternative LLL (linkage, leverage, and learning). In this framework, the important roles of network and interfirm linkage in FDI are introduced. In practice, the new/latecomer prefers a joint venture rather than a wholly-owned business, to obtain cooperation with a host partner and access local knowledge and networks. Based on the LLL perspective, Buckley and Casson (1998) also incorporated the linkage between host and home countries into the eclectic paradigm as an institutional asset. They reconstructed the eclectic paradigm theory with the idea that the determinants of international business activities can be classified by country-specific ownership advantages (CSAs), host country-specific location advantages, and the distance and linkage between the home and host countries. CSAs often significantly relate to outward FDI, since institutions and the macroeconomic environment in the home country can be transferred to firms' competitive advantages in internationalized processes. Host country-specific location advantages refer to the host country's locational attractions. Cross-country distance or linkage captures the interactive effect of a pair of home and host countries, which impacts the flow between them. Moreover, Buckley and Casson (1998) also stressed that linkage variables improve the explanatory power of the eclectic paradigm and are shown to be an adjunct, not an alternative.

Vertical FDI and horizontal FDI

Generally, an enterprise has several possible approaches for achieving growth. It can develop horizontally, which means creating new product lines, or vertically, into new activities, such as supplying or after-sales service. Also, it can expand market share by acquiring existing enterprises or exploiting foreign markets. Regarding motivations for going abroad, FDI can be categorised into vertical FDI and horizontal FDI. Considering vertical FDI, for instance, exporting accounts form a substantial part of the total cost. Assuming the inputs of manufacturing are the same in all countries, exporting

generates transport, insurance, communication, and exchange costs and trade barriers (tariffs and entry and license fees). Horizontal FDI sells products on the host country market instead of exporting, whereby the costs of geographical distance are lower. However, they need to pay the costs of adapting products to the local market and training local workers to use the MNE's technology. Apparently, these costs are highly related to linkage variables.

According to Hymer (1976), cross-national dissimilarity and distance could lead to liability of foreignness (LOF), which poses challenges to the lasting return and profit of MNEs. LOF was categorised into four aspects: costs due to spatial distance (travel, transport, and coordination), unfamiliarity with the local environment, differential treatment by the host country, and costs imposed by the environment of the home country (Zaheer 1995). From the perspective of business management, a concept that corresponds to LOF is the costs of doing business abroad (CDBA). Hymer (1976) believed that foreign firms would afford the CDBA because of the disadvantages and barriers regarding integration into the host market. MNEs may have unfamiliar hazards which are caused by the information disadvantage, requiring MNEs to spend additional time and money on obtaining knowledge to put them at a comparable level with local firms. Moreover, MNEs may suffer informal differential treatment if they are viewed as an outsider. This is possibly caused by environmental hazards (Choi et al. 2016; Henisz and Williamson 1999) or the unfamiliarity of foreign firms.

Besides geographic distance, economic and social costs could also be generated by cultural distance. Greater cultural distance leads to more unfamiliarity and relational hazards as consequences of rising LOF (Eden and Miller 2004). Scholars have found evidence that firms operating in culturally dissimilar countries have more difficulties transferring management routines and capabilities (Kostova and Roth 2002; Kostova and Zaheer 1999). Therefore, geographic and culture distance causes LOF to MNEs during international expansion. MNEs need to utilise their advantages or employ entry mode strategies to minimise the potential costs. Meanwhile, the effects of distance on LOF may also leverage on the location choice of FDI. In a specific range of choice, an MNE may choose a host region with less LOF over others. In other words, having a shorter distance from the home country is a competitive advantage for one region in attracting FDI.

Hypotheses

Market size is important for a firm's short-term and long-term sales and profits (Makino et al. 2004). For market-seeking FDI, a larger market size provides more opportunities to maximise the effects of economies of scale by supplying products and services to a new market (Dunning 1980). For efficiency-seeking FDI, many business activities are scale-sensitive, for instance, production, distribution, and per-unit costs. Hence, aiming at sustainable growth and future profit, firms are more likely to choose a host region with a larger market size. Given the huge disparity among regions in China and the findings of previous empirical studies (e.g., Aleksynska and Havrylchyk 2013; Bailey and Li 2015), a host region in China with a greater market size should receive more foreign investment compared to a province with a smaller market size.

Bi *et al. Frontiers of Business Research in China*   (2020) 14:12

Page 5 of 19

Differing from trade studies, the influence of spatial distance in FDI studies is not predictable. In terms of international business, long geographic distance represents high transportation costs for sending staff and shipping goods. Those high transportation costs seem to have opposite effects on vertical and horizontal FDI. For vertical FDI, costly shipping is a hindrance, since vertical FDI needs to export its products back to the home country and not solely sell them in the host market. However, for horizontal FDI, long geographic distance gives MNEs more motivation to conduct FDI rather than export to the foreign market, which demonstrates that FDI flow between a pair of regions separated by a longer geographic distance should be greater than with those separated by a shorter distance. Given the vast territory and complex terrain in China, the transportation costs for MNEs are undoubtedly huge. In 2012, FDI in the manufacturing industry in China still accounted for 43.74% of the total and took the largest share in the total 19 industries, which implies that vertical FDI dominates in China rather than horizontal FDI. Therefore, this paper proposes the first hypothesis:

> *Hypothesis 1: MNEs prefer regions with shorter geographic distance and larger market size regarding FDI location choice.*

Many studies have demonstrated the importance of cultural similarities for international business (Chen et al. 2017; Kuang and Wang 2018). According to Du et al.'s (2012) argument, FDI from a culturally closer country could effectively turn government intervention into help. A historical relationship creates a backdrop that is shared by the home and host region pair, and this shared history may make it more likely that the home invests in the host regions (Garrett 2016). In addition, Jones and Khanna (2006) argued that a history of business contact helps match the international buyers and sellers since foreign firms have social network resources in a host region sharing past business relations. Some regions in China had close business relationship with foreign countries in history, increasing the mutual understanding of each other's culture and institutions. Thus, the existence of a business history between a pair of home and host regions is a significant increasing factor of FDI. This paper uses the historical business tie to proxy the cultural distance and proposes the second hypothesis:

> *Hypothesis 2: MNEs prefer a culturally closer host region rather than a culturally remote region regarding FDI location choice.*

Empirical results suggest that regions with a more advanced economic level, higher income, and better infrastructure receive and conduct more FDI, which implies the importance of the demanding factor for attracting inward FDI (Belderbos et al. 2011). From the perspective of an investment supplier, a firm sourcing from a mature economic environment has abundant business experience of marketing and managing. That experience makes MNEs more competitive in a foreign economy compared to local and other foreign firms.

Although the positive relationship between economic development and FDI is confirmed by many studies, there is no consensus on the impact of the economic similarity

between home and host regions. Horizontal FDI produces services for both home and host regions and tends to be located in similarly endowed economies. However, for vertical FDI, a region with dissimilar economies is preferred, since headquarter services are completely separated from production to exploit factor cost differentials. Given that vertical FDI dominates in China, this paper proposes the third hypothesis:

> *Hypothesis 3: Regions with higher economic development receive and conduct more FDI, but MNEs in China prefer to invest in a dissimilar economy.*

In summary, we hypothesize that the promoting factors for FDI flow into China's various host regions are market size, GDP per capita of both home and host regions, and economic distance. The resistance factors include geographic distance and culture distance. Figure 1 illustrates the effects of distance factors on FDI inflow to regions in China.

Methodology

Data and sampling

The data of FDI inflow are only provided by some provincial governments in China. The dataset we used covers 20 of 31 Chinese provincial-level regions and the period from 2001 to 2012, though there are a few missing years for Shanxi, Shanghai, Hunan, Hainan, Ningxia, and Xinjiang.

The sample we chose is fairly representative. Although this dataset do not include all the host regions or investment sourcing countries, the primary contributors to China's FDI inflow are included in this sample, from the perspective of foreign investing countries. Figure 2 represents the percentages of total FDI in China from the sample countries from 2001 to 2012. These countries contributed over 80% of China's FDI in most of the years and over 70% in 2005, 2006, 2007, and 2008.

According to the classification of the Report on Foreign Investment in China, distribution of FDI can be divided into the east, central, and west areas. The sample includes 9 of 11 provincial regions in the east area, 6 of 8 regions in the central area, and 5 of
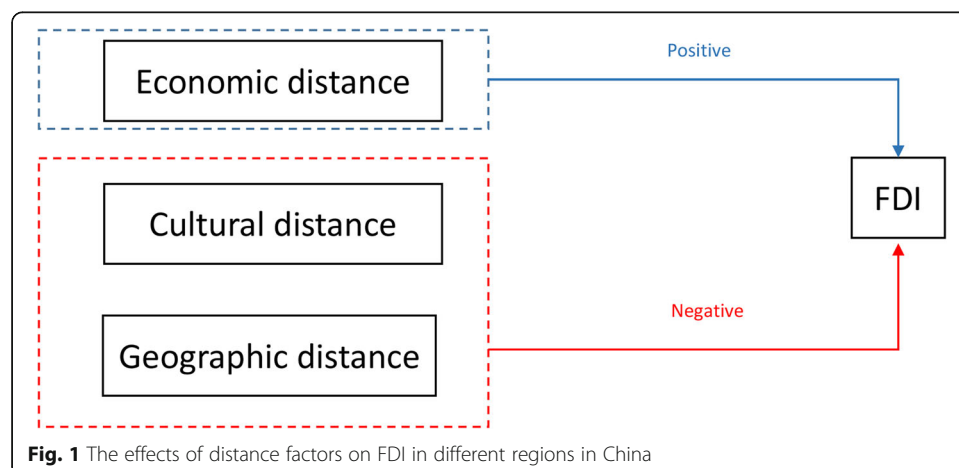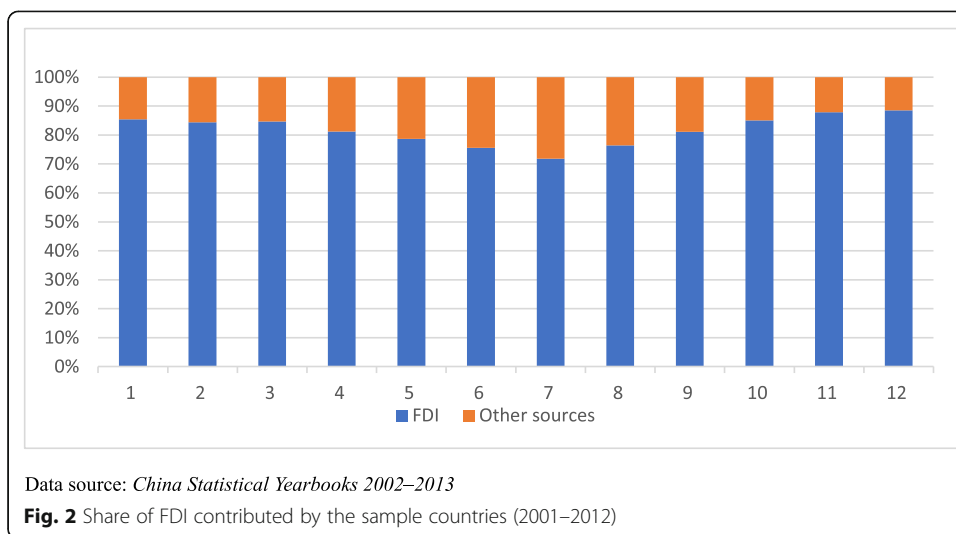

Fig. 1 The effects of distance factors on FDI in different regions in China

Data source: *China Statistical Yearbooks 2002–2013*

Fig. 2 Share of FDI contributed by the sample countries (2001–2012)

12 regions in the west area. Hence, from the perspective of host regions, the sample of 20 host regions in China is representative of FDI destinations.

All data are deflated using 2010 as the base year and employing China's deflator from the World Bank for the host region variable. Table 1 is a complete description of all variable definitions and sources. The raw data of home countries, including GDP, GDP per capita, population, and R&D, were collected from the

Table 1 Variable description and source

Variable	Description	Source
Dependent variable		
FDI	Logarithm of annual FDI inflow from a home country to a host province (USD10,000)	*China Statistical Yearbook*
Independent variables of home countries		
GDP_pc	Logarithm of GDP per capita (constant 2010 USD)	World Bank
POP	Logarithm of midyear estimates of all residents including legal status and citizenship	World Bank
RD	Medium- and high-tech industry (% manufacturing value added)	United Nations Industrial Development Organization,
		Competitive Industrial Performance database
Independent variables of host regions		
GDP_pc	Logarithm of real GDP per capita (constant 2010 RMB)	*China Statistical Yearbook*
POP	Logarithm of total population at the end of year	*China Statistical Yearbook*
BA	Takes the value of 1 if the region has duty free port, and 0 otherwise	Manual collections by the author
Linkage variables		
DIST	Logarithm of direct distance between the capital cities of pair regions	Google maps, and calculations by the author
CULT	Takes the value of 1 if the region ever had a business contact with foreign firms from 1840 to 1978, and 0 otherwise	Manual collections by the author
SIMI	An index measuring the economic similarity of a pair of regions	Calculated by the author

World Bank database. *China Statistical Yearbook* (various years) provided data relating to host regions, including GDP, GDP per capita, and population. The dependent variable *FDI* was collected from *China Statistical Yearbook* for various provinces separately. There are 5312 observations in our sample. Table 2 shows descriptive statistics for each variable.

Measurement of variables

In the studies of FDI location choice, FDI stock and FDI flow are two commonly used measurements of FDI flow. According to a survey paper conducted by Nielsen et al. (2017), who reviewed 153 quantitative studies on FDI location choice, there are 94 studies exploring FDI flow and 59 studies focusing on FDI stocks. Flow data could directly reflect how FDI accesses a specific region at a given time, and stock data could reveal the relationship between the FDI location and local characteristics, but not the entrance decision itself (Galan et al. 2007). For this study, we use FDI annual inflow as the dependent variable.

To measure geographic distance, we used the distance between the capital cities of the home and host regions in keeping with previous literature (e.g., Portes and Rey 2005). It is measured by the direct distance between the capital cities of pair regions according to the google maps. A great geographic distance creates high transporting costs, which encourages firms to conduct direct investment rather than export. However, greater distance is usually associated with unfamiliar language and culture, and higher transaction costs, which may detract from the FDI activities. Therefore, the effect of geographic distance on FDI is inconclusive.

To measure cultural similarity, previous studies have commonly used linguistic or religious proximity to gauge the cultural distance (Kuang and Wang 2018). However, these country-specific linkages are not suitable for a sub-national level study. In this study, a historical business tie between a home country and a host region is applied to capture cultural distance. We measure historical business tie by identifying if the pair regions have ever had a business link from 1840 to 1978. China started implementing reform and opening up as a national policy in 1978. Before then the foreign business link was unbalanced among sub-national regions. We take the value of 1 if the pair regions have ever had a historical business tie from year 1840 to 1978, and 0 otherwise,

Table 2 Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
$\ln(FDI + 1)$	5312	9.039	5.996	0	20.115
$\ln POP_{it-1}$	5003	9.715	2.145	3.027	16.919
$\ln POP_{jt-1}$	5312	8.294	0.625	6.438	9.268
$\ln GDP\_pc_{it-1}$	4884	10.058	1.059	5.602	11.626
$\ln GDP\_pc_{jt-1}$	5312	9.937	0.583	8.882	11.274
$\ln DIST_{ij}$	5312	8.567	0.835	4.780	9.781
$CULT_{ij}$	5312	0.035	0.183	0	1
$SIMI_{ijt-1}$	4831	0.223	0.161	0.001	0.500
RD_{it-1}	4643	0.390	0.186	0	0.880
BA_{jt-1}	5312	0.351	0.477	0	1

indicating the cultural distance since the historical business tie makes home and host regions culturally closer.

According to the research by Egger and Pfaffermayr (2004), the economic similarity is calculated by the following equation:

$$SIMI_{ijt} = 1 - \left| \frac{GDP_{it}}{(GDP_{it} + GDP_{jt})} \right|^2 - \left| \frac{GDP_{jt}}{(GDP_{it} + GDP_{jt})} \right|^2.$$

The effect of economic similarity on FDI inflow could be positive or negative. Theoretically, horizontal MNEs tend to be found in similarly endowed economies; however, vertical MNEs tend to occur in dissimilar economies. MNEs, who mainly offer services locally, in the host market, and which rely on economies of experience, scale, and standardisation, will prefer regions that have similar economic profiles, as they have to replicate their mature business model to obtain a competitive advantage, which is hard to achieve in an economy where customer incomes are very different. However, for particular industries, competitive advantage comes from economic arbitrage, the exploitation of cost and price differentials between markets.

Technology flow plays an important role as a driver of FDI flow in China. As a late developer, China imported technology from advanced countries, which have become significant sources of know-how and investment. Technology flow exists with both horizontal and vertical FDI, but vertical FDI is more willing to transfer technology, since it results in production facilities that are tailored to the production environment of the host economy. As countries upgrade their technology, they obtain more competitive advantages in technological products and exports. Firms in technologically leading countries drive this pattern through technology flow to next stage producers, mainly by FDI. Their motivation is to utilise their technological assets as much as possible after their domestic production advantage fades away. Thus, it is more feasible that FDI in China is driven by technology flow than by economies of scale or factor price differences. In this paper, R&D levels in the home country are measured as the sum of medium- and high-tech industry over manufacturing value-added.

A bonded area is a special economic zone inside the territory while outside the customs. It provides storage, export processing, and re-exporting services. Although the details of the preferential policy differ across various regions, a primary rule is that foreign goods imported from abroad into the bonded area are duty-free. To capture the attractive aspect of this policy for FDI, we generate a variable *BA* taking the value of 1 if a region has a bonded area, otherwise, it is 0.

Estimation approach

The empirical framework is estimated based on the traditional gravity model, which comes from Newton's law of gravitation, declaring that attractive force increases with the masses of two objects but reduces with the distance between them. It was widely used to explain economic flow, such as international trade and FDI (Anderson and van Wincoop 2003; Eaton and Kortum 2002; Feenstra 2015; McCallum 1995; Ravenstein 1885; Redding and Venables 2004; Savage and Deutsch

1960; Tinbergen and Hekscher 1962; Trefler 2002). The set-up of the gravity model can be interpreted that the potential flow between the supplied part i and destination $j(Y_{ij})$ is positively determined by the masses of goods or labour or other factors in i and j (X_iX_j), but negatively determined by the distance D_{ij}, which is shown as follows:

$$Y_{ij} = X_iX_j/D_{ij.}$$

Since 1961, the gravity model has been widely applied in empirical studies on trade flows and FDI (Anderson and van Wincoop 2003). Although the gravity model was initially implemented to study bilateral flow, it is also applicable for one-directional flow data because it can successfully explain how the interaction between the source and destination impacts the flow, and also capture the effects of source and destination variables. Therefore, the gravity model is applied in many empirical studies on one-directional FDI flow data. For instance, to investigate the effects of instability in home and host countries on FDI flow, the gravity model is employed by Chenaf-Nicet and Rougier (2016) to estimate the FDI flowing from Europe and the Mediterranean region to the four main recipients of FDI in the Middle East and North Africa from 1985 to 2009. In this paper, we also adopt the gravity model to explore the determinants of the FDI inflow to different provinces in China, with a special focus on distance.

Population and GDP are both classic variables for measuring masses of home and host regions in the gravity model in FDI studies. Previous literature has found that regional population is more representative than regional economic development in explaining the possibility of being a recipient of foreign capital (Flores and Aguilera 2007). Thus, this study uses population to measure the market sizes of home and host regions (Cuervo-Cazurra 2006; Henisz 2000). As suggested by the gravity model, the masses of both source and destination markets are promoting factors for bilateral flow. GDP per capita of the supplier and demander are often estimated in a gravity equation. The assumption behind this is that higher-income countries perform more expansion activities in foreign markets, since a higher income level implies a more advanced infrastructure, a developed legal system, and abundant human resources, which are all pulling factors for investment.

Focusing on distance factors, this study uses a pooled cross-sectional regression with various control variables based on the gravity model. The regression equation is as follows:

$$FDI_{ijt} = \beta_1DIST_{ij} + \beta_2CULT_{ij} + \beta_3SIMI_{ijt-1} + \beta_4X_{ijt-1} + \gamma_t + \varepsilon_{ijt,}$$

where i denotes home countries, j denotes host regions, and γ_t is the time effect. FDI_{ijt} symbolises the FDI flow from the home country i to the host region j in year t. $DIST_{ij}$ refers to the geographic distance between home countries and host regions. $CULT_{ij}$ is a dummy variable with the value 1 if the region ever had a business contact with foreign firms from 1840 to 1978, and 0 otherwise. $SIMI_{ijt-1}$ is a proxy that measures the similarity of the economies in the home country and host region. X presents other controlling variables, including population, GDP per capita, R&D, and bonded area. All independent variables are one-year lagged, since

the investment decisions made by MNEs are based on the figures in the previous year.

Although the gravity model is widely used to predict bilateral flow in empirical studies and also provides stable and significant evidence, there are several unsolved issues related to its empirical application. For instance, the unobserved heterogeneity and existence of zero flow in the dependent variable could reduce the feasibility of the logarithm estimation of the gravity model. Some previous literature introduced truncating and censoring estimates to deal with zero flow. However, the loss of information from zero observations leads to possible selection bias, since zero is usually not randomly distributed in trade (Linders and de Groot 2006; Westerlund and Wilhelmsson 2011).

Besides using a truncated OLS estimate, an alternative to deal with zero observations is to add a particular number onto the dependent variable, such as taking a natural logarithm of ($FDI + 1$). Although it has drawbacks where biased estimates could be generated by the increase in the dependent variable value, it is widely applied in empirical studies because of its success in maintaining the information from zero observations. In this paper, we adopt this conduction to treat the dependent variable.

Researchers have estimated the gravity model on panel data frameworks, aiming to investigate specific time or country effects (Egger and Pfaffermayr 2004). Fixed effects and random effects are two main techniques. For this research, although the fixed effects estimate is able to control unobserved heterogeneity, the random effects model is preferred since the fixed effects model drops some key constant regressors, such as distance and border variables.

In addition, Silva and Tenreyro (2006) proclaimed that the parameters of log-linearised model estimated by OLS could be highly misleading in the presence of heteroskedasticity. They argued that the gravity equation should be estimated in its multiplicative form and with proposed Poisson pseudo-maximum-likelihood (PPML). Hence, the dependent variable is introduced in level value but not log value, which successfully captures the information from zero observations. Although a PPML estimate was accepted by some researchers in empirical studies (e.g., Westerlund and Wilhelmsson 2011), it has to be mentioned that PPML estimates may present limited dependent variable bias when a significant part of the observations are censored.

The Heckman sample selection model is also widely used in gravity equation. It is a two-step estimation method. In the first step, a Probit equation is estimated to investigate whether flow occurs between the supplier and demander. In the second step, the expected values of the flow are estimated, based on an OLS estimate. The Heckman sample selection model requires exclusion variables that only determine the decision process. The disadvantage of the Heckman sample selection model is that the occurrence of the flow is not modelled independently from the decision of the volume of the flow.

Aiming to deal with zero observations, Tobit or censored regression is another alternative estimate for the gravity model. However, compared with the Heckman sample selection estimate, the disadvantage of employing Tobit or censored regression is that the probability of censoring and the value of observed outcome are determined by the same set of variables.

Bi *et al. Frontiers of Business Research in China*   (2020) 14:12

Page 12 of 19

Table 3 Results on various estimates

Y: $\ln(FDI_{ijt} + 1)$	OLS (Pooled)	OLS (Random effect)	Tobit	Heckman	PPML
$\ln DIST_{ijt}$	−3.032*** (−28.64)	−3.027*** (−11.26)	−3.776*** (−10.28)	−3.039*** (−29.23)	−3.057*** (−28.98)
$CULT_{ij}$	1.387*** (5.23)	2.282* (2.35)	2.480 (1.88)	1.799*** (6.53)	1.611*** (5.83)
$SIMI_{ijt-1}$	−4.159*** (−5.91)	−6.635*** (−6.08)	−9.202*** (−6.03)	−4.944*** (−7.53)	−5.250*** (−7.71)
$\ln POP_{it-1}$	0.549*** (8.52)	0.313* (2.43)	0.381* (2.16)	0.126** (7.77)	0.476*** (7.63)
$\ln POP_{jt-1}$	2.519*** (17.02)	2.267*** (8.19)	3.074*** (8.03)	2.383*** (17.56)	2.246*** (15.45)
$\ln GDP\_pc_{it-1}$	1.788*** (15.84)	1.337*** (5.91)	1.631*** (5.23)	1.694*** (15.34)	1.681*** (15.19)
$\ln GDP\_pc_{jt-1}$	3.606*** (17.87)	1.792*** (9.66)	2.379*** (9.25)	2.846*** (21.29)	2.452*** (14.37)
RD_{it-1}	3.386*** (6.77)	4.084*** (3.84)	5.116*** (3.51)	3.631*** (7.29)	3.724*** (7.46)
BA_{jt-1}	0.292 (1.30)	0.157 (0.66)	0.180 (0.54)	0.434*** (4.46)	0.793*** (3.65)
Constant	−46.22*** (−14.22)	−18.84*** (−4.86)	−29.56*** (−5.54)	−35.54*** (−16.55)	−30.35*** (−11.46)
Observation	4351	4351	4351	4351	4351

Notes. The *t* statistics are in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Therefore, each model has its advantages and drawbacks. Besides the pooled cross-sectional estimation, this study selects various models to estimate the gravity equation, including random effect, Tobit model, Heckman sample selection model, and PPML, to obtain a robust and reliable conclusion.

Empirical results

Main results

In Table 3, various models are employed. The first column presents the empirical results of the pooled cross-sectional OLS estimates with the dependent variable $\ln(FDI_{ijt} + 1)$. The second column reports the results from the random effect OLS estimates. The third column reports the panel-data regression results employing the Tobit model. The fourth column shows the results on the Heckman sample selection estimates. The last column presents the empirical results on the PPML estimates, with the form of the dependent variable reshaped.

Our key independent variables are distance factors. The coefficients of geographic distance ($DIST$) are statistically significant at the 1% level across all five models. The signs of the coefficients of geographic distance are the same, and the sizes of them are very similar. Since the dependent variable is $\ln(FDI_{ijt} + 1)$ in the regression model, the value of the coefficient cannot be interpreted directly as the amount to which the independent variables contribute to FDI flow. However, the results demonstrate a significant negative effect of the geographic distance from the home country on the provincial FDI in China, indicating that MNEs prefer to choose the location of FDI which is geographically close to their home country. On the other hand, the mass variable measured by population is significantly positive at the 1% level across five models, indicating FDI flow grows with market size of the province in China. These results confirm our Hypothesis 1.

Referring to cultural distance, the variable $CULT$ is confirmed to have a significant and positive impact on FDI flow. Except for the coefficients from the Tobit model, the results are statistically significant across all the columns. That implies that if the home and host regions share a historical business link, the FDI flow between them is greater than between a pair of regions without a history of business link. These results confirm our Hypothesis 2.

Regarding the economic distance, the coefficients of $SIMI$ are statistically significant at the 1% level across all five models. Although we observe that the sizes of the coefficients of economic similarity varies, the signs of them remain the same in all columns. The results demonstrate a significant negative effect of economic similarity, indicating that MNEs prefer to choose a location of FDI that is economically distant to their home country. On the other hand, positive effects of GDP per capita in the home countries and the host regions are observed in all columns, indicating FDI flow grows with the economic development in both host and home regions. These results confirm our Hypothesis 3.

RD is the control variable, and its coefficients are statistically significant at the 1% level across all five models. Although the sizes are different, the signs of the coefficients remain positive in all columns, which demonstrates that an investing country with a higher level of R&D conducts more outward FDI. Another controlling

Table 4 Results using GDP as an alternative measurement of market size

Y: $\ln(FDI_{ijt} + 1)$	OLS (Pooled)	OLS (Random effect)	Tobit	Heckman	PPML
$\ln DIST_{ijt}$	1.212*** (13.87)	0.947*** (4.46)	1.159*** (4.08)	0.987*** (10.82)	0.698*** (11.32)
$CULT_{ij}$	1.320*** (5.30)	2.481* (2.32)	2.701 (1.91)	1.185*** (4.69)	0.260 (1.26)
$SIMI_{ijt-1}$	−5.171*** (−6.83)	−7.528*** (−6.03)	−10.42*** (−6.02)	−0.445 (−1.19)	7.398*** (6.98)
$\ln GDP_{it-1}$	0.502*** (7.45)	0.230 (1.62)	0.275 (1.45)	0.398*** (5.67)	0.638*** (7.51)
$\ln GDP_{jt-1}$	0.882*** (5.35)	0.878*** (3.60)	1.207*** (3.63)	−0.100 (−0.86)	−0.947*** (−5.68)
$\ln GDP\_pc_{it-1}$	2.007*** (10.12)	1.147*** (4.53)	1.493* (4.33)	1.646*** (14.61)	1.520*** (9.94)
$\ln GDP\_pc_{jt-1}$	−3.136*** (−30.08)	−3.037*** (−10.24)	−3.794*** (−9.65)	−2.021*** (−12.95)	−0.971*** (−19.81)
RD_{it-1}	3.560*** (6.97)	4.198*** (3.68)	5.296** (3.44)	1.534*** (6.00)	−1.152** (−3.14)
BA_{jt-1}	1.159* (4.85)	0.129 (0.53)	0.199 (0.59)	0.669*** (6.21)	0.319 (1.00)
Constant	−17.42*** (−5.98)	−1.29 (−0.38)	−6.276 (−1.45)	−12.65*** (−5.88)	2.574 (1.34)
Observation	4351	4351	4351	4427	4351

Notes. The *t* statistics are in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 Results of quantile regression based on OLS estimates

Y: ln(FDI_{ijt} + 1)	Q(10%)	Q(25%)	Q(50%)	Q(75%)	Q(90%)	Q(99%)
$lnDIST_{ijt}$	0.542*** (4.66)	0.642*** (7.98)	0.695*** (11.95)	0.763*** (12.66)	0.863*** (11.68)	0.991*** (7.19)
$CULT_{ij}$	1.890*** (5.18)	0.913*** (3.62)	0.483** (2.65)	−0.0387 (−0.20)	−0.420 (−1.81)	−0.546 (−1.26)
$SIMI_{ijt-1}$	−3.280*** (−4.72)	−2.375*** (−4.94)	−2.078*** (−5.98)	−2.067*** (−5.74)	−1.834** (−4.15)	−1.253 (−1.52)
$lnPOP_{it-1}$	−0.017 (−0.26)	0.034 (0.75)	0.088** (2.72)	0.082* (2.45)	0.068 (1.65)	0.000 (0.00)
$lnPOP_{jt-1}$	0.861*** (5.86)	0.448*** (4.41)	0.148* (2.01)	−0.0315 (−0.41)	−0.155 (−1.66)	−0.323 (−1.86)
$lnGDP\_pc_{it-1}$	1.488*** (8.43)	1.596*** (13.06)	1.273*** (14.43)	1.204*** (13.16)	1.151* (10.26)	1.385*** (6.62)
$lnGDP\_pc_{jt-1}$	−1.200*** (−10.71)	−1.148*** (−14.79)	−1.364*** (−24.33)	−1.649*** (−28.40)	−1.752*** (−24.58)	−1.502*** (−11.30)
RD_{it-1}	2.948*** (5.60)	2.301*** (6.32)	2.210*** (8.40)	1.096*** (4.02)	−0.134 (−0.40)	−0.497** (−0.80)
BA_{jt-1}	0.214 (0.99)	0.330* (2.22)	0.624*** (5.80)	0.683*** (6.13)	0.861*** (6.29)	0.809** (3.17)
Observation	3109	3109	3109	3109	3109	3109

Notes. The *t* statistics are in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Bi *et al. Frontiers of Business Research in China*   (2020) 14:12

Page 15 of 19

variable *BA* indicates whether a region has a bonded area. There is a positive relationship between a bonded area and FDI, whereas less than half of the models show statistically significant results. The conclusion that a host region having a bonded area is more attractive for inward FDI is not fully established.

Robustness check

To confirm the robustness of the empirical results, we used GDP as an alternative measure of market size. The results in Table 4 are consistent with the main results. GDP in the home and host countries are the economy mass variables capturing the market size in alternative gravity regression. The coefficients of GDP in the home country remain positive; however, they are not statistically different from zero in the OLS or Tobit random effect. GDP in the host region shows a positive effect on inward FDI in most regressions, except for the Heckman estimates. The *CULT* variable is also found to have significant and positive coefficients in most columns except for Tobit and PPML regressions. The coefficients of economic similarity are significantly positive in all regressions except for the Heckman sample selection model. The results of *RD* in the home country are consistent with the main regressions. *BA* is positive in all regressions but only statistically significant in less than half of the models. Other independent variables, including GDP per capita in the home and host regions and geographic distance, show consistent results with the main regressions.

Table 5 reports the results of quantile regression based on OLS, which is the most widely used estimate in empirical gravity model studies. However, standard linear regression technique reflects the average effect of each regressor and the outcome variable based on the conditional mean function $E(y|x)$, assuming that this function is normal and has symmetrical distribution. Paniagua and Sapena (2014) found that OLS provides only a partial view of FDI, since large shares of the world's overall FDI are occupied by a small number of firms. In this case, Quantile regression is suitable to employ when the data concentrate at different levels. For this study, given that the majority of FDI flow to China is contributed from a limited number of countries, we applied a quantile regression to check the robustness of the main results.

Overall, the empirical results of quantile regression perform well and are consistent with the main results. Most of the variables are statistically significant, with the expected signs that vary with the quantiles. The results are presented by the 10%, 25%, 50%, 75%, 90%, and 99% quantiles. The effect of population in the host region decreases with the quantile, which means that market size is less important for foreign investment compared to smaller investment. GDP per capita in the host region, geographic distance, culture distance, and R&D in the home country follow the same trend. In addition, the coefficients of population, culture distance, economic similarity, and R&D are not statistically significant for large quantiles, indicating that the effects of these variables are not significant for large investments. The bonded area dummy variable shows the clearest upward trend, and it is positive in all quantiles.

Discussion

Besides the importance of the location choice of FDI at a national level, MNEs also make considerations regarding sub-national locational advantages offered by multi-level

local governments when selecting a proper site for investment (Nielsen et al. 2017). This is especially true for a country such as China, whose market is geographically wide and where the regional disparity is tremendous. That regional disparity includes differences in language, culture, climate, economic development, and government administration, which are all determinant factors for MNEs' location choice. However, the previous literature relating to sub-national level location choice of MNEs mainly focuses on the effects of some locational specific factors in the host region but overlooks the impact of relational factors between the source and recipient regions of investment. Thus, this study has first addressed the impact of distance on sub-national level location choice of FDI in China, which is missing in the previous literature.

Despite the slight differences, most of the empirical results in this study are shown as expected, and the results are consistent with our hypotheses. First, geographic distance significantly reduces FDI flow. Combining the positive effects of the mass variables, the negative effect of geographic distance confirms the existence of the gravity effect on FDI location choice in China. Second, economic similarity is significant and negatively related to FDI flow in all models. Together with the results of geographic distance, the preference of MNEs for an economically dissimilar region implies that vertical FDI dominates in China. Third, the significant and positive effect of *CULT* dummy variable shows that cultural distance impacts FDI location choice in China. Finally, the signs of others in the traditional gravity model are significant in most regressions, which is in accordance with past literature. GDP per capita in both home and host regions has significant positive effects on FDI flow. By comparing the coefficients of GDP per capita in the home country and host region, the effect of GDP per capita in the host region is stronger than in the home country, which implies that the pulling factors for inward FDI are stronger than the pushing factors for outward FDI. This is also confirmed by the coefficients of population, which are much greater in the host region than in the home country. The sign of the R&D variable in the home country is significantly positive, which means that the level of R&D in the home country is a pushing factor for outward FDI. The dummy variable *BA* positively relates to FDI flow but is not statistically significant, showing the impact of a bonded area on FDI flow is not confirmed.

Conclusion

This paper made estimates based on the gravity model with provincial FDI flow data from 2000 to 2012 and employed three concepts of distance: geographic distance, cultural distance, and economic distance. The regression results indicate that geographic distance and cultural distance have significant negative effects on FDI flow, whereas economic distance has a significant positive effect. That means that FDI prefers to locate in regions that are geographically and culturally close but economically distant from the home country, which further implies that FDI in China is dominated by vertical FDI. The results also confirm the existence of the gravity effect. FDI increases with market size in both home and host regions but decreases with distance. Other independent variables, such as the economic standard in the home and host regions, and the R&D level in the home country, are also significant determinants of FDI. However, the bonded area in the host region is not a clear contributing factor to FDI flow.

The empirical findings above provide implications for policymakers in China. First, the provincial government should put emphasis on attracting FDI from culturally close countries and from countries with advantages in R&D, since these countries have more motivation to invest than others. Second, the government should provide some institutional and policy support to encourage and promote horizontal FDI, which can reduce the negative effects of geographic and economic distances on FDI inflow. Although we provided a novel perspective by investigating the effect of distance on the FDI location choice in China, the measurements of the distance factors are far from perfect, indicating this study is still at an early stage. Future studies may benefit from using various measures of distance to obtain more solid conclusions.

Author details
[1]Business School, Renmin University of China, Beijing, China. [2]Business School, University of Nottingham, Nottingham, UK.

References
Aleksynska, M., & Havrylchyk, O. (2013). FDI from the south: The role of institutional distance and natural resources. *European Journal of Political Economy, 29*, 38–53.
Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review, 93*(1), 170–192.
Bailey, N., & Li, S. (2015). Cross-national distance and FDI: The moderating role of host country local demand. *Journal of International Management, 21*(4), 267–276.
Belderbos, R., Olen, W. V., & Zou, J. (2011). Generic and specific social learning mechanisms in foreign entry location choice. *Strategic Management Journal, 32*(12), 1309–1330.
Buckley, P. J., & Casson, M. C. (1998). Analysing foreign market entry strategies: Extending the internalization approach. *Journal of International Business Studies, 29*(3), 539–561.
Chen, Y., Wang, S. S., Tong, W. H. S., & Zhu, H. (2017). Economic freedom and IPO underpricing. *Frontiers of Business Research in China, 11*(4), 453–483.
Chenaf-Nicet, D., & Rougier, E. (2016). The effect of macroeconomic instability on FDI flows: A gravity estimation of the impact of regional integration in the case of Euro-Mediterranean agreements. *International Economics, 145*, 66–91.
Cheng L K, Kwan Y K. (2000). What are the determinants of the location of foreign direct investment? The Chinese experience. *Journal of International Economics, 51*(2), 379–400.
Choi, J. J., Lee, S. M., & Shoham, A. (2016). The effects of institutional distance on FDI inflow: General environmental institutions (GEI) versus minority investor protection institutions (MIP). *International Business Review, 25*(1), 114–123.
Cuervo-Cazurra, A. (2006). Who cares about corruption? *Journal of International Business Studies, 37*(6), 807–822.
Du, J., Lu, Y., & Tao, Z. (2012). Institutions and FDI location choice: The role of cultural distances. *Journal of Asian Economics, 23*(3), 210–223.
Dunning, J. H. (1980). Toward an eclectic theory of international production: Some empirical tests. *Journal of International Business Studies, 11*(1), 9–31.
Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica, 70*(5), 1741–1779.
Eden, L., & Miller, S. R. (2004). Distance matters: Liability of foreignness, institutional distance and ownership strategy. *Advances In International Management, 16*(4), 187–221.
Egger, P., & Pfaffermayr, M. (2004). The impact of bilateral investment treaties on foreign direct investment. *Journal of Comparative Economics, 32*(4), 788–804.
Feenstra, R. C. (2015). *Advanced international trade: Theory and evidence*. Princeton: Princeton University Press.

Flores, R. G., & Aguilera, R. V. (2007). Globalization and location choice: An analysis of US multinational firms in 1980 and 2000. *Journal of International Business Studies, 38*(7), 1187–1210.

Galan, J. I., Gonzalez-Benito, J., & Zuniga-Vincente, J. A. (2007). Factors determining the location decisions of Spanish MNEs: An analysis based on the investment development path. *Journal of International Business Studies, 38*(6), 975–997.

Garrett, J. Z. (2016). Explaining asymmetries in bilateral FDI flows. *International Review of Economics and Finance, 41*, 155–171.

Henisz, W. J. (2000). The institutional environment for multinational investment. *The Journal of Law, Economics, and Organization, 16*(2), 334–364.

Henisz, W. J., & Williamson, O. E. (1999). Comparative economic organization within and between countries. *Business and Politics, 1*(3), 261–277.

Hirsch, S. (1976). An international trade and investment theory of the firm. *Oxford Economic Papers, 28*(2), 258–270.

Hymer, S. H. (1976). The international operations of national firms: A study of direct foreign investment. *Journal of Development Economics, 4*(4), 387–400.

Jones, G., & Khanna, T. (2006). Bringing history (back) into international business. *Journal of International Business Studies, 37*(4), 453–468.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies, 19*(3), 411–432.

Kostova, T., & Roth, K. (2002). Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects. *Academy of Management Journal, 45*(1), 215–233.

Kostova, T., & Zaheer, S. (1999). Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. *Academy of Management Review, 24*(1), 64–81.

Kuang, W., & Wang, Q. (2018). Cultural similarities and housing market linkage: Evidence from OECD countries. *Frontiers of Business Research in China, 12*(2), 101–125.

Linders, G.-J., & de Groot, H. (2006). *Estimation of the gravity equation in the presence of zero flows. Tinbergen institute discussion paper, no. 06-072/3.*

Makino, S., Isobe, T., & Chan, C. M. (2004). Does country matter? *Strategic Management Journal, 25*(10), 1027–1043.

Mathews, J. A. (2006). Dragon multinationals: New players in 21st century globalization. *Asia Pacific Journal of Management, 23*(1), 5–27.

McCallum, J. (1995). National borders matter: Canada-US regional trade patterns. *The American Economic Review, 85*(3), 615–623.

Nielsen, B. B., Asmussen, C. G., & Weatherall, C. D. (2017). The location choice of foreign direct investments: Empirical evidence and methodological challenges. *Journal of World Business, 52*(1), 62–82.

Pan, Y. (1997). The formation of Japanese and US equity joint ventures in China. *Strategic Management Journal, 18*(3), 247–254.

Paniagua, J., & Sapena, J. (2014). Is FDI doing good? A golden rule for FDI ethics. *Journal of Business Research, 67*(5), 807–812.

Portes, R., & Rey, H. (2005). The determinants of cross-border equity flows. *Journal of International Economics, 65*(2), 269–296.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London, 48*(2), 167–235.

Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics, 62*(1), 53–82.

Savage, I. R., & Deutsch, K. W. (1960). A statistical model of the gross analysis of transaction flows. *Econometrica, 28*(3), 551–572.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics, 88*(4), 641–658.

Tinbergen, J., & Hekscher, A. (1962). *Shaping the world economy. Suggestions for an international economic policy*. New York: Twentieth Century Fund.

Trefler, D. (2002). The case of the missing trade and other mysteries: Reply. *The American Economic Review, 92*(1), 405–410.

Westerlund, J., & Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics, 43*(6), 641–649.

Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal, 38*(2), 341–363.

Publisher's Note