Identifying the effect of transshipment hubs on international trade[*]

Nicholas-Joseph Lazarou[†]

February 5, 2019

Abstract

Transshipment hubs such as Singapore have a ubiquitous role in facilitating the international trade of goods. Yet how their presence affects trade is under-explored. I blend network theory and geography in a simple representative firm trade model and find that hub formation is governed by geographical centrality and peripheral countries' potential to connect to them, manifested through levels of transport costs and infrastructure. The equilibrium characterisation of exports assists in empirically identifying their effect by treating observed exports as an outcome of the choice of route. I compare two capital distance variables, one direct, the other indirect which is interacted with an indicator signifying the presence of a hub on a route concluding that: i) Remote countries have on average 26% less impact of distance on exports if they trade via a hub compared to trading directly; ii) distance variables appear to capture information about the choice of route.

Keywords: International Trade; Economic Geography; Transport Costs; Network Formations.

JEL Classification Codes: F12, F14, L91, L14.

December 2018

---

[*] I am grateful to Héctor F. Calvo Pardo and Jackline Wahba for their faithful support and advice. I sincerely thank my external examiner Luca David Opromolla, Andreas Moxnes, Costas Arkolakis, Sotiris Blanas, James Feyrer, Marco Fugazza, Robert C. Johnson, Nina Pavcnik, Dimitra Petropoulou, Volodymyr Lugovskyy, Christian Volpe Martincus; participants of the 2011 PhD Workshop at the University of Southampton, 2011 UECE Lisbon meetings, 2012 CRETE, 2012 ETSG, 2015 Royal Economic Society, 2015 LETC FREIT conferences for their constructive recommendations and guidance. I am indebted to Daniel Rockmore and Andreas Moxnes who kindly provided me with an office at the Department of Mathematics at Dartmouth College where I developed parts of this paper. This work was partially carried out at the University of Southampton while the author was funded by the Economic & Social Research Council PhD studentship No. ES/1026169/1.

[†] Energy Institute, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom; and,
Council of Economic Advisors, Ministry of Finance of the Hellenic Republic, 5-7 Nikis Street, Athens GR10180, Greece. Email: N.Lazarou@ucl.ac.uk

Disclaimer: "The views and opinions expressed in this paper are of the authors and do not necessarily reflect the official policy or position of the Council of Economic Advisors of the Hellenic Ministry of Finance nor the incumbent Government of the Hellenic Republic."

1 Introduction

Transshipment hub ports such as Singapore, Shanghai or Hong Kong have a ubiquitous role in facilitating the international trade of goods.[1] The prominence of particular locations serving as transshipment hubs compared to other locations is the outcome of geographic advantage and concentration of production: Historical incidence, interregional trade and globalisation all seem to play a role in the development of hub ports (Krugman, 1993; Levinson, 2008). Yet do transshipment hubs, which possibly alter the geography of trade routes by virtue of their existence, have an observable effect on the international trade of goods?

The aim and contribution of this paper is the derivation of a set of testable expressions in order to measure any identifiable effect of the presence of hubs on trade flows. To achieve this, I embed the symmetric connections framework of Jackson and Wolinsky (1996) in a three country trade model of monopolistic competition á la Krugman.

The theory reveals that when transport costs to a remote destination are high, firms instead of lowering output or cease exporting to that location altogether, choose to export indirectly utilising a hub. The decision is made if a cost saving incentive exists compared to exporting directly to that destination. This finding stems from geographical incidence and the interplay between levels of transport costs and infrastructure.

If firms in peripheral countries export to each other using a hub, then large amounts of transport infrastructure are required in the hub to facilitate transshipment as well as domestic firms' exports to both destinations. Peripheral countries require less transport infrastructure since they do not tranship goods nor export directly to all destinations. Optimising the fixed costs of transport across countries reallocates labour input which is used to transport goods even further away for peripheral countries, and over shorter distances

---

[1] The total number of containers handled by these ports alone in 2008 stands at about 60 per cent of the volume of global containerised trade or about 6 per cent of global trade by weight. Since the 1950's similar ports the size of cities have emerged as a corollary of technological advances in shipbuilding, port infrastructure and the advent of containerisation. As these ports became increasingly interconnected across time, they enabled the transshipment of goods between remote locations, thus serving as hubs. For instance the port of Singapore today offers connections to 600 ports in 123 countries and accounts for 14 per cent of the world's total container transshipment throughput (Hummels, 2007; Levinson, 2008; UNCTAD Secretariat, 2009; Rodrigue, 2010; Rua, 2014; Bernhofen, El-Sahli and Kneller, 2015; Maritime and Port Authority of Singapore, 2018).

for the hub. Ever more distant peripheral countries find it harder to connect to the hub, but firms (and consumers in the other peripheral destination) prefer this situation compared to exporting less or nothing at all to their peripheral partners without using the hub.

The model's equilibrium exports characterisation admits a set of testable equations. The empirical strategy is to identify whether the effect of hubs on trade is observable and ameliorating for remote trading partners as the theory predicts. I compare the effect of two distance variables on exports, great circle capital distance and indirect capital distance via major ports of trading partners. Indirect distance is interacted with a dummy indicator variable if a route passes or not through a hub location.

Across three aggregation levels of trade data, I find that exports to a destination reduce on average 26% less for every 10% increase in distance when a hub is present on a route compared to exporting the same amount directly to this destination. Countries not having a hub intersecting the route tend to be less remote. More remote partners use at least one hub for exporting, but the greater the number of hubs employed, the lower the gains from their utilisation. The revealed choice to trade via a hub implies that distance variables explain information about the selection of routes when the interaction is unobserved and forms part of the error term. I attempt thus to uncover the pure effect of remoteness using time difference as an instrument which satisfies relevance and the exclusion restriction.

The international trade literature has not addressed whether transshipment hubs have any impact on trade volumes apart from a study by Fugazza (2015) whose results complement the findings of this paper. He augments the so called "gravity equation" of trade with three variables, the calculated sea distance between countries' major container ports,[2] an indicator of (in)direct maritime connections, and the estimated number of transshipments en route obtained from solving a shortest route problem. He finds that country level containerisable exports reduce by 42-55% if routes entail transshipment ports compared to routes that do not. Increasing the number of transshipments reduces exports but at a diminishing rate. Distance variables appear to overestimate the true impact of remoteness when information about the choice of route and use of transshipment ports is unmeasurable.

---

[2] And proximity to these ports in case of landlocked countries.

A plausible reason for the lack of research output may be the insufficient attention paid to the transport industry's structure combined with data paucity on transport costs and routes (Krugman, 1993; Fugazza, 2015). The most related studies that theoretically explain when a hub network forms, these being Krugman (1993), Hendricks, Piccione and Tan (1995), Starr and Stinchcombe (1992), Oum, Zhang and Zhang (1995), Mori (2012) are all characterised by an infeasibility to develop a testable prediction for the effect of hubs, validating why there is little empirical application of theoretical trade frameworks embedding economic geography (Allen and Arkolakis, 2014). I contribute by providing empirically testable specifications derived from a simple economic geography model, which uncover that the presence hubs is beneficial for remote trading partners as opposed to their absence.

This is enabled by relying heavily on Krugman (1993), wherein the interaction between increasing returns in manufacturing and low levels of transport costs to/from a location leads to emergence of a hub because of concentration of production, which implies strong historical incidence. I extend this work by removing concentration of production as a prerequisite, keeping locational advantage and the level of transport costs exogenous, and embedding relative changes in transport infrastructure in order to obtain the testable expression. The reasoning is that some of the less developed locations of the world, like Panama or Port Said obtain hub status conditional on location based on the distribution of trade routes. Concentration of production may then take place and reinforce the role of the hub but it is not a condition precedent for its formation.

The models of Hendricks, Piccione and Tan (1995), Starr and Stinchcombe (1992), Oum, Zhang and Zhang (1995) focus on air passenger transportation. They find that economies of density are a determinant of airline hub network formation as total costs per passenger on a route[3] decline with the number of passengers travelling on that route. Airline hub networks exhibit higher traffic densities than direct connections' networks. The distance travelled is longer but if economies of density are sufficiently large, transport costs decline. The combination of transport distance and density is the main determinant of the location of the hub and the spatial structure of the transport network (Mori, 2012).

---

[3] Attributed either to spreading fixed costs over a larger number of passengers or declining marginal costs.

The related empirical literature employs natural experiments to infer the marginal effect of transport costs on trade flows. The shocks that permit identification by isolating the causal direction are the closure of the Suez Canal (Feyrer, 2011), the closure of a bridge (Martincus, Carballo, Garcia and Graziano, 2014) and the eruption of the Eyjafjallajökull volcano (Besedeš and Murshid, 2015). Yet these shocks are few and far between; they are unanticipated and their effect has varying dispersion, making inference difficult to isolate and to generalise. This paper does not benefit from observing similar shocks. In fact, data limitations regarding transport infrastructure, product level trade routes, frequency and size of shipments of goods between ports substantially narrow the choice of empirical approach. Instead it proposes a simple mechanism using distance, the oft-employed proxy for the marginal cost of transport, to capture information about the choice of routes. If unobserved, this information is contained in the error term in which case instrumental variable estimation is used to quantify the marginal effect of remoteness.

The paper also contributes to a scarce literature that perturbs multiplicative trade costs, colloquially known as iceberg costs, to model the behaviour of an otherwise unobserved transport sector. This is a field of research which is considered to be under-explored as Melitz and Redding (2014) state.[4]

The policy outcome of this paper is that improving own and transit country infrastructure could contribute towards lowering transport costs to and from these countries while spurring the reallocation of factor inputs (Limão and Venables, 2001). Higher transport costs are attributed to geographic disadvantage and lack of proximity. Incumbent exporters have to absorb transport costs so as to access foreign markets, a situation that can prevent export-led development, reduce workers' wages and welfare (Amjadi, Reinke and Yeats, 1996). Prevention of market access for developing nations translates to losses from trade standing at about 68% lower GDP per capita on average (Redding and Venables, 2004).

The remainder of this paper is organised as follows. Section 2 provides the network setup and the economic environment in which it is embedded. The equilibria are characterised

---

[4] Key papers in this literature are Limão and Venables (2001), Clark, Dollar and Micco (2004), Hummels and Skiba (2004), Hummels, Lugovskyy and Skiba (2009), Harrigan (2010), Lux (2011), Rua (2014), Bernhofen, El-Sahli and Kneller (2015), Irarrazabal, Moxnes and Opromolla (2015).

in section 2.3. The empirical specifications for testing the existence of hub formations are derived in section 3 followed by a description of the dataset in section 4 and a discussion of the results in section 5. Section 6 concludes.

2 The model

In this section I develop the theoretical model which explains some of the stylised facts about hub formations presented in the introduction. I define the network notation used in the representative firm trade model and then proceed to state and discuss the characterisation of the equilibrium network formations.

2.1 Setup of the network

Consider a set of countries $K = \{1, 2, \ldots, k\}$ whose incumbent homogeneous manufacturing firms engage in costly exporting of their produce. Countries are said to be *directly connected* if one exports to another using no other intermediate country in the process. A network G is defined as the list of country pairs $\{i, j\}$ directly connected to each other. Connections are indexed using binary *graphs* $g \in G$ which obtain value equal to 1 if i and j are directly connected, or zero if they are *not directly connected.*

Each connection is characterised by costs and benefits that affect exporting firms equally. Should firms form a direct connection the cost of maintaining the connection is c and there is a benefit $0 \leq \delta \leq 1$ which can be thought of as a measure of market accessibility of the importing country. Firms also have the option of forming indirect connections to export their goods. Implicitly there must exist a direct connection to another country for the indirect connection to be feasible. Firms do not incur any additional costs to form an indirect connection, but since accessibility worsens they enjoy a decayed benefit $\delta^{t_{ij}}$, where $t_{ij} \geq 0$ is the integer number of connections between countries i,j. The net difference between the benefit and the cost of forming a particular connection is defined as $v_{ij} = \delta^{t_{ij}} - c_{ij|ij \in G}$.

By convention $g_{ii} = 0 \Rightarrow c_{ii} = 0$ and country i is autarkic. Further, $t_{ij} = 0$ occurs when there is no path that connects (in-)directly countries i and j. The simplest example is exhibited in Appendix A.1.

A proxy measure for fixed costs of exporting that is common across firms within a country is the *network participation share.* The set $N_i(G) = \{i \neq j | g_{ij} = 1\}$ with cardinality $n_i(G)$ contains country i's direct connections. The size of the network is $n(G) = \sum_i^N n_i(G)$. Country i's network participation share is $S_i = \frac{n_i(G)}{n(G)}$. An example is provided in Appendix A.2.

This setup allows a firm in country i which is considering exporting its products to assess i) how accessible foreign markets are (δ), ii) the cost to export (c), and decide the connection it will form.

2.2 Setup of the trade model

– *Description of the Economy*

Country $j \in K$ has a population L_j and two sectors which use labour as the only factor of production. One sector is responsible for the production of a single freely traded homogeneous good. This good is the numeraire. The other sector, labelled manufacturing, consists of a continuum of identically sized monopolist firms each producing a single differentiated variety of a good, that is subject to costly trade. In both sectors firms can freely enter or exit production. The population works in the sectors, moves freely across sectors but not across countries and consumes both goods. Each consumer is endowed with one unit of labour.

– *Consumer Demand*

The representative consumer of country $j \in K$ has preferences of the CES form over the continuum of differentiated varieties with an elasticity of substitution $\sigma \equiv \frac{1}{1-\theta} > 1$, where $0 < \theta < 1$ is the intensity of the preference.[5] Such preferences given the consumer's endowment, yield the demand function $q_{ij} = A_{ij} p_{ij}^{-\sigma}$ for a variety exported from country i with price p_{ij}. The demand can shift according to $A_{ij} \equiv \mu_j L_j P_j^{-1}$ which comprises the share μ of income L expended on purchasing the variety over the aggregate price $P_j =$

---

[5] As θ approaches 1 varieties become almost perfect substitutes; as θ approaches zero an increased number of varieties available to the consumer results in higher levels of utility.

$\sum_{j,i=1}^{N}\int_{\Omega}p_{ij}^{1-\sigma}(\omega)d\omega$ which would prevail if varieties $\omega\in\Omega$ were consumed as an aggregate good.

– *Production and Trade Costs*

The price of the numeraire good is normalised to 1 so that the wage is equal to the price of the good. Therefore the wage is set equal to 1 across countries, due to free trade, and across the two sectors.

Labour costs for producing a differentiated variety are split into two components, a marginal and a fixed cost, thus the sector is characterised by increasing returns. The marginal cost consists of a constant parameter $\gamma > 0$ and a variable trade cost, which are both common to all differentiated firms exporting to the same destination. The variable trade cost is the net difference between the benefit and cost of forming a connection, v_{ij} measured in units of the variety. In summary, to produce and sell a variety abroad, a firm's labour input is:

$$L(q)\equiv\gamma\frac{q_{ij}}{v_{ij}}+F_i\equiv\gamma\frac{q_{ij}}{\delta^{t_{ij}}-c_{ij|ij\in G}}+F_i,$$

where the distinction $v_{ii}=1$ is made for the case of domestic production. If a direct connection is formed the firm receives a benefit in the sense that its profits are increasing in δ, but decreasing in the cost c of transporting goods to the destination. In the case of an indirect connection, firm profits increase but at a decreasing rate since the benefit is discounted at no additional cost.

Partial equilibrium in manufacturing yields the firm's i) profit maximising price of a variety, equal to a constant markup over the unit cost, $p_{ij}=\frac{\sigma}{\sigma-1}\frac{\gamma}{v_{ij}}$ and ii) output, $Q_i\equiv\sum_{i,j=1}^{k}q_{ij}/v_{ij}=\frac{F_i}{\gamma}(\sigma-1)$. Because of full employment in the sector, the number of firms (and varieties) are $n_i=\frac{\mu_i L_i}{\sigma F_i}$. Consequently the aggregate price index becomes $\sum_{j,i=1}^{N}n_i p_{ij}^{1-\sigma}$.

– *Equivalence with additive costs framework*

At this stage, it is important to mention that the equilibria described in the next section are qualitatively equivalent to those of a rather cumbersome framework of homogenous firm trade presented in Appendix B. The two models yield similar conditions for hub manifestation, yet the one described in the Appendix is not capable of producing an empirically testable equation. Its sole purpose is to lend itself to replace the benefit of forming a connection δ as it is unmeasurable in the data, with the network participation share S_i, a type of fixed cost, ensuring all the while that the conditions for hub manifestation are not distorted. Then it is the interplay between transport costs and the fixed cost of transport which dictates if hubs are formed and also allows empirical estimation of identified coefficients.

The appended framework cannot provide a testable prediction because of the presence of an additive cost of transport d in manufacturing profit maximising prices $p_{ij} = \frac{1}{\theta}\left(\frac{\gamma}{\theta}\tau_{ij} + \phi d_{ij}\right)$, in addition to the multiplicative cost τ. The additive part implies the coefficient of transport costs ϕ cannot be separately identified (Irarrazabal, Moxnes and Opromolla, 2015). In both frameworks the profit maximising output is $Q_i \equiv \frac{F_i}{\gamma}(\sigma - 1)$ and the number of firms is $n_i = \frac{\mu_i L_i}{\sigma F_i}$, where $F_i \propto S_i$. The additive cost model's proposition regarding hub formations is qualitatively equivalent to the one stated in the next section and so both are presented in tandem.

2.3 Equilibria and network formations

In a similar manner to Krugman (1993) I assume that all countries have the same income expenditure share μ and size L which sets aside the home market effect between any two countries. It is normalised to 1 because the number of manufacturing firms within each country equalise. The zero profit condition in manufacturing for the technologically identical countries ensures that output is equal among all countries:

$$\sum_{1,j=1}^{k} \frac{q_{1j}}{v_{1j}} = \sum_{2,j=1}^{k} \frac{q_{2j}}{v_{2j}} = \cdots = \sum_{k,j=1}^{k} \frac{q_{kj}}{v_{kj}} \Rightarrow Q_1 = Q_2 = \ldots = Q_k \tag{1}$$

This result, albeit uninteresting, assists firms within a country to decide on the connection type through their respective countries' price indices. To state the proposition I set the number of countries $k = 3$ to ensure their symmetric placement on a plane, $c_{ij} = c_{ji} = c, t_{ij} = t_{ji} = t, i \neq j$.

- *Direct Connections Network* Because of the zero profit condition 1 and direct connections which imply $v_{ij} = \delta - c$, the price index of any of the three countries is:

$$P_j = n\left[\frac{\sigma}{\sigma-1}\gamma\right]^{1-\sigma}\left(1+2(\delta-c)^{\sigma-1}\right), j \in \{1,2,3\}$$

- *Indirect Connections Network*

If 1 and 2 are indirectly connected to each other via 3, which becomes the hub, and condition 1 must hold, the price index of 1 (or 2) becomes:

$$P_j = n\left[\frac{\sigma}{\sigma-1}\gamma\right]^{1-\sigma}\left(1+(\delta-c)^{\sigma-1}+(\delta^2)^{\sigma-1}\right), j \in \{1,2\}$$

where δ^2 is the benefit from connecting to the complement country via the hub country 3. Country 3's price index remains the same as in the direct connections case.

The two expressions equate if the zero profit condition happens to hold regardless of the network type. This occurs at the unique combination of costs and benefits from forming the connections, $\delta - c = \delta^2$, a point where there is indifference between network types.

If either of the two parameters δ or c vary, then the equilibrium manifests only within a specific type of network. Sorting transport costs $c \in \mathbf{C}$ in ascending order, where $\mathbf{C} = \{\dots, \bar{c}, \dots, \hat{c}, \dots\}$, suppose firms observe a very high transport cost c between countries 1 and 2. Firms in 1 or 2 could export varieties to one another using the direct connections network. Profits of firms in 1 become (firms in 2 obtain the same expression):

$$\pi_1^D = q_{11}(p_{11}-\gamma) + q_{12}^D(p_{12}^D - \frac{\gamma}{\delta-c}) + q_{13}^D(p_{13}^D - \frac{\gamma}{\delta-c}),$$

Alternatively they could choose to export to each other using 3 as a hub:

$$\pi_1^I = q_{11}(p_{11}-\gamma) + q_{12}^I(p_{12}^I - \frac{\gamma}{\delta^2}) + q_{13}^I(p_{13}^I - \frac{\gamma}{\delta-c})$$

where superscripts $D(I)$ denote (in)direct.

Profits under the hub network are always greater than the profits obtained by trading directly as $\pi_1^I > \pi_1^D \Rightarrow \delta - c < \delta^2$. This happens because of the very high cost of transport

c which forces firms in 1 and 2 to sever the direct connection between them in order to continue exporting, or exit the market. In this configuration, the zero profit condition holds when $c = \bar{c} = \inf\{c \in \mathbf{C} : \delta - c < \delta^2 \text{ and } 0 = \pi_1^I(\bar{c})\}$. Below $\bar{c}$, firms suffer losses if they keep trading indirectly via 3 but enjoy at least positive profits by trading directly. The relationship becomes $\pi_1^D > \pi_1^I \Rightarrow \delta - c > \delta^2$ and the equilibrium holds when $c = \delta$. The cost of adding a connection is less than the benefit the firm gains from shortening the connection of length two into a connection of length one.

When c is extremely high countries become autarkic. In this case firm profits from exporting are negative for both formations. When costs are $c > \delta + \delta^2$ indirect trading is prevented and because $c > \delta$, direct trading is also prevented. The threshold value for autarky is $c \geq \hat{c}$, where $\hat{c} = \inf\{c \in \mathbf{C} : c > \delta + \frac{\delta^2}{2}\}$.

The conditions for network formation coincide with those contained in Proposition 1 of Jackson and Wolinsky (1996) and are summarised as follows.

Proposition

For unique values of c in the set $\mathbf{C}$ while holding constant the benefit term δ, the network formation decisions for an exporting representative firm are:

i. A direct connections formation when $0 \leq c < \bar{c}$ where the equilibrium holds if $\delta = c$.

ii. A hub formation when $\bar{c} \leq c < \hat{c}$ where the equilibrium holds if $c = \bar{c}$ for a given $\delta < c$.

iii. Autarky if $\hat{c} \leq c$ for a given $\delta < c$ and there exists a range of autarkic equilibria.

iv. Indifference between direct or hub formations if $\delta = \bar{c}$.

v. Indifference between autarky and a hub formation if $\bar{c} = \hat{c}$.

vi. Indifference between autarky and any network formation if $\delta = \bar{c} = \hat{c}$.

For the interesting cases $i., ii.$ the exogenous levels of δ determine the equilibrium formation when a specific value of c is observed by firms in connecting countries. For sufficiently low levels of c, the cost of adding an extra connection yields a higher net benefit compared to only receiving a discounted benefit at the cost of one connection. And so firms always prefer a direct connection. Conversely, when it is too costly to maintain connections firms sever these and receive only the discounted benefit. This type of interplay is analogous to the outcomes of the additive costs model of Appendix B. Given a cost c, the extra benefit from severing or adding a connection can be thought of as the gain from deciding the optimal level of fixed costs of transport. The resulting proposition is:

1. *If the connecting country's transport firms decide to trade using the hub, they must increase trading distance to* $d'' = 2d > d'$. *This requires an increase in the ratio of fixed costs of transport between the hub and the connecting country, either by raising (lowering) the hub's (connecting country's) fixed costs of transport or any increasing combination of both.*
2. *If the hub is ever more distant from the connecting country, the ratio of fixed costs of transport must decrease either by lowering (raising) the hub's (connecting country's) fixed costs of transport or any decreasing combination of both.*

– *Discussion*

These network formations are *uniquely efficient* and *stable* based on the definitions of Jackson (2003). They are uniquely efficient because only one case prevails at a time and no other network can accommodate higher profits at a particular combination of δ and c. Given δ the hub network is *stable* for costs $\delta - \delta^2 \leq c$: Country 3 being in the center, becomes worse off if a connection is severed since utility for consumers and varieties traded decrease. Firms in 1 are adversely affected by this choice too as profits decrease. Therefore firms never choose to sever the connection with 3. If firms in 1 form a direct connection with 2 at this cost level instead of the indirect connection, profits decrease so firms would not connect. If they actually did, firms in 2 would sever their connection with the hub due to the high cost of maintenance. Thus the hub formation is *pairwise stable* but not necessarily unique as it can also rotate between countries. For lower costs c all direct connections are pairwise stable because firms are not willing to sever a connection. Firms in any two countries not directly connected benefit from forming a connection. The proof for these statements is available in the Technical Appendix.

Contrary to the Krugman (1993) three country trade model I remove historical incidence and concentration of production as prerequisites for hub manifestation. Locations not necessarily benefited from concentration of production, possibly created by historical incidence, can become hubs by virtue of geographic incidence and because of firm production in other, geographically disadvantageous, locations. Concentration of production could take place after the formation, but is not a condition precedent for the model's equilibria. In Krugman's framework the equilibrium arises by postulating concentration of production in a particular location, asymmetry in distance and employment of a defecting firm to explore other locations' production possibilities. When the cost of transport between two locations is less than between other locations the hub forms, concentration of production materialises, and is re-inforced. Herein I can simply postulate an excessive transport cost between all location pairs due to symmetry and gradually decrease it.[6] At some very high but not prohibitive transport cost, firms producing domestically can decide to increase profits by exporting to two or more countries. They do so by deciding the network formation that minimises exposure to the high costs they face and a hub network manifests.

This situation however has negative implications for the global economy, especially when considering locations that are far from established trade routes: Each firm's labour input would need to be increased and subsequently the number of firms or varieties traded are lower compared to a direct connections network. Consumers in importing countries enjoy less varieties and have lower utility levels as a consequence of high transport costs. Yet it is a strictly better situation than an autarkic equilibrium. These consequences are in line with the findings of Amjadi, Reinke and Yeats (1996) who report that Sub-Saharan African exporters have to absorb higher transport costs so as to be in the position to penetrate markets abroad. A large share of these countries' foreign exchange earnings instead of being used for productive investment, are allocated to paying the cost of transport.

In reality transport costs, and to some extent the transport network itself, depend on the pricing decisions and organisation of transport firms (Hendricks, Piccione and Tan, 1995). If exports using a hub network prevail, a large amount of transport infrastructure is required

---

[6] For expositional clarity henceforth the benefit δ is such that there is no indifference between formations nor it is too low to admit autarky.

in the hub to cope with increased traffic, while peripheral countries require only a fraction of this amount, otherwise it is underutilised. So the fixed costs of transport matter, reflected as a reduced network participation share for firms in connecting countries and a higher share for those in the hub. But distance, representing the marginal cost of transport, increases for firms in connecting countries. Otherwise the opposite should hold: For direct connections distance is shorter but fixed costs of transport infrastructure are higher. This juxtaposition is crucial for deriving the testable equations.

3 Empirical Strategy

In this section I lay out the econometric design and attempt to identify whether the presence of hubs on trade routes is observable and has an ameliorating effect as the theory predicts. By solving the endogeneity problem, I infer whether the effect of distance is attributed solely to remoteness, or if it also consists of information regarding the choice of the route.

Manufacturing exports take the familiar gravity characterisation:

$$x_{ij} = n_i p_{ij} q_{ij} = \mu_i L_i \mu_j L_j \frac{v_{ij}^{\sigma-1}}{\lambda_j} \frac{1}{\sigma F_i}, \tag{2}$$

where $\lambda_j = \sum_{j,l=1}^{N} n_l v_{lj}^{\sigma-1}$ is an aggregate index of network costs in j. As it is not possible to measure the benefit of forming a particular connection in the real world, I break down the problem in two parts, exploiting the qualitative equivalence of the two models' propositions. First, I assume the transport cost c is proportional to distance and follows the functional form $c_{ij} \propto d_{ij}^{\beta} \times \exp(\beta_0)$. Second, fixed costs are assumed to be proportional to the network participation share S_i which affects all firms of a country uniformly. Because of the similarity of the propositions I substitute v_{ij} with $d_{ij}^{\beta} \times \exp(\beta_0)$ and F_i with S_i in equation 2. When firms in i export to j they can do so directly or utilise a hub k. Taking logs, direct and indirect sectoral export volumes are:

$$\ln x_{ij}^{\mathbb{I}} = \beta_0 + \beta_1 \ln(\mu_i L_i) + \beta_2 \ln(\mu_j L_j) + \beta_3^{\mathbb{I}}(\sigma - 1)\big[\ln d_{ij}(\mathbb{I} - 1) - \ln(d_{ik} + d_{kj})\mathbb{I}\big] \\ - \beta_4^{\mathbb{I}} \ln S_i^{\mathbb{I}} - \ln(\sigma\lambda_j), \text{where } \mathbb{I} \equiv \begin{cases} D = 0, & \text{if direct} \\ I = 1, & \text{if indirect} \end{cases} \tag{3}$$

Country fixed effects control for country size and aggregate network cost indices (Chaney, 2005) but absorb variation of country specific fixed costs of transport crucially rendering $\beta_4^{\mathbb{I}}$ useless in explaining the effect of fixed costs when trading via the hub. Yet under such a specification one can conduct consistent and efficient estimation of the partial effects of the remaining regressors (Greene, 2008). I add a binary interaction term that indicates the presence of at least one hub location on the route. If a hub is involved when firms of a sector export to a destination, it must imply that transport costs were very high preventing direct connections or the opposite, holding exports constant. Gravity equation 3 obtains its testable forms which can be estimated using ordinary least squares:

$$\ln \left[x_{ij}^s\right]^D = A_{ij}^{\prime s} + X_{ij}^{\prime} B_1 - \beta_3^D(\sigma - 1)\ln d_{ij} + \epsilon_{ij}^s \tag{4}$$

$$\ln \left[x_{ij}^s\right]^I = A_{ij}^{\prime s} + X_{ij}^{\prime} \tilde{B}_1 - \ln(d_{ik} + d_{kj}) \left[\beta_3^I(\sigma - 1) + \iota \text{Hub}_{ij}\right] + \xi_{ij}^s, \tag{5}$$

where s denotes sector, prime denotes transpose, A_{ij}^s is a vector comprising a constant, country and sector dummies, X_{ij} is a vector of observed trade barriers between countries i and j other than distance and ϵ_{ij}^s, ξ_{ij}^s are both orthogonal to the independent variables and normally distributed. I assume that shocks affect trade flows within each country pair and so standard errors are clustered at this level.

The two testable equations are not comparable because they do not admit the same level of exports. However actual sectoral exports to a particular destination observed in the data are due to the prevailing configuration, $\ln x_{ij}^s = \max\{\ln \left[x_{ij}^s\right]^I, \ln \left[x_{ij}^s\right]^D\}$. A comparison across equations is enabled by employing the common dependent variable and applying interchangeably the right hand sides of 4 and 5. Then I can test whether the two partial effects of distance $\beta_3^D(\sigma - 1)$, $\left[\beta_3^I(\sigma - 1) + \iota\right]$ are significantly different from each other.

It may be possible that some hub locations are commodity/sector specific and so the variation would be absorbed by the sector dummy. The data unfortunately does not contain this information as routes are not commodity specific and so I perform experiments using all possible combinations of dummies.

According to the model's predictions, if a hub effect exists then the overall marginal effect of two countries trading via a hub is significantly less than the marginal effect of them trading directly in absolute value, thus ameliorating the negative impact of indirect distance. Yet it is not immediately apparent how the coefficient of indirect distance should behave relative to the direct distance counterpart, absent of any information regarding hub locations serving the route. This is calculated in section 5 using the available data described in the next section.

4 Data and summary statistics

I match export values from the UN Comtrade Database and maritime transport costs obtained from the OECD Maritime Transport Costs Database at the same Harmonised System (HS) classification level and year. Three aggregation levels of the Harmonised System are considered: The total (country) level of trade,[7] the chapter (HS 2) and subheading (HS 6) levels. The sample consists of 63 exporters, 42 importers, 408 trade partnerships which are listed in the Technical Appendix, and the time period is 1991-2007. The justification for this type of matching is twofold: First, 99% of the world's trade by weight (Hummels, 2007), and 90% of the volume of merchandised trade is transported by sea (OECD Trade and Agriculture Directorate, 2008; Korinek and Sourdin, 2010). Second, with the emergence of hub networks induced by the advent of containerisation, I assume the observed price maritime transport firms charge is a function of the network organisation (Rua, 2014). By conducting this operation some of the exports actually transported by sea and therefore by some form of network can be captured.

Each pair is assigned great circle capital distance, henceforth direct distance, and common border, common official language, past colonial relationship and active regional trade

---

[7] These are own constructs from country pair summations of chapter levels.

agreement binary dummies taken from the CEPII GeoDist dataset (Mayer and Zignago, 2011). Regional trade agreements for year 2007 are own constructs.

For capturing possible effects due to hubs on a route I build indirect capital distances, henceforth indirect distances. For a random subset of trading partners I measure the distance from the capital of the (ex-)importer to the closest domestic major (ex-)importing port.[8] Port to port distances are calculated using the U.S. National Imagery and Mapping Agency Distances Between Ports publication and the online resource Port World Distance Calculator. There are 408 country pairs at the total level of trade. A list of each country's main ports with associated sources in support of their domestic ranking, which is mainly based on throughput levels is available in the Technical Appendix. The indirect capital distance matrix is available upon request.

– *Assignment of Hubs*

Unfortunately there is no resource that describes the routes utilised to ship goods from origin to destination by transport mode. This leads to ambiguity as to the existence, followed by the identification of good-specific routes and their respective distance. Thus the distances considered herein are country pair specific.

The lack of routes also makes the selection of unanimously accepted hub locations difficult. Any location benefited either from geographical advantage or historical precedence can serve as a hub for a particular hinterland or as a transshipment hub (Krugman, 1993).[9]

I assess whether each indirect distance observation passes through one or more exogenously imposed locations.[10] These locations serve as hubs, which are big enough to potentially affect route configurations. They are selected on the basis of i) geographic centrality, ii) throughput and information regarding their status as a transshipment hub and are as follows:

---

[8] For instances where the country is landlocked the closest foreign major port is chosen.

[9] For example the port of Hong Kong ranks second globally in terms of container traffic (Rua, 2014), serves as a mediator for China's trade to the rest of the world (Feenstra and Hanson, 2004) yet information about the non-Chinese origin and type of goods in transit to other destinations via the port is scarce and fragmented.

[10] In Fugazza (2015) transshipment points are also exogenously obtained after solving Dijkstra's version of the shortest path algorithm. I consider a similar approach as a robustness check in section 5.2.

Figure 1: Hub locations, direct and indirect capital distance

- The ports of Arica and Paranagua (UNCTAD Secretariat, 1994).
- The area around Cape Town (African Development Bank, 2010).
- The Gibraltar area (Notteboom, Parola and Satta, 2014).
- Istanbul (Merk and Bagis, 2013).
- Panama Canal and Port Said (Notteboom, 2012; African Development Bank, 2010).
- Singapore (Port of Singapore Authority; Cullinane, Yap and Lam (2006)).

An indirect distance observation which involves passing from these areas obtains $\text{Hub}_{ij} = 1$, while no such distinction is made for the same direct distance observations. An example is shown in Figure 1 where hollow circles refer to hub locations. The direct capital distance between Beijing and Brasilia is 16,948 km. Indirectly, the distance from Beijing to Shanghai is 1,267 km, from Shanghai to Singapore 3,934 km and the indicator is assigned a value of 1. Then I add the distance from Singapore to Rio (16,366 km) and Rio to Brasilia (1,160 km), totalling 22,727 km. I also keep track of the number of hubs required to reach the destination (2). These are used in a specification described in section 5.2 that departs from the strict confines of the model's prediction and involves estimating the marginal effect of indirect distance and the number of hubs on a route.

The sample's time period is short enough to ensure that there are no changes to the

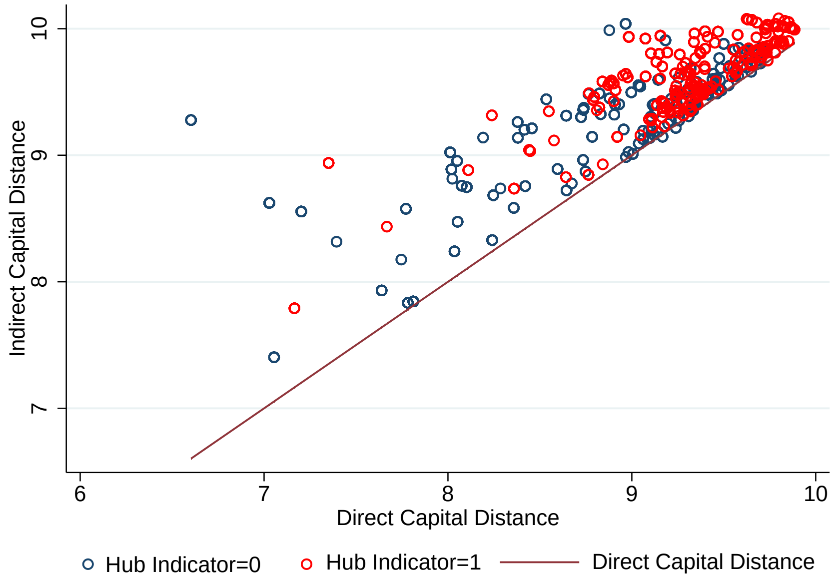


Figure 2: Distribution of distances and hubs

hub status of the selected locations.[11] Since the distance variables are not country-pair-commodity specific, it is possible that the number of chapters or subheadings for specific pairs gives a disproportional weight to certain routes and hub(s). For this reason I incorporate commodity fixed effects at both HS 2 and 6 levels.

The data have a weakness as transport costs for individual E.U. 15 countries are not available. Variables are aggregated to provide approximations at this level when necessary. The capital distance of the E.U. 15 area is measured from Brussels, the main export/import port is Rotterdam. Estimations are conducted with and without the presence of the E.U. 15 area and its biggest trading partner in the data, the United States, to ensure this is not the source of the findings.

The distribution of direct and indirect distances is shown in Figure 2. It reveals a systematic bias of hubs being used on longer routes and a diminishing dispersion of indirect distance over longer distances.

Table 1 contains the frequency of flows transiting a hub location at three aggregation levels. Particular routes appear to be voluminous as the number of commodities within a flow determines the volume of flows that pass from a hub: At the total trade level 48% of flows used one or more hub locations, but this number declines to 35% and 24% for the HS

---

[11] Estimations using shorter time periods did not alter the outcomes.

Table 1: Number of hubs required for flows to reach the destination.

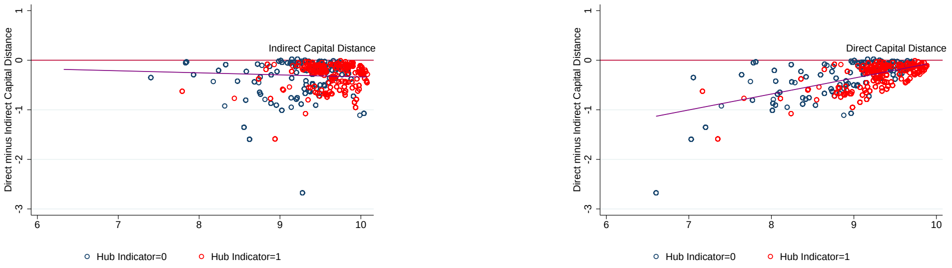
						Total Trade Level				Hub=1	Hub=0	Pass through	Reach destination
		Arica	Cape Town	Gibraltar	Istanbul	Panama Canal	Paranagua	Port Said	Singapore	Total			
1st Hub	Flows	123	197	782	46	283	48	325	217	2,021	2,175	631	1,390
	(/100)	0.06	0.10	0.39	0.02	0.14	0.02	0.16	0.11	1			0.71
2nd Hub		7	161	123	0	140	0	139	9	579		52	527
	(/100)	0.01	0.28	0.21	0	0.24	0	0.24	0.02	1			0.27
3rd Hub		29	0	0	0	19	4	0	0	52			52
	(/100)	0.56	0	0	0	0.37	0.08	0	0	1			0.03
Total		159	358	905	46	442	52	464	226	2,652			
(/100)		0.06	0.13	0.34	0.02	0.17	0.02	0.17	0.09	1			
Ranking		6	4	1	8	3	7	2	5				

						HS 2 Level				Hub=1	Hub=0	Pass through	Reach destination
		Arica	Cape Town	Gibraltar	Istanbul	Panama Canal	Paranagua	Port Said	Singapore	Total			
1st Hub	Flows	2,301	5,836	21,262	1,431	7,204	448	11,897	11,357	61,736	116,152	20,463	41,273
	(/100)	0.04	0.09	0.34	0.02	0.12	0.01	0.19	0.18	1			0.68
2nd Hub		29	9,015	6,726	0	2,421	0	1,197	10	19,398		1,065	18,333
	(/100)	0.001	0.46	0.35	0	0.12	0	0.06	0.001	1			0.30
3rd Hub		225	0	0	0	809	31	0	0	1,065			1,065
	(/100)	0.21	0	0	0	0.76	0.03	0	0	1			0.02
Total		2,555	14,851	27,988	1,431	10,434	479	13,094	11,367	82,199			
(/100)		0.03	0.18	0.34	0.02	0.13	0.01	0.16	0.14	1			
Ranking		6	2	1	7	5	8	3	4				

						HS 6 Level				Hub=1	Hub=0	Pass through	Reach destination
		Arica	Cape Town	Gibraltar	Istanbul	Panama Canal	Paranagua	Port Said	Singapore	Total			
1st Hub	Flows	8,543	19,799	120,229	4,977	44,210	631	66,147	123,307	387,843	1,261,195	173,109	214,734
	(/100)	0.02	0.05	0.31	0.01	0.11	0	0.17	0.32	1			0.56
2nd Hub		23	107,413	49,782	0	7,591	0	4,562	20	169,391		3,718	165,673
	(/100)	0	0.63	0.29	0	0.04	0	0.03	0	1			0.43
3rd Hub		385	0	0	0	3,245	88	0	0	3,718			3,718
	(/100)	0.10	0	0	0	0.87	0.02	0	0	1			0.01
Total		8,951	127,212	170,011	4,977	55,046	719	70,709	123,327	560,952			
(/100)		0.02	0.23	0.30	0.01	0.10	0.001	0.13	0.22	1			
Ranking		6	2	1	7	5	8	4	3				

2 and 6 levels. Approximately 65%, 33% and 2% of flows involve one, two and three hub locations respectively.

Figures 3a and 3b show how the difference between two distance variables increases slightly in indirect distance, but decreases in direct distance implying that direct distance observations approach in magnitude their indirect distance counterparts over longer distances in the sample. By overlapping the two distributions of distance I infer that the right tail of the distribution of direct distance approximates that of indirect distance. Consequently the variance of direct distance must be larger. Figure 4 plots the fitted size adjusted exports at the total trade level against the two distance variants plus indirect distance interacted with the hub indicator variable. The larger variance of direct distance may result in a smaller coefficient than indirect distance but indirect distance interacted with the indicator



(a) As a function of indirect distance (logs). (b) As a function of direct distance (logs).

Figure 3: Distance differentials

yields an even smaller coefficient.

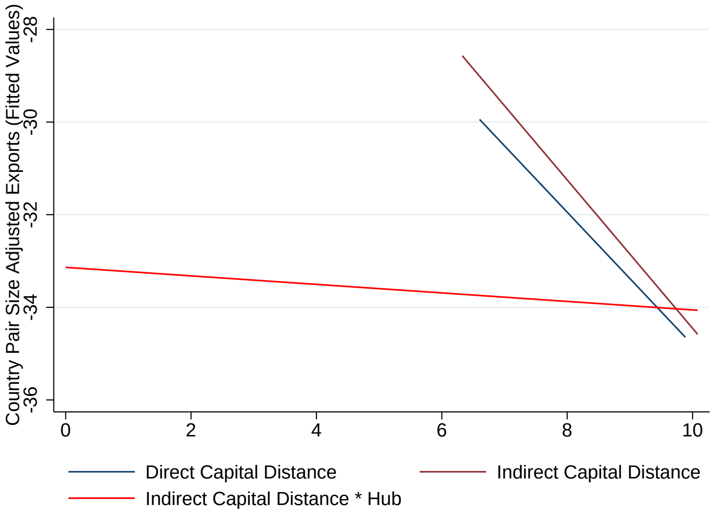


Figure 4: Size-adjusted exports as a function of distance (Total Trade Level, logs)

5 Empirical Results

5.1 Main Results

In all tables columns contain results corresponding to specific combinations of country, year, commodity fixed effects. Columns (1) - (6) and (7) - (12) pertain to the full sample and a sub-sample excluding the E.U. 15 - United States pair respectively. The tables containing the estimation results inclusive of all control variables and standard errors are relegated to the Technical Appendix.[12]

---

[12] I additionally carried out estimations using combinations of two and three control variables; the results validate those presented in the paper and are available upon request.

In Table 2 I compare the two estimated distance coefficients from applying equation 4 to the same sample. The null hypothesis states that the effect of the two distance variables on exports is indistinguishable. A t-test and a Paternoster test reveal that the null is not rejected in the majority of specifications across aggregation levels bar three exceptions: Columns (7) and (8) of the total trade and (8) of the HS 2 level (Paternoster, Brame, Mazerolle and Piquero, 1998). The coefficients of indirect distance appear weakly larger in absolute value than their direct distance counterparts in the full sample when the strictest combination of fixed effects[13] are employed but this is not consistently observable across other specifications and when excluding the E.U. 15 - United States pair. The goodness of fit across specifications in each column is identical up to and sometimes including the third decimal place. The coefficients of direct distance are in the range of surveyed estimates of the literature (Disdier and Head, 2008; Overman, Redding and Venables, 2001). The magnitudes of the indirect distance coefficient, although not comparable as the literature considers only direct distance, do not diverge much from this range due to their indistinguishable effect and similar levels to direct distance coefficients.

Table 3 shows the results from estimating gravity equations 4 and 5. The first row is a replication of the direct distance coefficients of Table 2. The second row of each aggregation level contains the marginal effect of indirect distance plus the interaction with the hub indicator of equation 5 for the same sample as the row above it. Entries in the third row are the percentage differences between exporting indirectly using a hub and directly, calculated as $\left[\frac{\hat{\beta}_{\text{Indirect Capital Distance}}+\hat{\beta}_{\text{Interaction}}}{\hat{\beta}_{\text{Direct Capital Distance}}}-1\right]*100$. Dashes pertain to a lack of result because of an insignificant interaction term. The last two rows contain p-values of two tests. The null hypothesis B is that the joint impact of the coefficient of indirect distance and the interaction is zero. The null hypothesis C states that the coefficient of direct distance is not statistically different from the coefficient of indirect distance plus the interaction coefficient.

When the interaction is significant, F-tests support that the null hypothesis B is rejected across specifications and aggregation levels. I find that in most cases where the interaction is significant, the joint impact of indirect distance and its interaction with the hub indicator

---

[13] (ex)importer-commodity and year, or (ex)importer-commodity-year fixed effects.

Table 2: Exports: Coefficients of direct distance and indirect distance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-1.5***	-1.473***					-1.408***	-1.412***				
Indirect Capital Distance	-1.4***	-1.37***					-1.144***	-1.14***				
R^2_{Direct}	0.937	0.947					0.912	0.932				
$R^2_{Indirect}$	0.932	0.942					0.908	0.929				
p-value Ho A	0.469	0.489					0.0195	0.0304				
p-value Ho A (P. T.)	0.565	0.599					0.139	0.179				
HS 2 Level												
Direct Capital Distance	-1.077***	-1.066***	-1.248***	-1.239***	-1.62***	-1.649***	-1.172***	-1.212***	-1.29***	-1.333***	-1.772***	-1.842***
Indirect Capital Distance	-1.11***	-1.105***	-1.325***	-1.323***	-1.77***	-1.829***	-1.048***	-1.049***	-1.196***	-1.203***	-1.61***	-1.655***
R^2_{Direct}	0.376	0.384	0.550	0.558	0.801	0.850	0.240	0.255	0.436	0.452	0.726	0.815
$R^2_{Indirect}$	0.374	0.382	0.548	0.557	0.801	0.851	0.239	0.253	0.436	0.451	0.724	0.814
p-value Ho A	0.768	0.712	0.541	0.470	0.218	0.160	0.169	0.0492	0.363	0.168	0.140	0.128
p-value Ho A (P. T.)	0.812	0.780	0.634	0.598	0.415	0.419	0.365	0.231	0.553	0.394	0.320	0.309
HS 6 Level												
Direct Capital Distance	-0.509***	-0.477***	-0.704***	-0.673***	-1.275***	-1.387***	-0.331***	-0.328***	-0.513***	-0.5***	-1.088***	-1.168***
Indirect Capital Distance	-0.414***	-0.396***	-0.684***	-0.667***	-1.511***	-1.684***	-0.305***	-0.283***	-0.482***	-0.459***	-1.024***	-1.082***
R^2_{Direct}	0.225	0.232	0.377	0.384	0.727	0.844	0.095	0.109	0.295	0.306	0.660	0.819
$R^2_{Indirect}$	0.223	0.230	0.375	0.382	0.726	0.845	0.095	0.109	0.295	0.306	0.659	0.818
p-value Ho A	0.450	0.494	0.890	0.963	0.258	0.226	0.742	0.525	0.694	0.555	0.640	0.642
p-value Ho A (P. T.)	0.496	0.550	0.901	0.968	0.346	0.355	0.812	0.655	0.784	0.690	0.743	0.744
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of direct capital distance and indirect capital distance and measures of goodness of fit from the estimation of equation 5 in the presence of controls for contiguity, common language, colony and common regional trade agreements. Ho A: Direct capital distance=Indirect capital distance. P. T. = Paternoster test. The regression output for each trade level and column is available in the Technical Appendix. *** p<0.01, ** p<0.05, * p<0.1.

is lower in absolute value than the impact of direct distance.

In all cases of the total trade and in columns (6) - (12) of the other aggregation levels, the null hypothesis C is rejected and I conclude that the joint impact is significantly less than the impact of direct distance. The hub indicator variable appears to be dampening the impact of indirect distance, while the goodness of fit is identical up to and sometimes including, the third decimal place.

The specifications of column 12 in Tables 2 and 3 only allow country pair variation. The results in Table 2 show that a 10% increase in direct distance reduces export volumes ceteris paribus by about 18% and 11% at the HS 2 and 6 levels respectively. When trading indirectly without assuming any presence of a hub the marginal effect on exports stands at about 16% and 10% but this difference is not statistically significant when i) one of the two coefficients is perceived as a parameter (t-test); ii) both are treated as sample estimates (Paternoster test). When I include the interaction with the hub indicator in the corresponding specifications

Table 3: Exports: Coefficients of direct distance, indirect distance plus the interaction term.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-1.5	-1.473					-1.408	-1.412				
Indirect Capital Distance + Interaction	-1.043	-1.021					-1.023	-1.034				
% Discount from trading via hub(s)	30.46	30.67					–	–				
R^2_{Direct}	0.937	0.947					0.912	0.932				
$R^2_{Indirect}$	0.934	0.944					0.909	0.929				
p-value Ho B	0	0					0	0				
p-value Ho C	0	0					0	0.01				
HS 2 Level												
Direct Capital Distance	-1.077	-1.066	-1.248	-1.239	-1.620	-1.649	-1.172	-1.212	-1.290	-1.333	-1.772	-1.842
Indirect Capital Distance + Interaction	-0.969	-0.995	-1.072	-1.099	-1.477	-1.554	-0.892	-0.931	-0.949	-0.991	-1.350	-1.404
% Discount from trading via hub(s)	–	–	14.14	11.30	8.841	5.745	23.99	23.23	26.50	25.63	23.82	23.77
R^2_{Direct}	0.376	0.384	0.550	0.558	0.801	0.850	0.240	0.255	0.436	0.452	0.726	0.815
$R^2_{Indirect}$	0.375	0.382	0.549	0.558	0.801	0.851	0.239	0.253	0.436	0.452	0.725	0.814
p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
p-value Ho C	0.426	0.579	0.2	0.278	0.339	0.571	0.008	0.007	0.004	0.003	0.003	0.013
HS 6 Level												
Direct Capital Distance	-0.509	-0.477	-0.704	-0.673	-1.275	-1.387	-0.331	-0.328	-0.513	-0.500	-1.088	-1.168
Indirect Capital Distance + Interaction	-0.459	-0.483	-0.649	-0.669	-1.268	-1.354	-0.293	-0.293	-0.378	-0.379	-0.754	-0.795
% Discount from trading via hub(s)	–	–	–	–	–	–	–	–	26.38	–	30.79	31.89
R^2_{Direct}	0.225	0.232	0.377	0.384	0.727	0.844	0.095	0.109	0.295	0.306	0.660	0.819
$R^2_{Indirect}$	0.223	0.230	0.375	0.382	0.726	0.845	0.095	0.109	0.295	0.306	0.660	0.819
p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
p-value Ho C	0.751	0.97	0.752	0.982	0.975	0.903	0.623	0.621	0.089	0.104	0.022	0.075
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of direct capital distance, indirect capital distance plus the interaction term with the hub indicator variable and measures of goodness of fit from the estimation of equations 5 and 6, in the presence of controls for contiguity, common language, colony and common regional trade agreements. Ho B: Indirect capital distance=Interaction=0. Ho C: Direct capital distance=Indirect capital Distance+interaction. The regression output for each trade level and column is available in the Technical Appendix.

of Table 3, the overall marginal effect of indirect distance reduces to about 14% and 8%. Compared to trading directly, trading via a hub results in a 23% and 32% less reduction in exports for every doubling of distance respectively. Generalising the outcome, these savings stand at 30%, 19% and 29% on average for each of the aggregation levels considered.

The size and signs of the distance variables and controls are in accordance with the literature.[14] However identification of the size and sign of the hub indicator coefficient when interacted with distance is unresolved.

---

[14] E.g. Limão and Venables (2001).

A possible explanation for these results is that the hub indicator is correlated positively with distance and by extrapolation negatively correlated with size adjusted exports. The negative correlation weakens over longer distances and turns positive possibly capturing the high volume of trade between East Asian and Atlantic countries. Such flows occur exclusively via Singapore or the Suez Canal and the coefficient of the interaction is picking up some of the pair specific variation associated with longer, voluminous flows which would otherwise be captured by the distance variable. The implication is that the distance variable appears to capture information not only associated with remoteness but also with route selection as inferred from the apparently beneficial impact of hubs. The revealed choice to trade via a hub location seems to have empirical support as it has a positive effect especially over longer distances, confirming the models' propositions.

– *Endogeneity*

So far I assumed that the interaction term which possibly captures information about route selection is exogenous. In reality, the revealed choice of firms to trade or not via hub locations implies that the outcome variable, exports, is correlated with the interaction. If thus distance coefficients capture information about routing decisions, then what is the effect of remoteness?

Performing a two stage least squares regression by instrumenting the dummy variable with a candidate instrument Z would result in a "forbidden regression".[15] I avoid this by relegating the interaction of equation 5 in the error term $\tilde{\xi}^s_{ij} = -\iota \text{Hub}_{ij} \ln(d_{ik} + d_{kj}) + \xi^s_{ij}$.

Since the effects of the two distance variables are indistinguishable based on the results contained in Table 2, I estimate $\beta^D_3(\sigma{-}1)$, $\beta^I_3(\sigma{-}1)$ by treating the interaction as unobserved. A candidate instrument that satisfies the exclusion restriction $Cov(\tilde{\xi}^s_{ij}, z)$ and produces linear fitted values of the first stage dependent variable is time difference. The instrument captures

---

[15] The first stage requires a non linear estimation on the hub indicator $\text{Hub}_{ij} = \Phi\left(\pi_1' X_{ij} + \pi_2' Z_{ij}\right)$, where Φ is the cumulative distribution function, to obtain the non linear fitted values $\widehat{\text{Hub}}_{ij}$. In the second stage I would substitute the indicator observations with the fitted values and interact the variable with indirect capital distance. The residuals would most likely be correlated with the fitted values and other covariates when applying ordinary least squares. To avoid this occurrence the fitted values $\widehat{\text{Hub}}_{ij}$ could be used as instruments themselves for Hub_{ij}. This would mean the second (linear) stage relies, albeit indirectly, on a non linear source of information (Angrist and Pischke, 2009).

Table 4: Exports: 2SLS impacts of direct and indirect capital distance using time difference as an instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-2.937***	-2.861***					-2.876***	-2.795***				
R^2_{Direct}	0.914	0.925					0.889	0.911				
Underidentification Test	16.15	19.55					10.44	13.58				
Cragg-Donald	195.4	165.5					178	157.1				
Indirect Capital Distance	-3.057***	-3.109***					-2.979***	-2.997***				
$R^2_{Indirect}$	0.905	0.909					0.863	0.880				
Underidentification Test	7.523	7.461					5.392	5.732				
Cragg-Donald	200.2	149					131.5	100.7				
p-value Ho A	0.902	0.825					0.920	0.856				
p-value Ho A (P. T.)	0.897	0.787					0.932	0.860				
HS 2 Level												
Direct Capital Distance	-1.508***	-1.549***	-2.070***	-2.075***	-3.442***	-3.355***	-1.503***	-1.538***	-1.855***	-1.868***	-2.937***	-2.849***
R^2_{Direct}	0.374	0.381	0.542	0.551	0.768	0.820	0.239	0.254	0.432	0.449	0.714	0.807
Underidentification Test	10.32	13.48	10.18	13.36	8.945	14.49	13.83	18.82	13.75	18.81	13.32	21.42
Cragg-Donald	7431	9563	7271	9385	5450	5015	11924	16131	11723	15906	9508	8800
Indirect Capital Distance	-1.128***	-1.231***	-1.543***	-1.646***	-2.428***	-2.590***	-1.330***	-1.414***	-1.641***	-1.718***	-2.503***	-2.608***
$R^2_{Indirect}$	0.374	0.382	0.548	0.557	0.797	0.846	0.238	0.251	0.433	0.448	0.716	0.805
Underidentification Test	12.75	13.87	12.70	13.82	12.05	16.53	11.89	13.50	11.86	13.48	12.10	16.74
Cragg-Donald	16993	18904	16736	18631	14105	10745	13604	16095	13388	15863	11936	9071
p-value Ho A	0.345	0.418	0.227	0.313	0.0908	0.220	0.540	0.653	0.480	0.603	0.328	0.597
p-value Ho A (P. T.)	0.518	0.555	0.502	0.532	0.451	0.510	0.704	0.766	0.732	0.782	0.651	0.767
HS 6 Level												
Direct Capital Distance	-0.0393	-0.197	-0.507	-0.639*	-1.896***	-2.107***	0.0875	-0.0537	-0.272*	-0.386***	-1.440***	-1.432***
R^2_{Direct}	0.222	0.231	0.377	0.384	0.723	0.839	0.093	0.109	0.294	0.306	0.659	0.819
Underidentification Test	11.46	15.89	11.39	15.98	11.40	23	15.23	21.45	15.35	22.22	12.10	29.20
Cragg-Donald	98783	142054	93474	135301	62577	40127	99486	140984	92198	131685	48770	26417
Indirect Capital Distance	-0.0293	-0.161	-0.376	-0.519	-1.366***	-1.850***	0.0798	-0.0515	-0.248*	-0.370***	-1.199***	-1.461***
$R^2_{Indirect}$	0.221	0.230	0.374	0.382	0.726	0.845	0.093	0.109	0.294	0.306	0.659	0.817
Underidentification Test	20.26	22.89	20.30	23	23.01	31.35	15.62	17.86	15.67	18.04	16.63	26.26
Cragg-Donald	275177	322512	263284	309032	190413	80088	119403	145506	110787	135625	69884	23507
p-value Ho A	0.978	0.912	0.720	0.727	0.149	0.506	0.964	0.989	0.854	0.907	0.222	0.921
p-value Ho A (P. T.)	0.988	0.945	0.822	0.805	0.467	0.701	0.978	0.992	0.902	0.935	0.699	0.951
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported 2SLS coefficients of direct capital distance and indirect capital distance and measures of goodness of fit from the estimation of equation 5 in the presence of controls for contiguity, common language, colony and common regional trade agreements. Country pairs with no time difference are assigned a minute value such that the logarithm is defined. Ho A: Direct capital distance=Indirect capital distance. P. T. = Paternoster test. The first and second stage regression output for each trade level and column is available in the Technical Appendix.

characteristics related to remoteness and information frictions of trading partners as in Egger and Larch (2013), hence its relevance, but does not contain information about which routes should be used to export commodities. Time difference observations are sourced from the CEPII Gravity dataset (Head, Mayer and Ries, 2010). I consider two distinctions for countries that have no time difference: i) The zero entry is left as is and ii) is replaced by an infinitessimal number. For brevity I present the second stage outcomes of case ii) in Table 4 while case i) and all first stage outcomes are available in the Technical Appendix.

The underidentification test and Cragg-Donald statistics show that across aggregation levels the instrument is relevant, not weak and the specification is exactly identified. Goodness of fit measures at HS 2 and HS 6 levels of trade are identical up to and sometimes including, the third decimal place. The coefficients of the two distance variables are significant in specifications using the strictest combination of fixed effects. The magnitudes are inflated compared with Table 2 which may be attributed to the small but significant first stage coefficients because I do not observe this when zero time differences are not replaced.

Indirect distance coefficients' magnitudes are overall weakly smaller in absolute value compared to those of the direct distance coefficients. The p-values of the t and Paternoster tests indicate that they are not significantly different, as in Table 2. I conclude that there appears to be evidence in support of routing decisions contained in the error term. If indeed the instrument satisfies the exclusion restriction, the reported distance coefficients capture variation devoid of at least routing decisions.

5.2 Robustness checks

So far the econometric experiments did not stray from the confines imposed by equations 4 and 5. In this section I address concerns of mispecification, data quality and the choice of hubs.

Comparing different distance variables across specifications might be a misspecified approach and an erroneous identifying strategy to check for the existence of a network effect. I consider a specification involving only indirect distance and distinguish between routes that do not involve a hub versus the complement, which correspond to rows I and II in Table 5 situated in Appendix C together with all remaining tables. Further, if there are hubs present on a route I add their number as a separate regressor and also interact it with distance in rows IIa and IIb. I use the same controls as in the previous regressions. The results show that exports tend to respond less to changes in distance when a hub location is not involved compared to the alternative. For pairs that have at least one hub intervening, the distance coefficients are dampened by significant interactions in IIb. The number of hubs as a separate regressor does not appear to have any explanatory power as only four less strict specifications (4, 8-10) have positively significant coefficients. The outcomes appear to support the main results.

The specification resembles that of Fugazza (2015), however he omits the interaction between distance and the number of transshipments. The latter's negative effect on country level exports diminishes in the number of transshipments, whilst I find an insignificant impact with mixed signs on country level (total trade) exports. We both confirm that inclusion of the number of transshipment hubs (and the interaction) reduces the impact of the indirect distance variable and making it less than the impact of indirect distance without involving transshipment hubs. Further comparisons between the two studies are not possible as Fugazza (2015) measures the impacts of transshipment hubs and maritime distance on country level exports only. It is not possible to know whether this situation is advantageous or not compared to the same partners trading directly without transshipment hubs intervening.

Regarding the quality of the data, I consider an alternative, artificial, indirect distance matrix constructed using the A* algorithm and repeat all experiments. The matrix contains the shortest routes between all locations, where the hub indicator is equal to one if the route passes within a few kilometres from every location that is a canal, passage or a transshipment hub, ultimately favouring geographic centrality rather than economic activity. Further details regarding the distance matrix are contained in Appendix C. Table 6 shows that the coefficients of indirect distance are lower than those of direct distance without considering the presence of hubs. The difference is statistically significant in most cases. In Table 7 the hub interaction term has a negative sign in all specifications when significant. Joint effects, when significant, have a magnified impact in absolute terms compared to the indirect distance coefficients of Table 6. Any discount from trading via a hub is due to the lower impact of indirect distance and not the interaction term. When instrumenting indirect distance with time difference, all distance coefficients are similar and statistically indistinguishable, except for column (12) of Table 8.

In Table 9 the number of hubs on a route have a positive effect when interacted with distance. There are instances where the number of hubs as a separate regressor has a negative and significant impact similar to Fugazza (2015). I conclude that utilising an alternative, artificial, distance matrix and choice of hubs results in weak support of the model's predictions. Further research is thus warranted possibly by collecting information on the geography of routes at the product level and establishing any product specific transshipment locations. To my knowledge such information is not currently available.

6 Conclusion

I merge a simple homogeneous firm trade model with the symmetric connections model of Jackson and Wolinsky (1996) to derive a set of specifications that identify the effect of transshipment hubs being present on trade routes. Hubs are formed by virtue of geographical centrality and peripheral countries' potential to connect to them, manifested through levels of transport costs and infrastructure.

To identify whether the mere existence of hubs impacts trade, I treat the observed level of exports as a revealed outcome of the choice of route used to ship goods. I compare the fit of two capital distance variables, one direct, the other indirect which is interacted with an indicator variable that signifies the presence of a hub location on a route. I find that i) remote countries have on average 26% less impact of distance on exports if they trade via a hub compared to trading directly; ii) distance variables capture information about routing decisions.

I conclude that geographically disadvantaged countries absorbing high transport costs achieve a more beneficial trading position when forming a transport network consisting of a hub. The connection with at least one proximal geographically advantaged partner improves market access, ameliorates exposures to these costs and leads to improvements in own and transit infrastructure.

Further research in this area could focus on the analyses of heterogeneity in fixed costs associated with transport infrastructure, and the geography of trade routes at the product level for which data is not currently available. This information would allow the calibration of a model that simulates the true data generating process and allows computation of counterfactuals.

References

African Development Bank (2010), African Development Report 2010. Oxford University Press.

Allen, T. and Arkolakis, C. (2014), 'Trade and the topography of the spatial economy', *The Quarterly Journal of Economics* **129**(3), 1085 – 1140. doi: http://doi.org/10.1093/qje/qju016.

Amjadi, A., Reinke, U. and Yeats, A. (1996), Did external barriers cause the marginalisation of Sub-Saharan Africa in world trade? Policy Research Working Paper, World Bank, No. 1586.

Angrist, J. and Pischke, J. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Bernhofen, D. M., El-Sahli, Z. and Kneller, R. (2015), 'Estimating the effects of the container revolution on world trade', *Journal of International Economics* pp. –. doi: http://doi.org/10.1016/j.jinteco.2015.09.001.

Besedeš, T. and Murshid, A. P. (2015), Experimenting with ash: The trade-effects of airspace closures in the aftermath of Eyjafjallajökull. Working Paper.

Chaney, T. (2005), Distorted gravity: Heterogeneous firms, market structure and the geography of international trade. Job Market Paper.

Clark, X., Dollar, D. and Micco, A. (2004), 'Port efficiency, maritime transport costs, and bilateral trade', *Journal of Development Economics* **75**(2), 417 – 450. doi: http://doi.org/10.1016/j.jdeveco.2004.06.005.

Cullinane, K., Yap, W. Y. and Lam, J. S. (2006), 'The port of Singapore and its governance structure', *Research in Transportation Economics* **17**, 285 – 310. doi: http://doi.org/10.1016/S0739-8859(06)17013-4.

Disdier, A.-C. and Head, K. (2008), 'The puzzling persistence of the distance effect on bilateral trade', *The Review of Economics and Statistics* **90**(1), 37 – 48. doi: http://doi.org/10.1162/rest.90.1.37.

Egger, P. H. and Larch, M. (2013), 'Time zone differences as trade barriers', *Economics Letters* **119**(2), 172 – 175. doi: http://doi.org/10.1016/j.econlet.2013.02.016.

Feenstra, R. C. and Hanson, G. H. (2004), 'Intermediaries in entrepôt trade: Hong Kong re-exports of chinese goods', *Journal of Economics & Management Strategy* **13**(1), 3 – 35. doi: http://doi.org/10.1111/j.1430-9134.2004.00002.x.

Feyrer, J. (2011), Distance, trade, and income – The 1967 to 1975 closing of the Suez Canal as a natural experiment. Society for Economic Dynamics, 2011 Meeting Papers, no. 1438.

Fugazza, M. (2015), Maritime connectivity and trade. UNCTAD Policy Issues in International Trade and Commodities Research Study Series No. 70.

Greene, W. H. (2008), *Econometric Analysis*, 6th edn, Pearson Education, New Jersey.

Harrigan, J. (2010), 'Airplanes and comparative advantage', *Journal of International Economics* **82**(2), 181 – 194. doi: http://doi.org/10.1016/j.jinteco.2010.07.002.

Head, K., Mayer, T. and Ries, J. (2010), 'The erosion of colonial trade linkages after independence', *Journal of International Economics* **81**(1), 1 – 14. doi: http://doi.org/10.1016/j.jinteco.2010.01.002.

Hendricks, K., Piccione, M. and Tan, G. (1995), 'The economics of hubs: The case of monopoly', *The Review of Economic Studies* **62**(No. 1), 83 – 99. doi: http://doi.org/10.2307/2297842.

Hummels, D. (2007), 'Transportation costs and international trade in the second era of globalization', *Journal of Economic Perspectives* **21**(3), 131 – 154. doi: http://doi.org/10.1257/jep.21.3.131.

Hummels, D., Lugovskyy, V. and Skiba, A. (2009), 'The trade reducing effects of market power in international shipping', *Journal of Development Economics* **89**(1), 84 – 97. doi: http://doi.org/10.1016/j.jdeveco.2008.05.001.

Hummels, D. and Skiba, A. (2004), 'Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture', *Journal of Political Economy* **112**(6), 1384 – 1402. doi: http://doi.org/10.1086/422562.

Irarrazabal, A., Moxnes, A. and Opromolla, L. (2015), 'The tip of the iceberg: A quantitative framework for estimating trade costs', *The Review of Economics and Statistics* **97**(4), 777 – 792.

Jackson, M. O. (2003), A survey of models of network formation: Stability and efficiency, Game theory and information, EconWPA.

Jackson, M. O. and Wolinsky, A. (1996), 'A strategic model of social and economic networks', *Journal of Economic Theory* **71**(1), 44 – 74. doi: http://doi.org/10.1006/jeth.1996.0108.

Korinek, J. and Sourdin, P. (2010), 'Clarifying trade costs: Maritime transport and its effect on agricultural trade', *Applied Economic Perspectives and Policy* **32**(3), 417 – 435. doi: http://doi.org/10.1093/aepp/ppq007.

Krugman, P. (1993), The hub effect: Or, threeness in interregional trade, *in* 'Theory, Policy and Dynamics in International Trade', Cambridge University Press, New York, pp. 29 – 37.

Levinson, M. (2008), *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*, Princeton University Press.

Limão, N. and Venables, A. (2001), 'Infrastructure, geographical disadvantage and transport costs.', *The World Bank Economic Review* **15**, 451 – 479. doi: http://doi.org/10.1093/wber/15.3.451.

Lux, M. (2011), Defying gravity: The substitutability of transportation in international trade, MPRA working paper, University Library of Munich, Germany.

Maritime and Port Authority of Singapore (2018), `http://www.mpa.gov.sg/web/portal/home/maritime-singapore/introduction-to-maritime-singapore/facts-and-trivia`. Accessed: 2018-01-20.

Martincus, C. V., Carballo, J., Garcia, P. M. and Graziano, A. (2014), 'How do transport costs affect firms exports? Evidence from a vanishing bridge', *Economics Letters* **123**(2), 149 – 153. doi: http://dx.doi.org/10.1016/j.econlet.2014.01.030.

Mayer, T. and Zignago, S. (2011), Notes on CEPII's distances measures: The GeoDist database, Working Papers 2011-25, CEPII.

Melitz, M. J. and Redding, S. J. (2014), Heterogeneous firms and trade, *in* E. Helpman, K. Rogoff and G. Gopinath, eds, 'Handbook of International Economics', Vol. 4 of *Handbook of International Economics*, Elsevier, pp. 1 – 54.

Merk, O. and Bagis, O. (2013), 'The Competitiveness of Global Port-Cities: The Case of Mersin - Turkey', *OECD Regional Development Working Papers* (1). doi: http://doi.org/10.1787/5k4c43014plt-en.

Mori, T. (2012), 'Increasing returns in transportation and the formation of hubs', *Journal of Economic Geography* **12**(4), 877. doi: 10.1093/jeg/lbr028.

Notteboom, T. E. (2012), 'Towards a new intermediate hub region in container shipping? Relay and interlining via the Cape route vs. the Suez route', *Journal of Transport Geography* **22**, 164 – 178. doi: http://doi.org/10.1016/j.jtrangeo.2012.01.003.

Notteboom, T., Parola, F. and Satta, G. (2014), State of the European port system - Market trends and structure update. EU PORTOPIA, D, 1.1, DT, 2013.05.11, working paper.

OECD Trade and Agriculture Directorate (2008), Clarifying trade costs in maritime transport. OECD Trade Committee Working Paper.

Oum, T. H., Zhang, A. and Zhang, Y. (1995), 'Airline network rivalry', *The Canadian Journal of Economics* **28**(4a), 836 – 857. doi: http://doi.org/10.2307/135934.

Overman, H. G., Redding, S. J. and Venables, A. J. (2001), The economic geography of trade production and income: A survey of empirics. Discussion Paper, CEPR, No. 2978.

Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998), 'Using the correct statistical test for the equality of regression coefficients', *Criminology* **36**(4), 859 – 866. doi: http://doi.org/10.1111/j.1745-9125.1998.tb01268.x.

Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of International Economics* **62**(1), 53 – 82. doi: http://doi.org/10.1016/j.jinteco.2003.07.001.

Rodrigue, J. P. (2010), Maritime transportation: Drivers for the shipping and port industries. OECD International Transport Forum 2012.

Rua, G. (2014), Diffusion of containerization. Finance and Economics Discussion Series 2014-88. Board of Governors of the Federal Reserve System (U.S.).

Starr, R. and Stinchcombe, M. (1992), Efficient transportation routing and natural monopoly in the airline industry: An economic analysis of hub-spoke and related systems. Discussion paper, University of California, San Diego, No. 92-25.

UNCTAD Secretariat (1994), Transit transport systems for Bolivia and Paraguay, issues actions and constraints. UNCTAD/LDC/91.

UNCTAD Secretariat (2009), Review of maritime transport.

A Exposition of a simple network

A.1 Connection types

For 3 equidistant countries on a plane, one possible type of network is direct connections between all participants, defined as $G = \{12, 21, 13, 31, 23, 32\}$. A second type is indirect connections between 1 and 2 and direct connections from and to country 3, $G = \{13, 31, 23, 32\}$. The net benefit between countries 1 and 2 when directly connected is $v_{12} = \delta - c = v_{21}$. If they are indirectly connected it is $v_{12} = \delta^2 - 0 = v_{21}$ however it is lower than the direct connections case. The indirect connection implies the existence of two costly direct connections: i) The connection between 1 and 3 and ii) between 3 and 2.

A.2 Network participation share

Consider the direct connections network for the 3 equidistant countries. Country 1's set of direct connections is $N_1 = \{12, 13\}$ with cardinality $n_1 = 2$. The total number of direct connections is 6, and country 1's network participation share is $S_1 = 1/3$. Equivalently for the case of an indirect connection between 1 and 2 we have $S_1 = 1/4$, and $F_3 = 1/2$ for country 3 which is directly connected to 1 and 2.

B Hub formations as an outcome of additive trade costs

A three country model with increasing returns in each country's transport sector is constructed in order to prove the existence of a trade off between an increase in distance due to indirect trading and a reduction in fixed costs of transport. This result, albeit derived in a more cumbersome manner, is qualitatively equivalent to the main model's result and assists to arrive at the empirically testable expressions 4 and 5.

– *Description of the Economy*

Instead of a network, it suffices to consider a set of countries $K = \{1, 2, 3\}$ on a plane where country 3 is equidistant from the other two countries, while 1 and 2 are further apart: $d_{13} = d_{23} = d < d_{12} = d'$.

All countries are identical in size L as in the main model. Each country has the two sectors of the main model and in addition a transport sector consisting of monopolistically competitive firms that produce transport services. Similar to Krugman (1993) mobility of labour between the numeraire and the increasing returns sector is allowed, however I impose a fixed share of labour in transport, exogenously setting the sector to full employment.

– *Consumer Demand and Production*

The demand function is identical to the main model, equal to $q_{ij} = A_{ij} p_{ij}^{-\sigma}$ as is the cost function $L^m(q) = \gamma q_{ij} + F_i^m$, $\gamma, F_i^m > 0$, where superscript m distinguishes the manufacturing sector.

– *Trade Costs and Transportation*

Prices of imported varieties comprise a multiplicative cost $\tau \geq 1$ and an additive transport cost f that is the profit maximising price set by transport firms (Hummels and Skiba, 2004; Irarrazabal, Moxnes and Opromolla, 2015): $p_{ij} = p_{ii}\tau_{ij} + f_{ij}$.

Each country's transport sector produces homogeneous transport services that manufacturing firms use as an intermediate input in order to export their products. This means that the transport revenue is obtained from manufacturing firms selling their products abroad, and distribute the revenue to transport firms. The cost function of a transport firm is $L^t(q) = d_{ij} q_{ij} + F_i^t$, where superscript t stands for transport, $d > 0$ is the marginal cost of transport proportional to distance, F_i^t is a varying fixed cost of transport infrastructure. I assume that the quantity each transport firm can carry is equal to the export quantity one manufacturing firm produces.

– *Partial Equilibria*

Profit maximising domestic prices in the manufacturing sector are $p_{ii} = \frac{\sigma}{\sigma-1}\gamma$. Note that in this setting the manufacturing firm can only optimise the domestic price of varieties. It delegates the decision of a connection type along with the associated costs to the transport firm. Manufacturing output stands at $Q_i \equiv \sum_{i,j=1}^{3} q_{ij}\tau_{ij} \equiv \frac{F_i}{\gamma}(\sigma - 1)$. The number of manufacturing firms is $n_i = \frac{\mu_i L_i^m}{\sigma F_i^m}$.

I assume the simultaneous pricing and output decisions of manufacturing and transport firms. Since labour shares are fixed (albeit exogenously) and the wage is equal across sectors, both types of firms have no reason to deviate from their optimal decisions after observing the other type's decision. Transport firms post profit maximising prices $f_{ij} = \frac{\sigma}{\sigma-1}\left(d_{ij} + \frac{\gamma\tau_{ij}}{\sigma-1}\right)$, which are a function of the domestic price of a variety and a markup over the marginal cost of transport due to the elasticity of import demand with respect to additive transport costs $-\frac{\partial q_{ij}}{\partial f_{ij}}\frac{f_{ij}}{q_{ij}} = \sigma\frac{f_{ij}}{p_{ii}\tau_{ij}+f_{ij}}$. The destination price of a variety can be expressed as $p_{ij} = \frac{\sigma}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\gamma\tau_{ij} + d_{ij}\right)$.

– *Consumption ratios*

To simplify the notation, henceforth $\sigma \equiv \frac{1}{1-\theta}$. I further assume that the multiplicative cost $\tau_{ij} = \tau$ is the same across all three countries. Given that γ is also the same, it allows the normalisation $\gamma = \tau = 1$.

The study of hub formations requires observing the fixed costs of transport of one country relative to either of the other two, and so the quantity sold abroad must be expressed relative to the amount sold domestically. Because countries are identical in size and technology, consumption ratios express consumption in a common unit across all three countries.

$$\frac{q_{ij}}{q_{ii}} = \left(\frac{p_{ij}}{p_{ii}}\right)^{\frac{1}{\theta-1}}$$

Firms in country 3 thus observe the ratio:

$$\frac{q_{31}}{q_{33}} = \frac{q_{32}}{q_{33}} = \left(\frac{\frac{\gamma}{\theta}\tau + d}{\gamma}\right)^{\frac{1}{\theta-1}}$$

Countries 1 and 2 when trading between them see:

$$\frac{q_{12}}{q_{11}} = \frac{q_{21}}{q_{22}} = \left(\frac{\frac{\gamma}{\theta}\tau + d'}{\gamma}\right)^{\frac{1}{\theta-1}}$$

and when trading with country 3:

$$\frac{q_{13}}{q_{11}} = \frac{q_{23}}{q_{22}} = \frac{q_{31}}{q_{33}} = \frac{q_{32}}{q_{33}}$$

– *Hub Formations*

So far I have been silent about the equilibrium transported quantity abroad. The zero profit condition in country 3's transport sector together with 3's consumption ratios admit:

$$f_{31} = f_{32} = d + F_3^t (q_{31} + q_{32})^{-1} \Longrightarrow q_{31} = \frac{1}{2} \frac{F_3^t}{\frac{1}{\theta} + d} \frac{\theta}{1-\theta} \tag{6}$$

Because 3 is equidistant from 1 and 2, it is straightforward to see that total exports are split equally between 1 and 2.

For country 1 (and by symmetry for 2), the zero profit condition leads to $f_{12}q_{12} + f_{13}q_{13} = d'q_{12} + dq_{13} + F_1^t$.

Expressing q_{12} in units of q_{13} and characteristing the transport price f, the expression becomes:

$$q_{13} = q_{31} = \frac{F_1^t}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \frac{\theta}{1-\theta} \tag{7}$$

Exports of 1 and 2 when expressed in units of 3, are clearly less than what 3 achieves because of its geographical advantage.

The left hand side of equations 6 and 7 are necessarily the same and so equating the two expressions yields the ratio of fixed costs between 1 and 3 (which is identical for 2 and 3).

$$\frac{F_3^t}{F_1^t} = \frac{2\left(\frac{1}{\theta} + d\right)}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \tag{8}$$

The type of connections, direct or indirect, can be studied through the interplay of fixed costs of transport in the two countries and the distance that separates them. Expression 8 is increasing in d' since $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} > 0$. It is decreasing in d since $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} < 0$.

Proof. Equation 8 has $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} > 0$:

$$\begin{aligned} \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= -\frac{\left[2\left(\frac{1}{\theta} + d\right)\right] \frac{\theta}{\theta-1} \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}-1} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}}{\left[\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}\right]^2} \Longleftrightarrow \\ \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= \frac{F_3^t}{F_1^t} \frac{\frac{\theta}{1-\theta} \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}-1} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} > 0 \end{aligned} \tag{9}$$

Since both fractions are positive. In addition $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} < 0$:

$$\begin{aligned}
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{2}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \\
&- \frac{\left[2\left(\frac{1}{\theta} + d\right)\right]\left(1 + \frac{1}{1-\theta}\left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}\right)}{\left[\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}\right]^2} \Longleftrightarrow \\
&\text{multiplying and dividing the first term with } \left(\tfrac{1}{\theta} + d\right) \text{ yields} \qquad (10) \\
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left[\left(\frac{1}{\theta} + d\right)^{-1} - \frac{\left(1 + \frac{1}{1-\theta}\left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}}\right] \Longleftrightarrow \\
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left(\frac{1}{\theta} + d\right)^{-1}\left[1 - \frac{\left(\frac{1}{\theta} + d + \frac{1}{1-\theta}\left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}} \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}}\right] < 0
\end{aligned}$$

The nominator of the fraction in brackets is greater than the denominator since products are scaled by $\frac{1}{1-\theta} > 1$; the expression in brackets is negative. □

To state the proposition note that 1 (or 2) is labeled the *connecting* country and 3 is the *hub.*

Proposition

1. *If the connecting country's transport firms decide to trade using the hub, they must increase trading distance to* $d'' = 2d > d'$. *This requires an increase in the ratio of fixed costs of transport between the hub and the connecting country, either by raising (lowering) the hub's (connecting country's) fixed costs of transport or any increasing combination of both.*

2. *If the hub is ever more distant from the connecting country, the ratio of fixed costs of transport must decrease either by lowering (raising) the hub's (connecting country's) fixed costs of transport or any decreasing combination of both.*

The decision of transport firms in country 1 to connect (in)directly cannot however affect a change in country 3's level of fixed costs of transport as the profit functions of transport firms are independent across countries. Therefore any change in the fixed costs of transport ratio $\frac{F_3^t}{F_1^t}$ is caused solely by firms in country 1. Intuitively if 1 decides to use a hub it must shed any excess transport infrastructure to accommodate higher marginal costs of transport, ceteris paribus. The hub must be in the position to accept increased traffic and thus have higher levels of transport infrastructure, but is benefited with lower marginal costs to destinations ceteris paribus. But the more remote a country is the harder it becomes to connect to the hub.

The proposition states possible routes a transport firm can choose by altering the fixed costs of transport. As such it is equivalent with the main model's outcomes where this interplay is observed through the benefits and the costs of forming a connection, but it necessitates strong assumptions regarding the non-variation of multiplicative costs, and adding a partitioned third sector in the classic model of homogeneous firm trade. Subsequently, the replacement of the unobserved benefit term δ with fixed costs $S_i \propto F_i^t$, retains the qualitative characteristics of the main model and allows empirical estimation.

– *The number of transport firms*

The model is complete by characterising the number of transport firms available in each country. This result has no impact on the predictions of the two models and thus is relegated to the Technical Appendix.

C Robustness Checks

– *Supplementary information*

A grid[16] containing all possible distances between country capitals to principal ports and intervening port to port distances is compiled. I assign an indicator if a particular route passes within a few kilometres of every possible location that is a canal, passage or a hub,

---

[16] Compiled using the following algorithm: 1. Calculate a visibility mesh from a set of land polygons, representing the world's land masses. 2 Add origin and destination to the visibility mesh (if e.g. origin is inland, then the only point visible is the nearest point on the coast). 3. Use the A* algorithm to calculate the shortest route through the visibility mesh. Visibility means that it's possible to travel in a straight line between points A and B, without going over land, for which a land mask is introduced consisting of 4-5,000 points. Vessels can travel along the coasts, which would not be entirely realistic, so it is expected the distances to represent a lower minimum. I would like to thank Michael Traut for providing these observations.

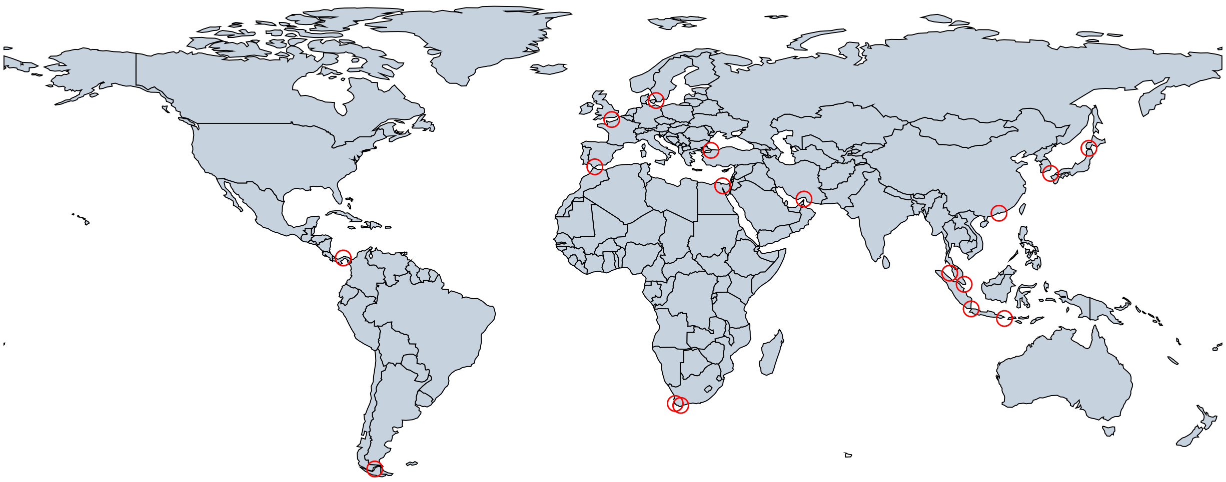


Figure 5: Hub Locations, (alternative distance matrix)

illustrated in Figure 5. I exclude landlocked exporting/importing countries from the sample. Since shortest routes are calculated, it is possible that transshipment hubs may become too remote, as is the case of Hong Kong for example: 183 routes out of the possible 6,631 routes that use a hub pass through this area, while 2,845/6,631 pass from the Panama Canal. In terms of flow number disaggregation at the HS 2 and 6 levels, only 7% and 13% of export flows use Hong Kong while 41% and 48% use the Panama Canal respectively, indicating that this sample is biased towards locations favouring geographic centrality.

Table 5: Exports: Impacts of i) indirect capital distance when no hubs are present on a route and ii) indirect distance when one or more hubs are present on a route.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total Trade												
I	Indirect Capital Distance, #Hubs=0	-1.937***	-1.919***					-1.557***	-1.533***				
	R^2	0.950	0.958					0.937	0.953				
IIa	Indirect Capital Distance,#Hubs>0	-0.656	-0.950					-1.316**	-1.682**				
	Interaction	-0.392	-0.163					-0.560	-0.297				
	Number of Hubs	0.0446	-0.0524					0.356	0.251				
	R^2_A	0.919	0.943					0.904	0.941				
IIb	Indirect Capital Distance,#Hubs>0	-0.721	-0.878					-1.758***	-1.985***				
	Interaction	-0.306	-0.261					0.113	0.164				
	R^2_B	0.919	0.943					0.904	0.941				
	p-value Ho A	0.041	0.123					0	0				
	p-value Ho B	0	0					0	0				
	HS 2 Level												
I	Indirect Capital Distance, #Hubs=0	-1.394***	-1.363***	-1.699***	-1.675***	-2.192***	-2.201***	-1.178***	-1.148***	-1.409***	-1.396***	-1.887***	-1.946***
	R^2	0.375	0.383	0.580	0.588	0.833	0.882	0.248	0.261	0.477	0.490	0.778	0.871
IIa	Indirect Capital Distance,#Hubs>0	-1.562***	-1.852***	-1.524***	-1.820***	-2.176***	-2.378***	-2.266***	-2.194***	-2.248***	-2.142***	-2.928***	-2.651***
	Interaction	0.138	0.0645	-0.00716	-0.105	0.417	0.244	0.222	0.143	0.118	0.00923	0.434	0.177
	Number of Hubs	0.151	0.240	0.253	0.372***	0.0744	0.217	0.297	0.240*	0.390**	0.351***	0.296	0.322
	R^2_A	0.253	0.267	0.434	0.449	0.766	0.868	0.200	0.213	0.406	0.423	0.734	0.851
IIb	Indirect Capital Distance,#Hubs>0	-1.769***	-2.147***	-1.870***	-2.276***	-2.274***	-2.622***	-2.614***	-2.467***	-2.703***	-2.541***	-3.247***	-2.974***
	Interaction	0.402**	0.468***	0.435**	0.520***	0.546***	0.599***	0.720***	0.541***	0.771***	0.592***	0.922***	0.694***
	R^2_B	0.253	0.267	0.433	0.449	0.766	0.868	0.199	0.213	0.406	0.422	0.734	0.851
	p-value Ho A	0	0	0	0	0	0	0	0	0	0	0	0
	p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
	HS 6 Level												
I	Indirect Capital Distance, #Hubs=0	-0.613***	-0.576***	-0.995***	-0.961***	-2.028***	-2.287***	-0.261**	-0.223**	-0.489***	-0.469***	-1.228***	-1.368***
	R^2	0.233	0.241	0.407	0.415	0.754	0.878	0.086	0.102	0.314	0.326	0.697	0.862
IIa	Indirect Capital Distance,#Hubs>0	-1.272***	-1.470***	-1.146***	-1.574***	-1.231***	-1.820***	-1.471***	-1.406***	-1.466***	-1.596***	-1.394***	-1.671**
	Interaction	0.665**	0.560**	0.522*	0.438*	0.553	0.383	0.565**	0.489**	0.445*	0.419*	0.502	0.276
	Number of Hubs	-0.123	-0.00140	-0.0975	0.0820	-0.104	0.0889	0.0322	0.0466	0.0588	0.0910	-0.0760	0.0577
	R^2_A	0.142	0.152	0.329	0.338	0.726	0.878	0.118	0.131	0.355	0.365	0.703	0.872
IIb	Indirect Capital Distance,#Hubs>0	-1.092***	-1.468***	-1.005***	-1.678***	-1.084***	-1.930***	-1.511***	-1.463***	-1.536***	-1.706***	-1.309**	-1.740**
	Interaction	0.447***	0.558***	0.351**	0.577***	0.369*	0.536**	0.618***	0.568***	0.543***	0.571***	0.375*	0.375
	R^2_B	0.142	0.152	0.329	0.338	0.726	0.878	0.118	0.131	0.355	0.365	0.703	0.872
	p-value Ho A	0	0	0	0	0	0	0	0	0	0	0.03	0.06
	p-value Ho B	0	0	0	0	0	0	0	0	0	0	0.05	0.08
	Fixed Effects												
	Exporter	Y		Y				Y		Y			
	Importer	Y		Y				Y		Y			
	Year	Y		Y		Y		Y		Y		Y	
	Exporter-Year		Y		Y				Y		Y		
	Importer-Year		Y		Y				Y		Y		
	Commodity			Y	Y					Y	Y		
	Exporter-Commodity					Y						Y	
	Importer-Commodity					Y						Y	
	Exporter-Commodity-Year						Y						Y
	Importer-Commodity-Year						Y						Y
	Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of i) indirect capital distance when no hubs are present on a route (rows *I*) and ii) indirect distance when one or more hubs are present on a route. *IIa* indicates a set of regressions including indirect capital distance and the interaction with the number of hubs on a route as independent variables. *IIb* indicates a set of regressions including indirect capital distance, the number of hubs on a route and their interaction as independent variables. Ho A: Indirect Capital Distance=Interaction=0. Ho B: Indirect Capital Distance=Interaction=0. The regression output for each trade level is available in the Technical Appendix.

Table 6: Exports: Coefficients of direct distance and indirect distance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-1.534***	-1.537***					-1.384***	-1.408***				
Indirect Capital Distance	-1.374***	-1.449***					-1.295	-1.333				
R^2_{Direct}	0.874	0.891					0.841	0.869				
$R^2_{Indirect}$	0.869	0.888					0.838	0.867				
p-value Ho A	0.0847	0.412					0.452	0.568				
p-value Ho A (P. T.)	0.263	0.576					0.606	0.694				
HS 2 Level												
Direct Capital Distance	-0.907***	-0.915***	-1.013***	-1.019***	-1.375***	-1.433***	-0.921***	-0.951***	-0.969***	-1.001***	-1.353***	-1.454***
Indirect Capital Distance	-0.666***	-0.664***	-0.768***	-0.765***	-1.115***	-1.342***	-0.735***	-0.746***	-0.784***	-0.799***	-1.148***	-1.329***
R^2_{Direct}	0.366	0.377	0.499	0.511	0.762	0.817	0.262	0.279	0.413	0.431	0.695	0.780
$R^2_{Indirect}$	0.361	0.372	0.493	0.505	0.754	0.811	0.257	0.274	0.408	0.426	0.689	0.775
p-value Ho A	0	0	0	0	0	0.433	0.0159	0	0.0298	0.0186	0.0939	0.382
p-value Ho A (P. T.)	0.0281	0.0181	0.0478	0.0351	0.0888	0.607	0.142	0.100	0.194	0.152	0.273	0.551
HS 6 Level												
Direct Capital Distance	-0.481***	-0.501***	-0.616***	-0.637***	-1.175***	-1.345***	-0.406***	-0.434***	-0.497***	-0.522***	-1.062***	-1.271***
Indirect Capital Distance	-0.193***	-0.2***	-0.308***	-0.312***	-0.709***	-1.108***	-0.238***	-0.243***	-0.308***	-0.311***	-0.628***	-1.113***
R^2_{Direct}	0.214	0.223	0.365	0.373	0.704	0.826	0.103	0.116	0.303	0.314	0.649	0.807
$R^2_{Indirect}$	0.211	0.220	0.360	0.368	0.694	0.817	0.101	0.115	0.301	0.312	0.644	0.803
p-value Ho A	0	0	0	0	0	0	0	0	0	0	0	0
p-value Ho A (P. T.)	0.0305	0.0173	0.0251	0.0131	0.0127	0.336	0.0755	0.0338	0.0419	0.0204	0.00324	0.418
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of direct capital distance and indirect capital distance and measures of goodness of fit from the estimation of equation 5 in the presence of controls for contiguity, common language, colony and common regional trade agreements. Ho A: Direct capital distance=Indirect capital distance. P. T. = Paternoster test. The regression output for each trade level and column is available in the Technical Appendix.

Table 7: Coefficients of direct distance versus coefficients of indirect distance plus the interaction term and associated test outcomes.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-1.534	-1.537					-1.384	-1.408				
Indirect Capital Distance + Interaction	-1.536	-1.625					-1.556	-1.615				
% Discount from trading via hub(s)	-0.177	-5.773					-12.39	-14.67				
R^2_{Direct}	0.874	0.891					0.841	0.869				
$R^2_{Indirect}$	0.869	0.888					0.840	0.869				
p-value Ho B	0	0					0	0				
p-value Ho C	0.977	0.418					0.157	0.129				
HS 2 Level												
Direct Capital Distance	-0.907	-0.915	-1.013	-1.019	-1.375	-1.433	-0.921	-0.951	-0.969	-1.001	-1.353	-1.454
Indirect Capital Distance + Interaction	-0.744	-0.745	-0.858	-0.858	-1.202	-1.455	-0.840	-0.868	-0.883	-0.920	-1.330	-1.546
% Discount from trading via hub(s)	17.99	18.57	15.37	15.79	–	–	8.699	8.727	8.884	8.070	1.644	-6.412
R^2_{Direct}	0.366	0.377	0.499	0.511	0.762	0.817	0.262	0.279	0.413	0.431	0.695	0.780
$R^2_{Indirect}$	0.361	0.372	0.493	0.505	0.754	0.811	0.258	0.274	0.408	0.426	0.689	0.775
p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
p-value Ho C	0.0102	0	0.0229	0.0159	0.0291	0.811	0.340	0.317	0.343	0.365	0.837	0.432
HS 6 Level												
Direct Capital Distance	-0.481	-0.501	-0.616	-0.637	-1.175	-1.345	-0.406	-0.434	-0.497	-0.522	-1.062	-1.271
Indirect Capital Distance + Interaction	-0.190	-0.210	-0.304	-0.319	-0.681	-1.058	-0.285	-0.293	-0.355	-0.363	-0.715	-1.172
% Discount from trading via hub(s)	–	–	–	–	–	–	–	–	–	–	32.71	–
R^2_{Direct}	0.214	0.223	0.365	0.373	0.704	0.826	0.103	0.116	0.303	0.314	0.649	0.807
$R^2_{Indirect}$	0.211	0.220	0.360	0.368	0.694	0.817	0.101	0.115	0.301	0.312	0.644	0.803
p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
p-value Ho C	0	0	0	0	0	0	0	0	0	0	0	0
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of direct capital distance, indirect capital distance plus the interaction term with the hub indicator variable and measures of goodness of fit from the estimation of equations 5 and 6, in the presence of controls for contiguity, common language, colony and common regional trade agreements. Ho B: Indirect capital distance=Interaction=0. Ho C: Direct capital distance=Indirect capital Distance+interaction. The regression output for each trade level and column is available in the Technical Appendix.

Table 8: Exports: 2SLS impacts of direct and indirect capital distance using time difference as an instrument.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Trade												
Direct Capital Distance	-1.525***	-1.554***					-1.317***	-1.348***				
R^2_{Direct}	0.875	0.892					0.841	0.870				
Underidentification Test	24.94	24.02					22.12	21.26				
Cragg-Donald	1134	900.9					1131	909.3				
Indirect Capital Distance	-1.251***	-1.256***					-1.078***	-1.096***				
$R^2_{Indirect}$	0.870	0.888					0.838	0.867				
Underidentification Test	40.65	39.42					36.63	34.98				
Cragg-Donald	1762	1529					1767	1474				
p-value Ho A	0.222	0.225					0.327	0.349				
p-value Ho A (P. T.)	0.560	0.572					0.628	0.650				
HS 2 Level												
Direct Capital Distance	-0.841***	-0.874***	-1.094***	-1.130***	-1.820***	-2.063***	-0.692**	-0.751**	-0.830**	-0.893**	-1.461***	-1.719**
R^2_{Direct}	0.366	0.377	0.499	0.51	0.760	0.812	0.261	0.278	0.412	0.43	0.695	0.779
Underidentification Test	9.713	9.809	9.586	9.685	8.755	9.786	9.370	9.298	9.271	9.205	8.419	9.233
Cragg	18497	17577	18303	17399	15859	9713	16479	15542	16380	15459	14560	9331
Indirect Capital Distance	-0.596***	-0.609***	-0.774***	-0.786***	-1.245***	-1.380***	-0.488***	-0.523***	-0.584***	-0.621***	-0.993***	-1.152***
$R^2_{Indirect}$	0.361	0.372	0.493	0.505	0.754	0.811	0.256	0.273	0.407	0.425	0.688	0.774
Underidentification Test	20.79	21.29	20.65	21.15	20.01	22.59	19.88	19.80	19.76	19.69	18.86	20.37
Cragg-Donald	41175	39672	40892	39407	37492	28025	38432	36438	38305	36326	35957	25504
p-value Ho A	0.247	0.231	0.145	0.135	0.0310	0.0315	0.266	0.246	0.216	0.201	0.0643	0.0663
p-value Ho A (P. T.)	0.554	0.542	0.521	0.510	0.478	0.492	0.613	0.600	0.610	0.599	0.546	0.556
HS 6 Level												
Direct Capital Distance	-0.0262	-0.0936	-0.307	-0.367	-1.368***	-1.780***	0.0404	-0.0143	-0.180	-0.224	-1.076***	-1.489***
R^2_{Direct}	0.211	0.221	0.363	0.372	0.704	0.824	0.100	0.114	0.302	0.313	0.649	0.806
Underidentification Test	10.01	9.258	9.920	9.159	11.92	17.70	8.407	7.793	8.150	7.543	7.808	12.13
Cragg-Donald	140505	131586	136150	127349	102721	38079	104656	96332	99746	91737	64132	19242
Indirect Capital Distance	-0.0207	-0.0731	-0.242	-0.285	-1.045***	-1.411***	0.0304	-0.0106	-0.135	-0.165	-0.734***	-1.032***
$R^2_{Indirect}$	0.211	0.220	0.360	0.368	0.693	0.816	0.100	0.114	0.301	0.312	0.643	0.803
Underidentification Test	22.34	21.40	22.21	21.26	27.70	38.57	20.26	19.14	20.09	18.92	23.27	33.59
Cragg-Donald	263481	245140	255518	237729	180882	85310	178470	162904	171334	156379	109685	48214
p-value Ho A	0.983	0.938	0.780	0.740	0.190	0.249	0.957	0.984	0.756	0.697	0.0625	0.0853
p-value Ho A (P. T.)	0.990	0.964	0.863	0.836	0.502	0.548	0.977	0.992	0.862	0.828	0.484	0.484
Fixed Effects												
Exporter	Y		Y				Y		Y			
Importer	Y		Y				Y		Y			
Year	Y		Y		Y		Y		Y		Y	
Exporter-Year		Y		Y				Y		Y		
Importer-Year		Y		Y				Y		Y		
Commodity			Y	Y					Y	Y		
Exporter-Commodity					Y						Y	
Importer-Commodity					Y						Y	
Exporter-Commodity-Year						Y						Y
Importer-Commodity-Year						Y						Y
Sample	Full						No USA-EU15 Bilateral Trade					

Reported 2SLS coefficients of direct capital distance and indirect capital distance and measures of goodness of fit from the estimation of equation 5 in the presence of controls for contiguity, common language, colony and common regional trade agreements. Ho A: Direct capital distance=Indirect capital distance. P. T. = Paternoster test. The first and second stage regression output for each trade level and column is available in the Technical Appendix.

Table 9: Number of hubs.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total Trade												
I	Indirect Capital Distance, #Hubs=0	-1.408***	-1.376***					-1.139***	-1.152***				
	R^2	0.880	0.913					0.880	0.914				
IIa	Indirect Capital Distance,#Hubs>0	-1.855***	-1.830***					-1.729***	-1.810***				
	Interaction	0.278	0.188					-0.0129	-0.0127				
	Number of Hubs	-0.0452	0.0118					0.0592	0.0704				
	R^2_A	0.880	0.902					0.833	0.865				
IIb	Indirect Capital Distance,#Hubs>0	-1.795***	-1.845***					-1.804***	-1.901***				
	Interaction	0.211*	0.206					0.0760	0.0953				
	R^2_B	0.880	0.902					0.833	0.865				
	p-value Ho A	0	0					0	0				
	p-value Ho B	0	0					0	0				
	HS 2 Level												
I	Indirect Capital Distance, #Hubs=0	-0.712***	-0.709***	-0.789***	-0.785***	-1.172***	-1.391***	-0.632***	-0.640***	-0.644***	-0.651***	-0.998***	-1.263***
	R^2	0.356	0.370	0.495	0.510	0.771	0.858	0.292	0.308	0.443	0.461	0.737	0.842
A	Indirect Capital Distance,#Hubs>0	-0.863***	-0.810***	-1.191***	-1.149***	-1.680***	-1.882***	-0.844***	-0.807***	-1.081***	-1.067***	-1.614***	-1.835***
	Interaction	0.0630	0.0145	0.249*	0.208*	0.371**	0.328	-0.0864	-0.136	0.0681	0.0283	0.139	0.221
	Number of Hubs	0.134**	0.158**	0.0244	0.0467	-0.0497	-0.0212	0.116	0.139*	0.0348	0.0524	-0.0144	-0.0702
	R^2_A	0.372	0.384	0.506	0.520	0.775	0.847	0.234	0.253	0.395	0.416	0.697	0.814
IIb	Indirect Capital Distance,#Hubs>0	-1.077***	-1.064***	-1.230***	-1.223***	-1.601***	-1.848***	-1.013***	-1.011***	-1.131***	-1.144***	-1.593***	-1.730***
	Interaction	0.277***	0.268***	0.288***	0.282***	0.292***	0.294***	0.0923	0.0798	0.122	0.109	0.116	0.110
	R^2_B	0.371	0.384	0.506	0.520	0.775	0.847	0.234	0.253	0.395	0.416	0.697	0.814
	p-value Ho A	0	0	0	0	0	0	0	0	0	0	0	0
	p-value Ho B	0	0	0	0	0	0	0	0	0	0	0	0
	HS 6 Level												
I	Indirect Capital Distance, #Hubs=0	-0.223***	-0.208***	-0.334***	-0.315***	-0.670***	-1.319***	-0.173**	-0.156**	-0.218***	-0.195***	-0.365**	-1.141***
	R^2	0.206	0.214	0.378	0.386	0.723	0.882	0.119	0.130	0.331	0.340	0.700	0.868
IIa	Indirect Capital Distance,#Hubs>0	-0.305	-0.286	-0.558***	-0.521***	-1.028***	-1.305***	-0.692***	-0.594***	-0.869***	-0.749***	-1.275***	-1.456***
	Interaction	0.135	0.0752	0.222*	0.149	0.280**	0.177	0.297**	0.191**	0.362***	0.232***	0.297**	0.170
	Number of Hubs	0.0227	0.0551	-0.00856	0.0282	-0.0497	0.0572	-0.109*	-0.0674*	-0.136**	-0.0894***	-0.101*	-0.109
	R^2_A	0.214	0.224	0.366	0.375	0.728	0.871	0.088	0.104	0.310	0.323	0.667	0.853
IIb	Indirect Capital Distance,#Hubs>0	-0.348**	-0.391**	-0.542***	-0.575***	-0.932***	-1.408***	-0.552***	-0.505***	-0.693***	-0.630***	-1.136***	-1.288***
	Interaction	0.176**	0.175**	0.207**	0.200**	0.188**	0.279*	0.144	0.0950	0.170*	0.104	0.146*	-0.0151
	R^2_B	0.214	0.224	0.366	0.375	0.728	0.871	0.088	0.104	0.310	0.323	0.667	0.853
	p-value Ho A	0.193	0.0397	0	0	0	0	0	0	0	0	0	0
	p-value Ho B	0.0753	0.0434	0	0	0	0	0	0	0	0	0	0
	Fixed Effects												
	Exporter	Y		Y				Y		Y			
	Importer	Y		Y				Y		Y			
	Year	Y		Y		Y		Y		Y		Y	
	Exporter-Year		Y		Y				Y		Y		
	Importer-Year		Y		Y				Y		Y		
	Commodity			Y	Y					Y	Y		
	Exporter-Commodity					Y						Y	
	Importer-Commodity					Y						Y	
	Exporter-Commodity-Year						Y						Y
	Importer-Commodity-Year						Y						Y
	Sample	Full						No USA-EU15 Bilateral Trade					

Reported coefficients of i) indirect capital distance when no hubs are present on a route (rows *I*) and ii) indirect distance when one or more hubs are present on a route. *IIa* indicates a set of regressions including indirect capital distance and the interaction with the number of hubs on a route as independent variables. *IIb* indicates a set of regressions including indirect capital distance, the number of hubs on a route and their interaction as independent variables. Ho A: Indirect Capital Distance=Interaction=0. Ho B: Indirect Capital Distance=Interaction=0. The regression output for each trade level is available in the Technical Appendix.