# Endogenous hub formations in international trade*

Nicholas-Joseph Lazarou†

Energy Institute
University College London

**Abstract**

Transshipment hubs such as Singapore have a ubiquitous role in facilitating the international trade of goods. Yet when and why should a transshipment hub emerge remains largely unexamined. We incorporate network theory in a trade model of monopolistic competition with representative firms and identify that the determinants of optimal network formation are i) the level of transportation costs, ii) increasing returns in transportation, iii) locational advantage. Empirical evidence across three aggregation levels of trade data suggests that if remote partners select a route passing through a hub, the doubling of distance reduces exports by 26% less on average compared to them selecting a direct route. Thus geographically disadvantaged countries that absorb high transportation costs could improve upon these by trading via a hub.

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

JEL Classification Codes: F12, L91, L14.

*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, 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 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. Email: N.Lazarou@ucl.ac.uk

# 1 Introduction

Are trade volumes affected by transshipment hubs? To answer this important question, we conjecture when and why such hubs emerge within a trade model à la Krugman (1980), deriving an appropriately specified set of gravity equations amenable to empirical testing. They reveal that a hub's presence impacts beneficially remote trading partners by providing a discount in the transportation cost relative to them trading without the hub's utilisation.

Today hub ports such as Singapore and Rotterdam have pre-eminent functions as transshipment areas for goods transported by sea. They have emerged since the 1950's as a corollary of technological progress and innovations in shipbuilding and port infrastructure. These permitted stronger scale effects in the carriage of goods, the reduction of labour -and by correlation port capital- costs and heralded the advents of containerisation and intermodal transportation. Such factors in turn induced the prominence of hub and spoke transport networks (Hummels, 2007; Levinson, 2008; Rodrigue, 2010; Rua, 2014; Bernhofen, El-Sahli and Kneller, 2015).

The significance of particular locations serving as transshipment hubs compared to less attractive 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). Trade route distributions under such network types are defined by unit transportation costs diminishing in vessel size or size of aggregate port infrastructure: Bigger ships operate on relatively thicker trade routes, whilst smaller ships deliver the quantity demanded to and from the relatively larger transshipment ports (Krugman, 1993; Hummels, Lugovskyy and Skiba, 2009).

Uncovering how these distributions ultimately affect trade volumes becomes the key objective of this study: In what follows we briefly outline the model's intuition, its predictions and empirical outcome in order to discuss its position in and contribution to the extant trade literature.

We employ a simple trade model with representative firms embedded in a network wherein connectivity is assessed through costs and benefits using the symmetric connections model of Jackson and Wolinsky (1996). When transportation costs to a remote destination are high firms, instead of lowering their output or exiting this particular market, can choose to trade via a hub provided there is a cost saving incentive, with the opposite holding true. We also contribute by proving that this parsimonious model is qualitatively equivalent to a more cumbersome analysis presented in Appendix B. The latter incorporates additive transportation costs, a transport sector operating under increasing returns, partially asymmetric trade to emulate remoteness and an assumption that fixed costs associated with transportation can vary within countries.

We document a significant beneficial impact stemming from the existence of hub ports across three aggregation levels of Harmonised System export flows. The model's equilibrium aggregate exports characterisation provides the empirically testable gravity formulations. We alternately compare two distance variables, one being great circle capital distance and the other being capital distance using sea routes, termed direct and indirect capital distance respectively. For indirect capital distance we assign an indicator variable that serves to detect if a route passes through an exogenously defined hub location associated with a sea port. Interacting the binary variable with indirect distance yields that the marginal effect of partners trading via a hub becomes significantly less in absolute value than the marginal effect of those partners trading directly. This represents a discount in the cost of transport due to the interaction.

Across aggregation levels, a doubling of distance reduces exports by 26% less if a shipping route passing through a hub is selected compared to the alternative. Countries that do not have a hub intersecting their route are found to be less remote. But for more distant partners remoteness is counteracted by the presence of the hub location on the route. Weak evidence also suggests that the greater the number of hubs employed on a route the lower the gains from their utilisation. Yet the hub indicator variable acts also as a proxy for route selection by firms and hence there is correlation with the error term. The endogeneity is addressed by introducing time difference, which is perceived to be agnostic to transport routing decisions, as a variable to satisfy the exclusion restriction.

This paper contributes to a scarce literature that perturbs the iceberg trade cost vector in order to model behaviours stemming from transport technology and derive associated economic inference. The international trade literature has not addressed whether transshipment hubs have any impact on trade volumes. A plausible reason for the lack of studies may be the insufficient attention paid to the structure of the transport industry (Krugman, 1993) combined with data paucity on transportation costs and trade routes.

We invoke network theory to answer this question. Chaney (2014) presents us with a dynamic model of information frictions in trade which, when accounting for geographic space, explains the distribution of the number and location of countries reached by exporters. The model we present does not impede the direct or remote searches leading to the creation of information hubs from which exporters radiate away to more distant contacts. Because the model is static in firm behaviour and geographic space, it takes this dynamic process as given, hence making the comparison of a world trading only with direct search versus the combination of two search methods. It then quantifies a cost saving incentive for firms to export using the latter. Yet it is not able to uncover any details about the distribution of the number of countries that can be reached using a specific transport network, and so this paper focuses on changes in the intensive margin.[1]

Within this static microcosm of the Chaney (2014) model, there are a number of papers in the neighbourhood to mine. Most related is Krugman (1993) wherein the interaction between increasing returns and low levels of transportation costs to and from a particular location leads to the emergence of a hub region by virtue of concentration of production, which implies also a strong historical incidence. We extend this work by dropping the concentration of production prerequisite and keeping locational advantage and the level of transportation costs exogenous. Whilst concentration of production jointly with geographic advantage may be of importance under air transport, the same need not apply for maritime transport which this paper takes into consideration:[2] Some of the less developed regions of the world, such as Panama or Port Said, obtain hub status conditional only on location based on the distribution of trade routes. Concentration of production may then take place

[1]A distinction must also be made that a transport network may not necessarily align with a network of information diffusion as described in Chaney (2014). For example London can be perceived as a major information hub in terms of diffusion of information, yet for transportation purposes, the port of London handles less throughput than, for example, the port of Rotterdam.

[2]Maritime transport is responsible for 99% of the world's trade by weight and 90% of merchandised volume (Hummels, 2007; OECD Trade and Agriculture Directorate, 2008; Korinek and Sourdin, 2010).

but is not a condition precedent.

The models of Hendricks, Piccione and Tan (1995), Starr and Stinchcombe (1992) and Oum, Zhang and Zhang (1995) focus on air passenger transport by a particular carrier. They find that economies of density are a determinant of airline network formation as total costs per passenger on a route[3] decline with the number of passengers travelling on that route. Hub networks exhibit higher traffic densities than direct connections' networks. The distance travelled is longer, but if economies of density are sufficiently large, the total costs of satisfying demands may be lower. The common observation across the aforementioned four models of trade explaining hub formations is the infeasibility to develop testable predictions for their existence which also validates why there is little empirical application of theoretical research in economic geography (Allen and Arkolakis, 2014). We contribute by providing an empirically testable prediction of a simple economic geography model that finds support in the data.

The related empirical literature utilises natural experiments to infer the marginal effect of transportation costs on trade flows. The shocks that assist identification of the effect by isolating the causal direction become the closure of the Suez Canal (Feyrer, 2011), the eruption of the Eyjafjallajökull volcano (Besedeš and Murshid, 2015) and the closure of a bridge (Martincus, Carballo, Garcia and Graziano, 2014). However these shocks are few and far between; they have varying degrees of unanticipation and dispersion of the impacts that makes inference rare and difficult to completely isolate. This paper does not benefit from observing similar shocks. Instead it proposes a simple mechanism that uncovers information about routing decisions contained in the error term and proposes a way to control for these through the exclusion restriction.

By doing so, it complements research that signals the importance of embedding transport technology in trade. Hummels (2007) documents that there is little, if any, maritime transportation cost declines with the opposite holding for airborne trade during the post-war period. Hence trade growth cannot be explained sufficiently by transportation cost reductions during the second era of globalisation.[4] On the contrary this seems to be the prevalent

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

[4]Baier and Bergstrand (2001) also find that declining transportation costs contributed to about 8% of

reason for the first era as Pascali (2014) reveals that transportation cost reductions induced by the introduction of the steamship contributed to about 54% of the export growth between 1850-1905. Bernhofen, El-Sahli and Kneller (2015) quantify the impact of containerisation for international trade and report that while it has strongly benefited developed economies, it has had relatively small long term effects for developing countries. This finding can be explained by higher transportation costs that developing countries face, derived from higher product prices and levels of protection, lower competition on routes and import demand elasticities, due to presence of market power in shipping as Hummels, Lugovskyy and Skiba (2009) uncover. Rua (2014) provides evidence that the diffusion[5] of containerisation during the post-war period depended on network size, average income in the network and domestic infrastructure development. With respect to inefficiencies of the latter, Clark, Dollar and Micco (2004) illustrate the nested impact they have on trade through transportation costs. The impacts of, and substitutability between, transportation modes on trade flows are presented in Harrigan (2010) and Lux (2011).

Lastly, this paper, through the analysis in Appendix B that augments the assumptions used for defining the empirical specification, can be classified in the strand of literature that highlights the prevalence of additive trade costs in addition to iceberg costs such as Irarrazabal, Moxnes and Opromolla (2015), Hummels and Skiba (2004) and Hummels, Lugovskyy and Skiba (2009). This is a field of research which is considered to be under-explored as Melitz and Redding (2014) state.

The implications of this study are directed towards developing countries. Incumbent exporters have to absorb higher transportation costs so as to penetrate markets abroad. This situation prevents export-led development, reducing workers' wages and inducing a welfare impact (Amjadi, Reinke and Yeats, 1996). Higher transportation costs are attributed to geographic disadvantage and lack of proximity. Prevention of market access for developing nations translates to losses from trade of about 68% lower GDP per capita on average (Redding and Venables, 2004). Therefore improvement of own and transit country infrastructure together with hub formations could result in lower transportation costs (Limão and

post-war mean trade growth among 16 developed countries.

[5]In the sense of adoption speed and usage.

Venables, 2001).

The remainder of this paper is organised as follows. Section 2 provides simple notions of economic and social networks that will be used in the model of section 3. The equilibrium is characterised in section 4. The empirical specifications used for testing the existence of hub formations are derived in section 5. Data description ensues in section 6 followed by results in section 7. Robustness checks are performed in section 7.2 and in appendix C. The last section concludes.

## 2 Setup of the network

Consider a set of countries $K = \{1, 2, ..., k\}$ which engage or not in international trade through manufacturing firms. Countries can be *directly connected*, if they have a direct trading relationship using no other intermediary country. Thus, a network $G$ is defined as a list of pairs of countries $\{i, j\}$ that are directly connected to each other. Each connection can be represented as a *graph* $g \in G$. And hence if country $i$ and country $j$ are directly connected $g$ takes a value of 1 or if they are not directly connected $g$ becomes 0. The existence of a *direct connection* will be denoted as $g_{ij} = 1$, and $g_{ij} = 0$ if countries $i$ and $j$ are *not directly connected*.

Each *connection* is associated with costs and benefits. These costs and benefits affect firms that choose to enter the export market in each country. Two connection types can exist, a *direct* and an *indirect* connection. If a direct connection is formed by a firm then it must incur a cost $c$. There is a benefit $0 \leq \delta \leq 1$ between $i$ and $j$ that can be perceived as a measure of the firm's market access to country $j$. The firm has the additional option to form an indirect connection. Implicitly, there must already exist a direct connection to another country for the indirect connection to be feasible. It is formed without cost and the firm receives only a pure benefit $\delta^{t_{ij}}$, where $t_{ij} \geq 0$ is the integer number of connections between countries $i$,$j$. This construction allows a firm in country $i$ entering the export market to probe i) how easy it is to access a partner ($\delta$) and ii) the cost to export to that partner ($c$). It can also consider whether to connect directly to the destination country and incur this cost. Or alternatively, it can consider connecting indirectly. In the latter case it avoids the cost but receives a discounted benefit as the accessibility decreases. The difference between the benefit of forming a connection and the associated cost is thus defined as:

$$v_{ij} = \delta^{t_{ij}} - c_{ij|ij \in G}$$

By convention we have $g_{ii} = 0 \Rightarrow c_{ii} = 0$ since $g_{ii} = 0$ is not a connection in the network $G$ and country $i$ remains autarkic. Further, $t_{ij} = 0$ if there is no path that connects directly or indirectly countries $i$ and $j$. An exposition of this construct is described in Appendix A.

Countries are also characterised by their participation share in the network depending on the types of connections they form. The network participation share will be perceived as the fixed cost associated with the network. While the participation share is not employed in the theoretical model, it will assist in the gravity equation specification and the parallel model of Appendix B in lieu of the unobservable benefit of forming an indirect connection. Denote the set of country $i$'s direct connections in a network as $N_i(G) = \{i \neq j | g_{ij} = 1\}$. The cardinality of this set is $n_i(G)$. The size of the network is $n(G) = \sum_i^N n_i(G)$. The participation share of $i$ in the network is simply $F_i = \frac{n_i(G)}{n(G)}$. An example is provided in Appendix A.

For the remainder of this paper, countries are symmetrically spaced: $ij = ji$ therefore $c_{ij} = c_{ji} = c, t_{ij} = t_{ji} = t$.

# 3 Setup of the trade model

Symmetric countries produce goods using only labour. Country $n$ has a population $L_n$ and two sectors. One sector is responsible for the production of a single homogeneous good that can be traded freely. This good is the numeraire. The other sector produces a continuum of differentiated varieties of a good that can be traded at a cost. Each specific variety is produced by a single monopolist. 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 goods. Each consumer is endowed with one unit of labour.

***Demand*** $\triangleright$ A representative consumer receives utility $U$ from consuming $q_0$ units of the numeraire and $q$ units of the differentiated variety $\omega$ both of which may be produced domestically or be imported. Her preferences are given by a C.E.S. utility function over the continuum of differentiated varieties:

$$U = q_0^{1-\mu} \left[ \int_{\omega \in \Omega} q_{ij}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}\mu}$$

where $\sigma > 1$ is the elasticity of substitution between pairs of varieties, $1-\mu$ is the share of income expenditure on 0, and $\Omega$ is the mass of available differentiated goods. Maximising utility subject to exhausting her labour income share, the representative consumer in country $j$ has demand for differentiated goods:

$$q_{ij} = \frac{\mu_j L_j p_{ij}^{-\sigma}}{\sum_{j,i=1}^{N} \int_{\Omega} p_{ij}^{1-\sigma}(\omega) d\omega}$$

where the denominator represents an aggregate price if the set of differentiated goods was consumed as an aggregate good.

***Production and Trade Costs*** $\triangleright$ Good 0 is the numeraire homogeneous good. Its price is normalised to 1, so that the wage is equal to the price of the good. In this respect the wage is set equal to 1 across countries due to free trade, and across the two sectors within each country.

One firm can produce one variety of the differentiated good. Labour costs for differentiated goods are split between a marginal and a fixed cost and thus the sector is characterised by increasing returns. In the aggregate trade -gravity- equation this assumption will be relaxed by incorporating the share of each country's participation in the network as a fixed cost proxy. The marginal cost consists of a constant parameter $\gamma > 0$ and a variable trade cost. The variable trade cost is the net benefit term that stems from forming a connection to another country. For domestic consumption the net benefit becomes by construction equal to the value of unity.

To produce and sell a variety $\omega$ either domestically or abroad, the firm in country $i$

employs labour input:

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

As such, if a direct connection $ij$ is formed the firm receives a benefit $\delta$ in the sense that $\frac{\partial \Pi}{\partial \delta} > 0$. In addition it incurs a transportation cost $c$ as $\frac{\partial \Pi}{\partial c} < 0$. In case of an indirect connection, the firm receives a decayed benefit $\delta^{t_{ij}}, t_{ji} > 1$ and incurs no cost at all.

The firm solves its maximisation problem constrained by the quantity demanded. It sets its optimal price equal to a constant markup over the unit cost $p_{ij} = \frac{\sigma}{\sigma-1} \frac{\gamma}{v_{ij}}$. Positive profits incentivise firms to enter the sector exhausting the profit margin. At zero profits, each firm produces output $Q_i \equiv \sum_{i,j=1}^{k} q_{ij}/v_{ij} = \frac{F_i}{\gamma}(\sigma - 1)$.

Over all varieties $\omega$ produced in each country, the total labour input must equal the labour share in the increasing returns sector: $\int_{\omega \in \Omega} L_{q,i}(\omega) d\omega = \mu_i L_i$. Since each firm produces one variety, the number of firms becomes finite and equal to $n_i = \frac{\mu_i L_i}{\sigma F_i}$. Consequently the aggregate price index can be characterised as $\sum_{j,i=1}^{N} \int_{\Omega} p_{ij}^{1-\sigma}(\omega) d\omega = \sum_{j,i=1}^{N} n_i p_{ij}^{1-\sigma}$.

# 4 Equilibria and comparative static experiments

In this section equilibria are characterised for varying network formations, symmetric geographical placement of countries, optimal prices given trade costs and traded quantities. The network specific notions of stability and efficiency are defined and are then proved for each formation observed in the Online Appendix.[6]

***Two Country Equilibrium*** $\triangleright$ The equilibrium is characterised by the zero profit condition across two countries due to free entry and exit of firms within each country. The net benefit term associated with the two countries becomes $v_{12} = \delta - c = v_{21}$ because of symmetry of the two direct connections $g_{12}$ and $g_{21}$. Then it must be that profits are $\pi_1 = \pi_2 = 0$. Given that fixed costs of production are equal and countries differ only in their size, the zero profit condition can be written compactly as:

[6]All material relegated to the Online Appendix is available on the following webpage: Online Appendix for "Endogenous hub formations in international trade".

$$\sum_{1,j=1}^{2} q_{1j}(p_{1j} - \frac{\gamma}{v_{1j}}) = \sum_{2,j=1}^{2} q_{2j}(p_{2j} - \frac{\gamma}{v_{2j}}) \Rightarrow Q_1 = Q_2$$

Domestic prices are equal across countries as $\gamma$ is a common constant. Prices abroad differ only by the net benefit term which is symmetric. By expanding the price indices in country 1 and 2 given domestic prices and prices abroad, in equilibrium the home market effect is yielded:

$$\frac{n_1}{n_2} = \frac{\frac{\mu_1 L_1}{\mu_2 L_2} - (\delta - c)^{\sigma-1}}{1 - \frac{\mu_1 L_1}{\mu_2 L_2}(\delta - c)^{\sigma-1}} > 0$$

Introducing a network leaves things unchanged in the standard model of trade with two countries (Krugman, 1980). Yet a notable remark is that the decision to export entails an additional inherent condition for the firm. Provided the cost of transport will never exceed the benefit and as long as profits cover the fixed costs, the firm will always be favourable towards establishing a connection.

The connection is beneficial for society as utility increased due to the greater number of varieties available to consumers. Clearly there is autarky when $c > \delta$ and the benefit is greater the more proximal countries 1 and 2 are. Thus in equilibrium if the relative size of country 1 increases there is a more than proportional increase in the relative number of domestic firms given the net benefit of forming a direct connection. The condition holds as long as $(\delta - c)^{\sigma-1} < \frac{\mu_1 L_1}{\mu_2 L_2} < (\delta - c)^{1-\sigma}$ and $n_1$ and $n_2$ are non-zero.

***Three Country Equilibrium*** $\triangleright$ Similarly to Krugman (1993), the three country example entails a strong simplifying assumption that will enable characterisation of the equilibrium. That is, all countries (and shares) have the same size $L_1 = L_2 = L_3 \equiv L$ so that the number of firms is also equalised: $n_1 = n_2 = n_3 \equiv n$. Essentially the home market effect between any two trading partners in a three country world is normalised to 1.

The impact of the network structure becomes apparent when a firm has to consider whether it will form a direct or an indirect connection to a trading partner. In order to form an indirect connection it must have implicitly formed a direct connection with another partner. Exploiting the symmetry assumption, any decision that a firm in country 1 may make is an equivalent decision for a firm in the complement countries. Therefore we focus on the decision of a firm in country 1 that has the option to trade directly with country 2 or indirectly with country 2 via country 3.

Similarly as in the two country case, the equilibrium with three countries is characterised by the zero profit condition $\pi_1 = \pi_2 = \pi_3 = 0$ irrespective of the types of connections formed and written as:

$$\sum_{1,j=1}^{3} \frac{q_{1j}}{v_{1j}} = \sum_{2,j=1}^{3} \frac{q_{2j}}{v_{2j}} = \sum_{3,j=1}^{3} \frac{q_{3j}}{v_{3j}} \Rightarrow Q_1 = Q_2 = Q_3 \tag{1}$$

While this expression may not be of particular interest, it is employed to extract inference regarding selection among two available network formations. This will occur through the differences across the price indices when alternate formations occur. Consider first the case of a network consisting only of direct connections.

***Direct Connections Network*** $\triangleright$ The net benefit term becomes $v_{ij} = \delta - c$ for all pairs in the network. The zero profit condition and the assumption of no home market effect equalises production output across countries. The price index that any country faces is an aggregate measure of domestic prices and imported prices given the types of connections established. For country 2 for example it can be expressed as:

$$P_2 = n\left(\frac{\sigma}{\sigma-1}\gamma\right)^{1-\sigma}\left(1 + 2(\delta - c)^{\sigma-1}\right)$$

***Indirect Connections Network*** $\triangleright$ In the case of indirect connections between countries 1 and 2 and direct connections with country 3, the same equilibrium condition (1) must hold. The price index with one indirect connection and one direct reads:

$$P_2 = n\left(\frac{\sigma}{\sigma-1}\gamma\right)^{1-\sigma}\left(1 + (\delta - c)^{\sigma-1} + (\delta^2)^{\sigma-1}\right)$$

where the term $\delta^2$ indicates the benefit from having a hub location intervening between countries 1 and 2. The two equilibria will be identical by the zero profit condition and the assumption of no home market effect if there are unique values of benefit $\delta$ and cost $c$ such that the two price indices are equalised across the two networks. This single point accommodates indifference between network formations; otherwise a specific network formation would prevail and the zero profit condition would be violated for one or both of the two network formations as will be shown below. Equalising the two price indices across formations, there is a unique pair of transportation cost $c$ and benefit $\delta$ that admits the equilibrium condition:

$$\delta - c = \delta^2$$

This unique cost level eliminates any benefit from choosing one particular formation such that the firm becomes indifferent between network formations.

It may also be the case that for a given value of benefit $\delta$ the values of transportation costs admit an equilibrium where only direct or indirect connections are formed. Consider a set of transportation costs ranked in ascending order, $\mathbf{C} = \{\ldots, \bar{c}, \ldots, \hat{c}, \ldots\}$ and $c \in \mathbf{C}$. Suppose that a high cost shock, $c$, is introduced between country 1 and 2. The two countries could continue trading directly. The profits for a firm in country 1 trading with 2 and 3 are (notation $D$ denotes a direct connections network):

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

Whilst the profits for the same firm if it chose to trade indirectly with country 2 using country 3 as a hub (notation $I$ denotes a network with one indirect connection) become:

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

The decision of the firm to change network formation arises by minimising losses given a constant benefit $\delta$ and a variable cost $c$. The indirect network formation prevails if the cost from forming a direct connection is very high. Then the firm may decide to sever the direct connection and begin trading indirectly. In this way it has the option to remain in the market otherwise see its profits decrease and exit the market. Setting the equilibrium condition to $0 = \pi_1^I > \pi_1^D$ determines when the indirect connections network formation arises. Solving the inequality yields the simple relationship $\delta - c < \delta^2$. Then denote as $c = \bar{c}$

the infimum of high transportation costs such that the inequality holds and the equilibrium condition is satisfied, $\bar{c} = \inf\{c \in \mathbf{C} : \delta - c < \delta^2\}$ and $0 = \pi_1^I(\bar{c})$. Then the equilibrium network is the indirect connections network. Given a high transportation cost $\bar{c}$ or above (as long as the cost is not high enough to induce autarky), it is more sensible for the firm to choose a hub network formation with the equilibrium holding only when $c = \bar{c}$. The hub formation thus minimises each country's exposure to transportation costs.

Alternatively, when there is a low transportation cost $c < \bar{c}$, the direct connections network prevails and a firm suffers losses if it is trading indirectly. Setting the equilibrium condition to $0 = \pi_1^D > \pi_1^I$ determines when the direct connections network emerges. As expected, it gives the simple solution $\delta - c > \delta^2$ implying $c < \bar{c}$. The equilibrium $0 = \pi_1^D$ will be satisfied when $c = \delta$. When connectivity costs are low it becomes beneficial to form all direct connections. The cost of adding a connection is less than the benefit the firm gains from shortening the connection of length two ($\delta^2$) into a connection of length one.

When the transportation cost is extremely high, none of these formations arise and countries become autarkic. The autarkic equilibrium requires that firm profits are negative for both formations simultaneously. If costs are such that $c > \delta + \delta^2$ then indirect trading is prevented and because $c > \delta$ direct trading is prevented. For the equilibrium to be autarky for all partners, due to symmetry, it must be that simultaneously both of the above statements are true. This holds when $c$ obtains the threshold value $\hat{c}$ or higher, where $\hat{c} = \inf\{c \in \mathbf{C} : c > \delta + \frac{\delta^2}{2}\}$.

These conditions coincide with Proposition 1 of Jackson and Wolinsky (1996) and are summarised compactly as follows.

For unique values of $c$ in the set $\mathbf{C}$ and holding constant the benefit term $\delta$, the network formation decisions for a representative firm in the symmetric trade model with increasing returns 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 $\delta = \bar{c} = \hat{c}$.

These formations are *uniquely efficient* in the sense that each case is a prevailing case and no other network can accommodate higher profits. If costs are forbidding it does not make sense for a firm (or for a consumer at the receiving end) to proceed with trading (consuming) a specific variety. The empty network, or autarky is the only efficient outcome of the three country problem. If costs are high but less than the autarkic level for a given level of $\delta < c$, the only efficient network is the hub network. Autarky would have lower utility levels and direct connections would give lower profits for the firm and lower consumption. For sufficiently low transportation costs, the cost of adding an extra connection is less than the firm's gain from replacing an indirect connection to a direct connection. And so it will always prefer to have a direct connection at these costs. The same applies for the consumer.

The hub network is *stable* for cost values consistent in the range of $\delta - \delta^2 \leq c$: country 3 being in the center, becomes worse off if a connection is severed since utility for consumers there decreases. A firm in country 1 similarly is adversely affected by this choice. The indirect connection is severed and the varieties traded decrease. Profits for the firm decrease. Therefore a firm will never choose to sever the connection with country 3. Suppose also that a firm in country 1 forms a direct connection at this cost level with country 2 instead of the indirect connection via country 3. Profits from this configuration become less and thus the firm will never choose to do so. If it actually did, a firm in country 2 would have to sever another direct connection with country 3 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 transportation costs all direct connections are pairwise stable as no country would be willing to sever a connection. Therefore any two countries which are not directly connected benefit from forming a connection.[7]

[7] See Jackson (2003) for the intuition behind the definitions of efficiency and stability. The proof for these statements is available in the Online Appendix.

This approach develops a very simple economic concept. Contrary to the Krugman (1993) three country trade model, countries which are not necessarily benefited from concentration of production, possibly created by historical incidence, can yield hub network formations as well. This arises by incidence merely of geographic advantage and as a form of hedge.

In the Krugman model of trade with three countries, the equilibrium arises by postulating concentration of production and a defecting firm to survey other countries' production possibilities. Instead, herein one can simply postulate an excessively large transportation cost and start to decrease it. At some autarkic liberating level, where for expositional clarity the benefit $\delta$ is such that there is no indifference between formations nor it is too low to admit autarky, the profit of a firm producing only domestic goods can be increased. This happens because there is a benefit from entering the export market. The firm decides to export, due to the positive profit margin. But it also decides the formation that minimises exposure to the exceedingly high costs it faces. The network formation will be a hub configuration with indirect connections and the number of firms enter the market driving profits to zero at a unique level $c = \bar{c}$ for a given $\delta < c$.

Each firm's labour input would need to be increased since $\frac{\partial L(q)}{\partial c} > 0$ and subsequently the number of firms or varieties would need to be decreased compared to a case when $c < \bar{c}$ holding $\delta$ constant ($\frac{\partial n}{\partial L(q)} < 0$). If it happened to be that costs are lower, and specifically when $\delta = c$, each firm's labour input $L(q)$ would be decreased freeing up units of labour. Due to the full employment condition, the available labour input translates into an increase in the number of firms or traded varieties. More firms entering the sector decrease profits to zero at the unique level $\delta = c$, yielding an equilibrium where trading only occurs directly between countries.[8]

# 5 Empirical Strategy and Endogeneity

This section derives the empirically testable expression for aggregate trade in the manufacturing sector of a country with representative firms. However, in place of iceberg transporta-

[8]See also Krugman (1979) for similar comparative static experiments.

tion costs, the net benefit term embedding network structure is introduced stemming from the theory. The presence of the net benefit term implies the existence of a trade off: ceteris paribus, a shorter distance to the destination dictates a lower cost and higher benefit whilst trading indirectly is associated with evermore discounted benefits at no additional costs.

An empirical problem arises however as the benefit from forming a connection is unmeasurable and does not reflect a realistic representation of a transportation cost. In order to yield a testable prediction, the aforementioned trade off is modelled as an endogenous routing decision that is common for all firms across sectors as exports are observed at high aggregation levels in the data. This decision can either be made by manufacturing firms themselves or a transport firm which posts the optimal shipping price that firms take as given. It will be expositorily easier to test the predictions of the model by employing the latter line of argument.

A firm's decision to trade via a hub is the result of transportation costs being sufficiently less to and from the hub versus the alternative decision to trade directly to the destination (Krugman, 1993). The derivation of lower total transportation costs depends on the form of the cost function of a transport firm which is assumed to incorporate a network structure and exhibits increasing returns to scale. If exports using a hub network prevail, fixed costs of transport are reduced because the network participation share for a trading partner is decreased. But distance, representing the marginal cost of transport, increases. This setting will be preferred against a network formation with direct connections holding constant export volumes. Otherwise the opposite should hold: for direct connections distance is shorter but fixed costs are higher and the overall costs for transporting the same export volume would be lower. We claim and prove in Appendix B that such a trade off is an outcome of increasing returns to scale in the transportation sector. This alternative setting confirms qualitatively, yet in a more cumbersome way, the simplistic theoretical exposition presented in the previous section.

For a particular sector, the gravity equation is defined as the value of exports from $i$ to $j$ of all firms belonging in this sector. It is equivalent to $x_{ij} = n_i p_{ij} q_{ij}$ or,

$$x_{ij} = \mu_i L_i \mu_j L_j \frac{v_{ij}^{\sigma-1}}{\theta_j} \frac{1}{\sigma F_i}, \text{ where } \theta_j = \sum_{j,l=1}^{N} n_l v_{lj}^{\sigma-1} \tag{2}$$

and $\theta_j$ is an aggregate index of network costs in $j$, derived from the price index. $F_i$ represents country $i$'s network participation share which we assume to be positively correlated with fixed costs. Since it is not possible to measure the net benefit from forming a connection, we decompose the problem into two parts. First, the transportation cost proxy is replaced with the distance between $i$ and $j$ assuming that transportation costs are of the form $c_{ij} = d_{ij}^{\beta} \times exp(\beta_0)$. Second, for every network formation there are fixed costs correlated with the network structure and burden all firms uniformly in a specific country.

Hence the gravity equation of exports between two countries involving also a hub will incorporate an increase in distance. It will also involve a reduction in fixed costs by lowering country $i's$ network participation share which acts as a benefit of forming this particular indirect connection. But if there is no hub involved, the gravity equation is just the standard outcome of the trade model with representative firms and fixed costs are proportional to forming and maintaining direct connections to all partners.

Suppose the hub location is $k$. A sector's firms in country $i$ trade with $j$ via $k$. Denote $x_{ij}^{I}$ as the aggregate exports using indirect connections, or the hub $k$. If $k$ is not involved, aggregate exports using direct connections are denoted $x_{ij}^{D}$ . Writing equation 2 in logarithmic form, sectoral export volumes for direct and indirect trading become respectively:

$$\ln x_{ij}^{D} = \beta_0 + \beta_1 \ln(\mu_i L_i) + \beta_2 \ln(\mu_j L_j) - \beta_3(\sigma - 1) \ln d_{ij} - \beta_4 \ln F_i^{D} - \ln(\sigma\theta_j) \tag{3}$$

and,

$$\ln x_{ij}^{I} = \beta_0 + \beta_1 \ln(\mu_i L_i) + \beta_2 \ln(\mu_j L_j) - \tilde{\beta}_3(\sigma - 1) \ln(d_{ik} + d_{kj}) - \tilde{\beta}_4 \ln F_i^{I} - \ln(\sigma\theta_j) \tag{4}$$

We can control for country size and the impact of relative prices, herein the aggregate network cost indices $\theta_j$, by using country fixed effects as in Chaney (2005). This operation however will absorb the variation of country specific fixed costs crucially rendering $\beta_4$ and $\tilde{\beta}_4$ useless in explaining any cost saving benefits of trading via the hub. Yet under such a specification we can conduct consistent and efficient estimation of the partial effects of the remaining regressors (Greene, 2008). Interacting the distance variable in equation 4 with a binary variable indicating the presence of a hub, allows to observe the marginal effects stemming from the presence of hub locations on a route. If, for this sector, there is a hub involved between two countries trading, it must imply that transportation costs were very high thus preventing direct connections with the opposite holding true holding export volumes constant. Equations 3 and 4 thus obtain their testable form as follows:

$$\ln \left[x_{ij}^{s}\right]^{D} = A_{ij}^{\prime s} + X_{ij}^{\prime} B_1 - \beta_3(\sigma - 1)\ln d_{ij} + \epsilon_{ij}^{s} \tag{5}$$

$$\ln \left[x_{ij}^{s}\right]^{I} = A_{ij}^{\prime s} + X_{ij}^{\prime} \tilde{B}_1 - \ln(d_{ik} + d_{kj}) \left(\tilde{\beta}_3(\sigma - 1) + \tilde{\gamma}\mathrm{Hub}_{ij}\right) + \xi_{ij}^{s} \tag{6}$$

Where prime denotes transpose, $A_{ij}^{s}$ is a vector comprising a constant, a set of country ($i$ or $j$) and sector ($s$) dummies, $X_{ij}$ is a vector of trade barriers between countries $i$ and $j$ other than distance and $\epsilon_{ij}^{s}$, $\xi_{ij}^{s}$ are both orthogonal to the independent variables and normally distributed. We assume that the shocks affect trade flows within each country pair and so all observations are clustered at this level. It may be possible that some hub locations are commodity/sector specific and thus the variation would be absorbed by the sector dummy. While we are not in the position to observe this in the data since routes are not commodity specific, we perform experiments using all possible combinations of dummies deducing that the source of the variation at least for the dataset used herein, is pair specific.

The two testable equations are not comparable as the level of export volumes that depend on the routing decision are different. But as the sectoral export volumes to a destination that are observed in the data are the maximum of the two equations, $\ln x_{ij}^{s} = \max\{\ln \left[x_{ij}^{s}\right]^{I}, \ln \left[x_{ij}^{s}\right]^{D}\}$, a comparison across equations is enabled by employing the common dependent variable. This allows for testing whether the two partial effects of distance are significantly different from each other.

While it is not immediately apparent how the coefficient of indirect distance should behave relative to the direct distance counterpart, the model dictates that the interaction should have an ameliorating effect on the negative impact of the former variable. The aim therefore is to test if the overall marginal effect of trading via a hub is significantly less than the marginal effect of trading directly, in absolute values, reflecting a discount in transportation costs due to the interaction. With this strategy in mind, we build the dataset followed by a discussion on expected magnitudes of the consistent estimators of $\beta_3$ and $\tilde{\beta}_3$.

# 6 Construction of the dataset

The dataset is constructed by matching export data obtained from the UN Comtrade Database with corresponding information on maritime transport costs of the same Harmonised System (HS) classification and year. The latter are sourced from the OECD Maritime Transport Costs Database.

The justification for selecting this particular matching experiment 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 and spoke networks induced by the advent of containerisation, it is not strict to assume that the observed price of shipping services is a function of the network organisation of the transport sector (Rua, 2014). By conducting this operation some of the global export volume that has been transported by sea and therefore by some form of network can be captured.

All observations not having matching exports and transportation costs are removed.[9] Each surviving trade partnership is assigned a measure of great circle capital distance and common border, common official language, past colonial relationship and active regional trade agreement binary dummies from the CEPII GeoDist dataset (Mayer and Zignago,

[9] This operation may not be innocuous. To this end, I have repeated all experiments without the matching of exports with observed maritime transport costs. The results are in support of those presented herein however the identification of an effect stemming from the organisation of a particular transport network disappears as more than one modes of transport are candidates for shipping goods. These results are available upon request.

2011).[10]

We construct manually a measure of indirect capital distance in addition to geodesic capital distance.[11] For a random subset of trading partners presented in the Online Appendix we measure the distance from the capital of the exporter/importer to the closest domestic major exporting/importing port. For cases of landlocked countries the closest major foreign port is chosen, through which it may proceed to trade. Port to port distances are measured using the U.S. National Imagery and Mapping Agency Distances Between Ports publication and the online resource Port World Distance Calculator.

The selection of each country's main ports is made on the basis of throughput. A list of each country's main ports and associated sources in support of their domestic ranking is available in the Online Appendix.[12]

Unfortunately there does not exist a source that describes the routes utilised to ship goods from an origin to destination by each transport mode. Nor is there a proxy for the number of ships on a specific route within a specific time period. This leads to ambiguity as to the existence, followed by the identification, and therefore selection, of good-specific routes and their respective measurement of distance. Thus the distances that are described herein are country pair specific.

The lack of route description also makes the selection of unanimously accepted hub locations difficult. Any location benefited either from geographical advantage or historical precedence (Krugman, 1993) can serve as a hub for a particular hinterland or as a transshipment hub. 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. Moreover hubs can be large or small enough to be potential determinants of the distribution of trade routes internationally, regionally and domestically.

Each indirect capital distance observation is assessed in order to infer whether the route

[10]Regional trade agreements for year 2007 are own constructs.
[11]Henceforth great circle capital distance will be distinguished as direct capital distance.
[12]The indirect capital distance matrix is available upon request.

requires passage through one or more elements of an exogenously defined set of locations. These locations are termed as hubs, which are big enough to potentially affect the route distribution for the international trade of goods. 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:

- The ports of Arica, Chile and Paranagua, Brazil: Due to their entrepôt status for Bolivia and Paraguay respectively (UNCTAD Secretariat, 1994).
- The area around Cape Town, South Africa: Because of its position which facilitates the transit of cargo for South-South trade (African Development Bank, 2010).
- The Gibraltar area: As the port of Algeciras is considered a transshipment hub and a European hub (Notteboom, Parola and Satta, 2014).
- Istanbul: Which serves as a transshipment area for countries with access to the Black Sea (Merk and Bagis, 2013).
- Panama Canal, Panama and Port Said, Egypt: Because of their geographical location and transshipment facilities at the entrances of the canals (Notteboom, 2012; African Development Bank, 2010).
- Singapore: As it is the world's largest transshipment hub (Port of Singapore Authority, Cullinane, Yap and Lam (2006)).

If the route requires passage through one or more of these areas then the indirect capital distance variable is assigned an indicator value equal to one or zero otherwise. No such distinction is made for direct capital distance. Thus we have built the variables required to alternately test equations 5 and 6 and subsequently compare them. To provide an example, Figure 1 illustrates the distinction between the two distance variables. Hollow circles refer to the hub locations used to test the empirical specification. The direct capital distance between Beijing and Brasilia is 16,948 km. Indirectly, the distance from Beijing to Shanghai is 1,267 km, and from Shanghai to Singapore 3,934 km, where the indicator is assigned a

Figure 1: Hub locations, direct and indirect capital distance

value of 1. Then add the distance from Singapore to Rio (16,366 km) and Rio to Brasilia (1,160 km). The observation for indirect distance finally becomes 22,727 km. We also keep track of the number of hubs required to reach the destination which in this case is 2. These are used in a specification that departs from the strict confines of the testable prediction. It involves calculating the marginal effect of indirect capital distance and the number of hubs on a route. It is further discussed in section 7.2.

The time period of the sample is 1991-2007, which is short enough to ensure that there are no changes to the hub status of the selected locations. Estimations using shorter time periods did not alter the outcomes. Three aggregation levels of the Harmonised System are employed in the empirical experiments: The total (country) level of trade,[13] the chapter (HS 2) and subheading (HS 6) levels. Information on maritime transport costs at the heading level was not available thus no matching occurs. Since the distance variables are not country-pair-commodity specific, it is possible that the number of chapters or subheadings traded between pairs gives a disproportional weight to particular routes, and therefore hub(s), relative to alternatives. This is dealt with by incorporating commodity fixed effects and combinations thereof at both HS 2 and 6 levels.

The data come with a weakness in the sense that transportation costs for individual E.U. 15 countries are not observed. For this reason, and when otherwise not available, all other

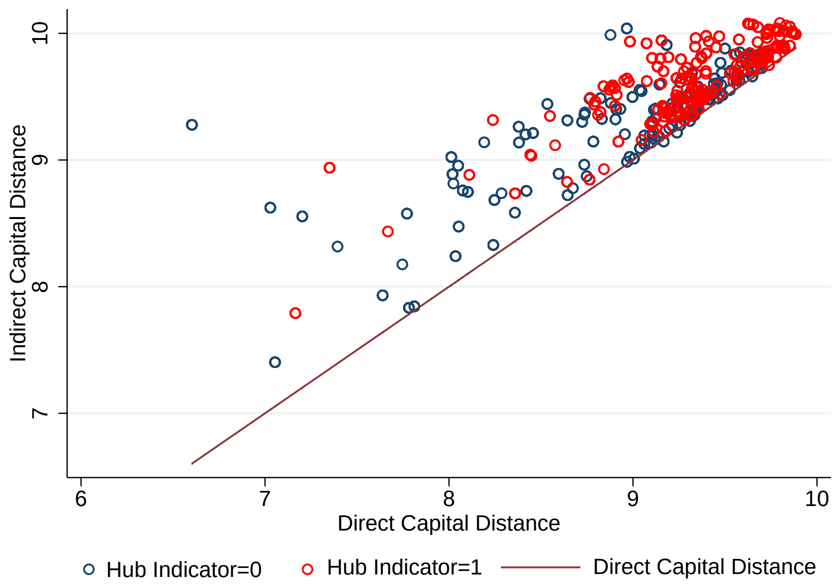
[13]These are own constructs from country pair summations of chapter levels.

units are aggregated to provide approximations at the aggregate level. The capital distance of the E.U. 15 area with the rest of the world is then measured from Brussels and its main export/import port becomes Rotterdam. Estimations will be 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.

## 6.1 Description of the data

The distribution of direct and indirect capital distances is exhibited in figure 2, and reveals a systematic bias of routes utilising a hub over longer distances. It also illustrates that the dispersion of indirect capital distance diminishes the further two locations become.

Figure 2: Distribution of distances and hubs




This distribution is dissected in table 1, which depicts the breakdown of frequency of flows transiting a hub area at the three aggregation levels. Particular routes are found to be thick since the number of different commodities within a flow plays a role in determining the volume of flows that pass from a hub: At the total trade level 48% of trade flows required passage from one or more hub locations, but this number declines to 35% and 24% for the HS 2 and 6 levels respectively. Approximately 65%, 33% and 2% of flows involve passage through one, two and three hub locations respectively.

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 | | | | |

Addressing further the systemic bias observed in figure 2 which was associated with longer trade routes, figures 3 and 4 exhibit how the difference between two distance variables increases slightly by changes in indirect capital distance, but decreases in direct capital distance. The implication is that direct capital distance observations approach in magnitude their indirect capital distance counterparts over longer distances in the sample. Thus by overlapping the two distributions of distance, we infer that the right tail of the distribution of direct capital distance approximates that of indirect capital distance and as a consequence, the variance of direct capital distance must be larger.

Such a finding prepares for the fact that the coefficient of indirect capital distance could possibly be weakly larger than the coefficient of direct capital distance, in absolute value.

Figure 3: Distance differential as function of indirect distance (logs).

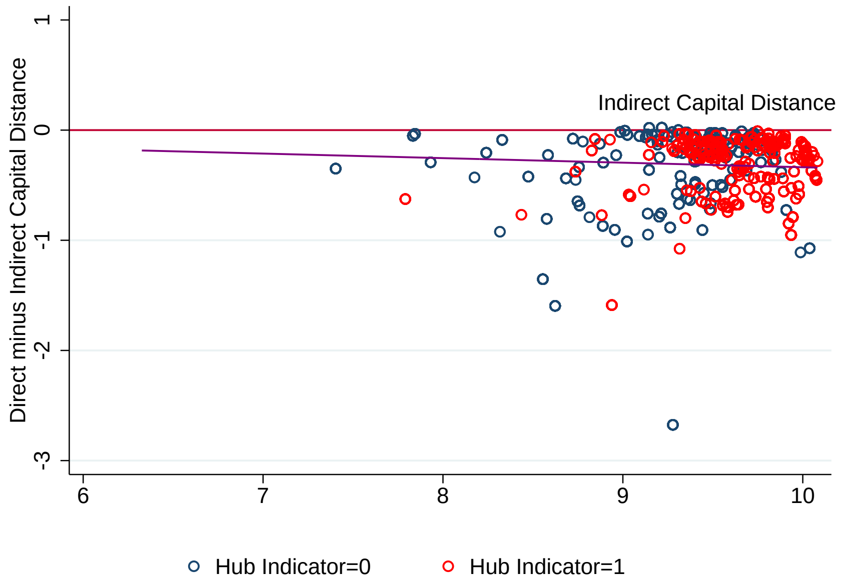
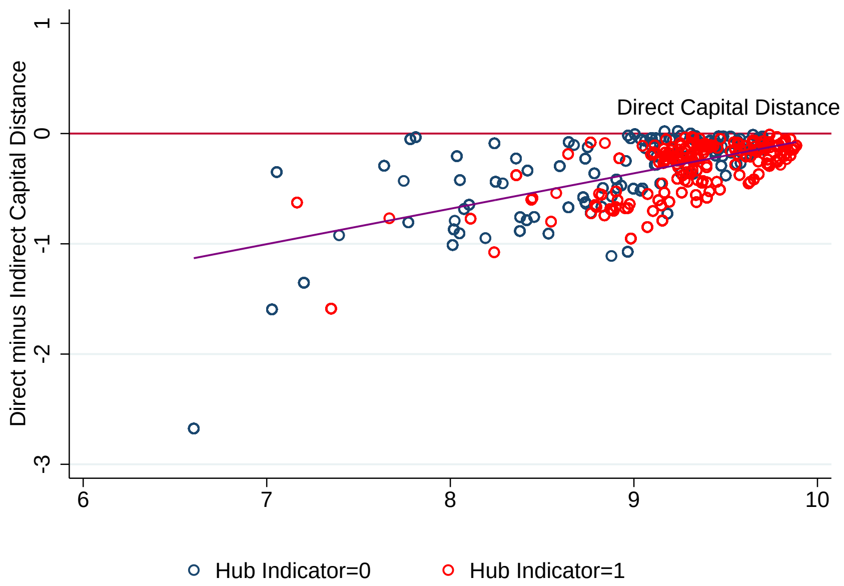


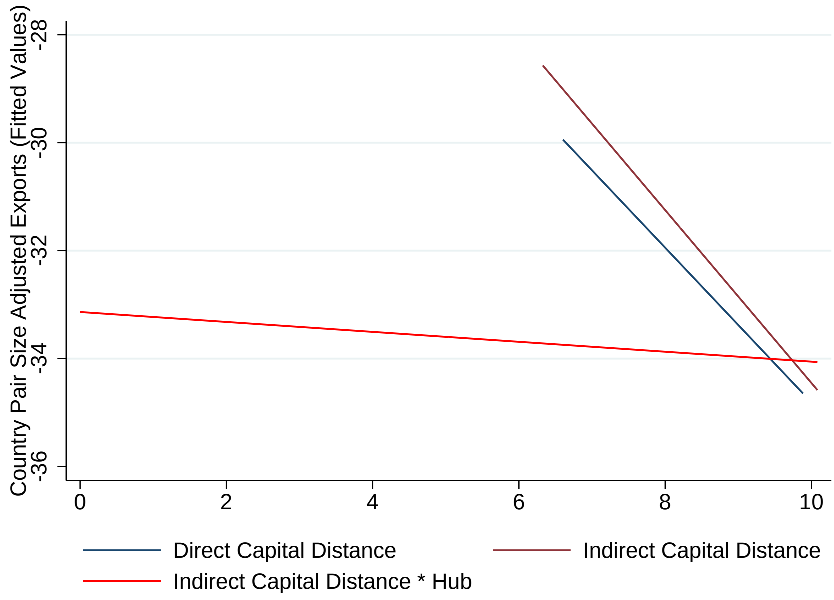

Figure 4: Distance differential as function of direct distance (logs).




Adding the marginal effect of the interaction of indirect capital distance with the hub indicator dummy permits inference as to whether the total marginal effect of indirect capital distance has a different impact on the same export volume. This is obtained by testing which of the three individual marginal effects is lower in absolute value and whether this difference is statistically significant.

Figure 5 concludes the section by preliminarily revealing that the interaction term induces

Figure 5: Size-adjusted exports as a function of distance (Total Trade Level, logs)




the overall reduction of the indirect capital distance effect compared to that of direct capital distance at the total trade level.

# 7 Results

The analysis commences with a comparison of the distance coefficients $\beta_3$ and $\tilde{\beta}_3$ at the total trade, HS 2 and 6 levels and the outcomes of testing for their significant difference in magnitudes. It is achieved by estimating equation 5 alternately by including the direct or indirect capital distance variables.

Table 2 summarises the coefficients of these variables for twelve different estimations across the three aggregation levels in the presence of common border, official language, past colonial relationship and regional trade agreement variables. Estimations using combinations of two and three of these control variables were carried out and the results validate those presented in table 2.[14] All estimations are conducted using ordinary least squares and standard errors are allowed to be correlated within country pair clusters. Columns (1) - (6) and (7) - (12) are characterised by varying combinations of country, year and commodity fixed effects. The latter group of experiments exclude the E.U. 15 area trading with the

[14]These results are available upon request.

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 | | | Raw | | | | | | 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 Online Appendix. *** p<0.01, ** p<0.05, * p<0.1.

United States. The tables containing the estimation results inclusive of all control variables and country pair clustered standard errors are relegated to the Online Appendix.

The null hypothesis states that the impact 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, with three exceptions in 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 capital distance are found to be weakly larger in absolute value than their direct capital distance counterparts in the raw sample when exporter/importer-commodity-/year fixed effects are employed yet this is not consistently observable across other specifications and when excluding the E.U. 15 - USA pair. The goodness of fit comparison between specifications including either the direct or indirect capital distance variables is identical up to, and in some cases including, the third decimal place. Lastly the coefficients for direct distance all fall within the range of surveyed estimates in the

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** | | | **Raw** | | | | **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 Online Appendix.

literature (Disdier and Head, 2008; Overman, Redding and Venables, 2001). While we cannot classify the magnitudes of the indirect capital distance variable, their indistinguishable impact and their levels suggest that they do not diverge from the acceptable ranges.

Table 3 exhibits the outcomes of equations 5 and 6. The respective samples and combinations of fixed effects that were outlined in Table 2 are preserved. The individual tables containing the regression results are available in the Online Appendix. The first row replicates the direct capital distance coefficients from table 2. The second row of each aggregation level depicts the marginal effect of indirect capital distance plus the interaction with the hub indicator function of equation 6 for the same sample as the row above it. The third row illustrates the discounted cost of trading indirectly compared to trading directly.[15] Dashes pertain to a lack of result owing to the fact that the interaction term is insignificant. First we invoke a set of F-test results whereby the null hypothesis $B$ that the joint impact of the coefficient of indirect distance and the interaction is zero is rejected across estimations and aggregation levels, when the interaction is significant.

The table reveals that in most cases, and when the interaction is significant, the joint impact of indirect capital distance and its interaction with the hub indicator is lower in absolute value than the impact of direct capital distance. In all cases of the total trade and in columns 6 - 12 of the latter two aggregations levels, the joint impact is significantly less than the impact of direct capital distance as the null hypothesis $C$ is rejected. It is deduced that the hub indicator appears to be dampening the impact of indirect capital distance, while the goodness of fit is identical up to, and sometimes including, the third decimal place.

The HS 2 and 6 level outcomes in column 12 of Tables 2 and 3 allow pair variation only. In the first instance it is found that a 1% increase in direct capital distance reduces export volumes ceteris paribus by about 1.8% and 1.1% respectively. When trading indirectly (without assuming any presence of a hub) the marginal effect on exports stands at about 1.6% and 1%, however 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).

Including the interaction with the hub indicator, we now observe that the overall marginal effect of indirect capital distance has reduced to about 1.4% and 0.8% for every 1% increase. Compared to trading directly, trading via a hub can have ameliorating effects as there is a 23% and 32% saving 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 results further indicate that the performance and signs of the capital distance variables and controls, are those expected in the majority of specifications and are in accordance

[15] Calculated as $\left(\frac{\hat{\beta}_{\text{Indirect Capital Distance}}+\hat{\beta}_{\text{Interaction}}}{\hat{\beta}_{\text{Direct Capital Distance}}}-1\right)*100.$

to the magnitudes observed in the empirical trade literature.[16] Yet identification of the sign and magnitude of the marginal effect of the hub indicator when interacted with distance is unresolved. A possible explanation for the observed outcome is that the hub indicator is correlated positively with distance and by extrapolation negatively correlated with size adjusted (or controlled for) exports. Yet the negative correlation weakens over longer distances and turns positive possibly capturing the high volume of trade between the East Asian countries with the countries in the Atlantic basin. These flows occur exclusively via a hub such as Singapore or the Suez Canal. Therefore it is understood that the coefficient of the interaction appears to be derived by picking up the pair specific variation associated with longer trade flows and higher volumes. This fact cannot be identified using traditional distance variables in empirical trade experiments.

The model's prediction of the existence of benefits arising from trading via a hub location is confirmed. Whilst the reasoning for utilising a hub area lies not in the explanatory power provided by hub indicator variable itself, this acts only as a proxy capturing lower bilateral network costs, we are able to recover a positive impact on trade flows.

## 7.1 Addressing Endogeneity

The hub indicator variable employed in the analysis so far was assumed to be uncorrelated with the error term. It served to reduce the omitted variable bias by acting as an element of the vector of trade barriers. Yet because it is a proxy for endogenous route selection as the theoretical exposition asserted, one cannot rule out correlation with the independent variable and correlation with measures of remoteness.

Performing a two stage least squares regression by instrumenting the endogenous dummy variable with an appropriately selected instrument $Z$ would most likely lead to a case of "forbidden regression": The first stage entails estimating non linearly the hub indicator as $\mathrm{Hub}_{ij} = \Phi\left(\pi_1' X_{ij} + \pi_2' Z_{ij}\right)$, where $\Phi$ is the cumulative distribution function, and obtaining the non linear fitted values $\widehat{\mathrm{Hub}}_{ij}$. These could be utilised in the second stage by substituting the indicator and interacting it with the indirect capital distance variable but the residuals

[16]E.g. Limão and Venables (2001).

would most likely be correlated with the fitted values and covariates when applying ordinary least squares. To avoid this occurrence the non linear fitted values $\widehat{\text{Hub}}_{ij}$ could be used as instruments themselves for $\text{Hub}_{ij}$. But this again would imply that the second (linear) stage relies, albeit indirectly, on a non linear source of information (Angrist and Pischke, 2009).

We avoid this risk by considering the estimation of $\beta_3$ of equations 5 and 6 and relegating the interaction of the latter equation in the error term $\tilde{\xi}^s_{ij} = \tilde{\gamma}\text{Hub}_{ij}\ln(d_{ik} + d_{kj}) + \xi^s_{ij}$. Given the magnitudes of the two distance variables' marginal effects are indistinguishable through table 2, the problem is transformed to estimating $\beta_3$, $\tilde{\beta}_3$ when the interaction, or any other indicator that dictates trading via a specific route, be it direct or indirect, is 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: It captures characteristics related to remoteness and information frictions of trading partners in the spirit of Egger and Larch (2013), and in that sense, is an element of the vector of trade barriers in the gravity equation, yet it is agnostic about the routes utilised to ship commodities between country pairs.

The time difference variable comes from the CEPII Gravity dataset (Head, Mayer and Ries, 2010). To address for country pairs that do not have time difference, two sets of experiments are used: One where this observation is left untreated, and one where the zeros are replaced by an infinitesimal number. The distance coefficients summarised in table 4 refer to the second stage of the latter case, while their untreated counterparts and all first stage output are available in the Online Appendix. The experiments were also performed using combinations of two and three controls and exhibit similar results.[17]

The underidentification test and Cragg-Donald statistics indicate that across aggregation levels the candidate excluded instrument is relevant and not weak and the specification is exactly identified. The measures of goodness of fit are identical up to, and sometimes inclusive of, the third decimal place. Coefficients of the two distance variables are significant when all sources of heterogeneity are accounted for leaving only identification of the effect through country pair variation. The magnitudes are inflated relative to their ordinary least

[17]These are available upon request.

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** | **Raw** | | | | | | **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 Online Appendix.

squares counterparts and this may be attributed to the small but significant values of the coefficients of the first stages, as this case is not observed when zero time differences are left untreated.

The indirect capital 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, similarly to the corresponding p-values in table 2. The implication is that the reported magnitudes are devoid of information pertaining to route selection and convey information solely about remoteness.

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** | | | | | | | | | | | | |
| | Indirect Capital Distance, #Hubs=0 | -1.937*** | -1.919*** | | | | | -1.557*** | -1.533*** | | | | |
| | $R^2$ | 0.950 | 0.958 | | | | | 0.937 | 0.953 | | | | |
| A | Indirect Capital Distance,#Hubs>0 | -0.656 | -0.950 | | | | | -1.316** | -1.682** | | | | |
| A | Interaction | -0.392 | -0.163 | | | | | -0.560 | -0.297 | | | | |
| A | Number of Hubs | 0.0446 | -0.0524 | | | | | 0.356 | 0.251 | | | | |
| A | $R^2_A$ | 0.919 | 0.943 | | | | | 0.904 | 0.941 | | | | |
| B | Indirect Capital Distance,#Hubs>0 | -0.721 | -0.878 | | | | | -1.758*** | -1.985*** | | | | |
| B | Interaction | -0.306 | -0.261 | | | | | 0.113 | 0.164 | | | | |
| B | $R^2_B$ | 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** | | | | | | | | | | | | |
| | 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 |
| A | 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*** |
| A | 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 |
| A | 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 |
| A | $R^2_A$ | 0.253 | 0.267 | 0.434 | 0.449 | 0.766 | 0.868 | 0.200 | 0.213 | 0.406 | 0.423 | 0.734 | 0.851 |
| B | 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*** |
| B | 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*** |
| B | $R^2_B$ | 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** | | | | | | | | | | | | |
| | 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 |
| A | 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** |
| A | 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 |
| A | 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 |
| A | $R^2_A$ | 0.142 | 0.152 | 0.329 | 0.338 | 0.726 | 0.878 | 0.118 | 0.131 | 0.355 | 0.365 | 0.703 | 0.872 |
| B | 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** |
| B | 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 |
| B | $R^2_B$ | 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** | **Raw** | | | | | | **No USA-EU15 Bilateral Trade** | | | | | |

Reported coefficients 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. *A* indicates a set of regressions including indirect capital distance and the interaction with the number of hubs on a route as independent variables. *B* indicates a set of regressions s 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 Online Appendix.

## 7.2 Robustness of findings

The conducted experiments remain close to the theoretically derived testable equation and appear to validate the predictions of the model. Comparing different distance variables across specifications to check for the existence of a network effect may be a case of mispecification. A candidate model involving only one distance variable is considered. We estimate the marginal effect of indirect capital distance for a subsample characterised by no hubs intersecting the route between exporter and importer. We repeat for the complement set by also estimating the marginal effect of the number of hubs on the route and the interaction with indirect distance: It is possible that the existence of a hub location may amplify traded volumes but the presence of more than one hub location may diminish such gains due to the additional time required to reach the destination.

Table 5 summarises the findings. Within each aggregation level the first row refers to coefficients of indirect capital distance for the subsample that does not have a hub location on the route. Rows distinguished by $A$ pertain to the complement and include the distance and categorical variable and their interaction. Rows identified by $B$ consist only of the distance variable and its interaction with the number of hubs on the route. The estimation results are available in the Online Appendix.

The table reveals that trading partners which do not trade via a hub have a much lesser impact on exports than partners who do. For the latter, this impact is dampened as the interaction term is increasing in the number of hubs to the destination in parts $B$ of the table. The number of hubs does not seem to play an independent role in shaping trade flows except for three out of twenty eight experiments. Use of a hub location can amplify traded volumes but trading partners will never be as competitive as their less remote counterparts who do not need to utilise a hub location. This observation supports the findings of the model-driven specification.

Further concerns regarding the choice and identity of hub locations together with issues associated with sample selection bias are addressed in Appendix C.

# 8 Conclusion

We study the reasons that lead to formation and, the impacts of transportation hub networks in international trade by merging the symmetric connections model of Jackson and Wolinsky (1996) in a trade model of monopolistic competition with representative firms. Firms commence exporting and choose the network formation that is minimising their exposure to transportation costs. When transportation costs are extremely high countries remain closed. Upon their gradual reduction, exporting firms create hub networks that are associated with higher levels of costs and then form direct connections. The equilibrium is characterised given a fixed benefit value for which a unique transportation cost leads to satisfaction of the zero profit condition.

The theoretical prediction is tested using an endogenous route selection process. Firms choose the formation that maximises the volume of output given two measures of capital distance, one direct and one indirect with the latter interacted with a hub binary indicator. Trading via a hub is found to be preferable over longer distances where the interaction term provides an ameliorating effect on the distance barrier reducing its overall impact to being less than that of trading directly. As the indicator function is endogenous, we use time difference as the excluded instrument since it does not contain information about routes that could define distance to the destination.

The driving force behind hub formations appear to be economies of scale in transportation due to sectoral export volumes and their interaction with distance. Thus transportation costs on high-volume trading routes tend to be low. Transportation costs on low-volume trading routes, tend to be higher and remoteness here plays a crucial role affecting negatively the output of exporting firms. Additional factors could be directional imbalances penalising countries which cannot provide return cargoes, costs for importing and exporting commodities and exercise of market power.

It is concluded that geographically disadvantaged countries absorbing high transportation costs achieve a more beneficial trading position when forming a transportation network by utilising 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.

Potential avenues for future research in this area become i) measuring the change in the extensive margin of trade following the emergence of a hub location; ii) deriving a functional form for the hub indicator by observing the origin and destination of commodities passing through, along with the number, size and direction of trade of vessels on routes; iii) analysing the heterogeneity in fixed costs associated with infrastructure for which information is not currently available.

# 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.

Baier, S. L. and Bergstrand, J. H. (2001), 'The growth of world trade: tariffs, transport costs, and income similarity', *Journal of International Economics* **53**(1), 1 – 27. doi: http://doi.org/10.1016/S0022-1996(00)00060-X.

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.

Chaney, T. (2014), 'The network structure of international trade', *American Economic Review* **104**(11), 3600 – 3634. doi: http://doi.org/10.1257/aer.104.11.3600.

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. (2004), *Advanced International Trade: Theory and Evidence*, 1st edn, Princeton University Press, Princeton.

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.

Fujita, M., Krugman, P. and Venables, A. (1999), *The Spatial Economy: Cities, Regions and International Trade*, 1st edn, The MIT Press, London.

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. and Mayer, T. (2014), Gravity Equations: Workhorse, toolkit, and cookbook, *in* E. Helpman, K. Rogoff and G. Gopinath, eds, 'Handbook of International Economics', Vol. 4 of *Handbook of International Economics*, Elsevier, pp. 131–195.

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. (2008), *Social and Economic Networks*, Princeton University Press, Princeton.

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. (1979), 'Increasing returns, monopolistic competition, and international trade', *Journal of International Economics* **9**(4), 469 – 479. doi: http://doi.org/10.1016/0022-1996(79)90017-5.

Krugman, P. (1980), 'Scale economies, product differentiation and the pattern of trade', *The American Economic Review* **70**(5), 950 – 959.

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.

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.

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, C.E.P.R., No. 2978.

Pascali, L. (2014), The wind of change: Maritime technology, trade and economic development. Department of Economics and Business, Universitat Pompeu Fabra, Economics Working Papers, No. 1428.

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.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, 1st edn, The MIT Press, London.

# A Exposition of a simple network

## A.1 Connection types

For $N = 3$ symmetrically placed countries assume that countries 1 and 2 are at the edges and 3 is in the middle. There can be two types of available networks. One network formation is direct connections between all participants. Then the network is defined as $G = \{12, 21, 13, 31, 23, 32\}$. The second formation is an indirect connection between 1 and 2 and direct connections from and to country 3, such that $G = \{13, 31, 23, 32\}$. The net benefit term between countries 1 and 2 becomes for the case of direct connections $v_{12} = \delta - c = v_{21}$. For the case of an indirect connection between 1 and 2 we have: $v_{12} = \delta^2 - 0 = v_{21}$. The latter indirect connection implies the existence of two 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 countries. Country 1's set of direct connections is $N_1 = \{12, 13\}$ and the cardinality of the set is $n_1 = 2$. The total number of direct connections is 6, and country 1's fixed costs associated with the network are $F_1 = 1/3$. Equivalently for the case of indirect connections between 1 and 2 we have $F_1 = 1/4$, since country 3 in the middle is burdened by the additional share $F_3 = 1/2$.

# B A model of increasing returns to scale in transportation and hub formations

A three country model with increasing returns in each country's transportation sector is constructed in order to prove the existence of a trade off between an increase in distance due to indirect trading and the reduction in fixed costs of transportation operating under increasing returns. This theoretical finding, albeit derived in a more cumbersome manner, has qualitatively the same effects as the theoretical exposition of the paper and yields the same conclusions that lead to the empirical prediction.

## B.1 Model setup

The model opens directly in costly trade. Instead of a network, it will suffice to consider a set of countries $K = \{1, 2, 3\}$ that exist in a world where there is symmetry to and from country 3 and an asymmetry between countries 1 and 2. Country 3 shall be in the middle in order to be consistent with the main theoretical model. The asymmetry is measured in terms of distance and therefore: $d_{13} = d_{23} = d < d_{12} = d'$.

All countries are identical technologically and in size. The latter assumption is imposed as in Krugman (1993) in order to set aside the home market effect. An arbitrary country has population $L$ and three sectors, Agriculture, Manufacturing and Transport. The agricultural good is homogeneous, produced under constant returns, and is the numeraire good. The manufacturing good is produced under monopolistic competition and some quantity of the good produced is exported to the other two countries using domestically produced transport services. Transport services are produced under monopolistic competition and are utilised solely for transporting the exporting volume to the importer. As in Krugman (1993) we can allow for mobility of labour between the constant returns and increasing returns sectors but need to impose a fixed labour share in transportation. As such labour is exhausted in employment in the three sectors.

***Demand*** $\triangleright$ Agents in country $i \in K$ notwithstanding their sector of occupation, consume differentiated varieties of agricultural and manufacturing goods attaining utility,

$$U = q_0^{1-\mu}\left[\int_{\omega\in\Omega} q_{ij}(\omega)^{\frac{\sigma-1}{\sigma}}d\omega\right]^{\frac{\sigma}{\sigma-1}\mu}$$

where $q_0$ is consumption of the aggregate agricultural good, $1-\mu$ is the share of income expenditure on 0, $\sigma > 1$ is the elasticity of substitution between pairs of varieties and $\Omega$ is the mass of available goods. Maximising utility subject to exhausting her labour income share, the representative consumer in country $j$ has demand for differentiated goods:

$$q_{ij} = \frac{\mu_j L_j p_{ij}^{-\sigma}}{\sum_{j,i=1}^{N}\int_\Omega p_{ij}^{1-\sigma}(\omega)d\omega}, j \in K$$

It will be notationally convenient to define $\theta = \frac{\sigma-1}{\sigma}$, $0 < \theta < 1$ as the intensity of the preference; as $\theta$ approaches 1 varieties become almost perfect substitutes. As $\theta$ approaches zero an increased number of varieties results in higher utilities.

The demand function can then be rewritten as $q_{ij} = \frac{\mu_j L_j p_{ij}^{\frac{1}{\theta-1}}}{P}, j \in K$, where P will represent the price index.

***Manufacturing Production and Trade Costs*** $\triangleright$ Good 0, the agricultural good, is the numeraire homogeneous good. Its price is normalised to 1, the wage rate is then equal to the price of the good. In this respect the wage rate is equal to 1 across countries due to free trade, and across the three sectors within each country.

One manufacturing firm can produce one variety of the differentiated good using labour and transportation as an intermediate input only for exporting. Labour costs for differentiated goods are split between a marginal and a fixed cost and thus the sector is characterised by increasing returns.

To produce and sell a variety $\omega$ either domestically or abroad, the firm in country $i$ employs labour input:

$$L^m(q) = \gamma q_{ij} + F_i, \quad j \in K, \gamma, F_i^m > 0$$

There is full employment in manufacturing so that the sum of labour used in manufacturing production constitutes the sector's labour share.

***Pricing Regime*** $\triangleright$ *C.i.f.* prices of imported goods consist of a multiplicative iceberg cost

$\tau_{ij} \geq 1$ and an additive transportation cost $f_{ij}$ that is the optimal price set by the transport firm (Hummels and Skiba, 2004; Irarrazabal, Moxnes and Opromolla, 2015):

$$p_{ij} = p_{ii}\tau_{ij} + f_{ij}$$

***Transportation Production*** $\triangleright$ Equivalently to manufacturing, the transportation sector produces a continuum of differentiated transport varieties that are used as an intermediate input in manufacturing in order to facilitate exports. One transport variety is utilised to transport the output of one exporting variety.[18] Each country uses transport services produced domestically. Each specific transport variety is produced by a single firm using labour as its input. All firms have the same cost function, can freely enter or exit production and each consumer is endowed with one unit of labour. The production function is

$$L^t(q) = d_{ij}q_{ij} + F_i^t, \quad i \neq j, d, F_i^t > 0$$

where $F_i^t > 0$ is a varying overhead/fixed cost by location, $d_{ij} > 0$ is a constant marginal cost of transport production that will be associated with distance to the trading partners. $q_{ij}$ denotes the quantity of output that each transport firm can carry and comprises of the total export volume produced by one manufacturing firm.

## B.2 Partial equilibrium in manufacturing

Manufacturing firms in country $i$ maximise profits subject to feasible output. Provided $i \neq j \in K$, their profit function is defined as

$$\Pi_i^m = p_{ii}q_{ii} + \sum_{i,j=1}^{3} p_{ij}q_{ij} - w\gamma q_{ii} - \sum_{i,j=1}^{3} w\gamma\tau_{ij}q_{ij} - \sum_{i,j=1}^{3} f_{ij}q_{ij} - wF_i^m$$

where the transport revenue obtained from exporting to country $j$ is passed directly to the transport firm. Maximising profits subject to the demand for a domestic good, the profit

[18]This assumption could be too strong. I have shown elsewhere, but omit to prove herein, that if one permits homogeneity of degree greater or less than 1, then transport services can be used to carry more than or less than the exporting output produced by one manufacturing firm. Nevertheless the qualitative results would remain unchanged. This proof can be provided upon request.

maximising price becomes $\frac{p_{ii}}{w} = \frac{\gamma}{\theta}$ which constitutes the MR=MC condition. Free entry and exit of firms results in zero long term profits for each manufacturing firm and fulfills the P=AC condition $\frac{p_{ii}}{w} = \gamma + \frac{F_i^m}{x_i}$, where $x_i$ is the total output produced by each firm.

The equilibrium is characterised when simultaneously marginal revenue equals marginal cost and price equals average cost. The equilibrium manufacturing output is constant amounting to:

$$x_i = \sum_{i,j=1}^{3} q_{ij}\tau_{ij} = \frac{F_i^m}{\gamma}\frac{\theta}{1-\theta}$$

The number of firms can then be derived due to full employment in the manufacturing sector:

$$n_i^m = \frac{\mu_i L_i^m}{F_i^m}(1-\theta)$$

## B.3 Partial equilibrium in transportation

Transport firms, simultaneously to manufacturing firms, maximise profits subject to feasible total export output produced by manufacturing firms. They obtain their revenue via the *c.i.f.* price of the manufacturing good and the intermediate input assumption. Provided $i \neq j \in K$, their profit function is

$$\Pi_i^t = \sum_{i,j=1}^{3} f_{ij}q_{ij} - \sum_{i,j=1}^{3} wd_{ij}q_{ij} - wF_i^t$$

We assume the simultaneous pricing and output determining behaviours of manufacturing and transport firms. Equivalently the manufacturing firm would observe the equilibrium value of $f$ as both entities play simultaneously and have no reason to deviate from their optimal decisions, since labour shares are fixed and the wage is equalised across sectors.

Transport firms yield prices that are a function of the *f.o.b.* price and a markup over the marginal cost of transport due to the elasticity of import demand with respect to transportation costs:[19]

[19] $\epsilon_f = -\frac{\partial q_{ij}}{\partial f_{ij}}\frac{f_{ij}}{q_{ij}} = \sigma\frac{f_{ij}}{p_{ii}\tau_{ij}+f_{ij}}$

$$\frac{f_{ij}}{w} = \frac{d_{ij}}{\theta} + \frac{\gamma\tau_{ij}}{\theta}\frac{1-\theta}{\theta}$$

Free entry and exit of firms result in zero long term profits. However the imposed asymmetry between countries 1 and 2 will prevent the export shares being equal for all countries in $K$. Crucially this fact may give rise to a hub formation.

The transportation price characterisation allows then to write the *c.i.f.* price, reminding that $w = 1$ for all countries:

$$p_{ij} = \frac{1}{\theta}(p_{ii}\tau_{ij} + d_{ij}) = \frac{1}{\theta}(\frac{\gamma}{\theta}\tau_{ij} + d_{ij})$$

## B.4 Hub formations driven by the zero profit condition in transportation

***Consumption Ratios*** $\triangleright$ It will be useful at this point to define consumption ratios as viewed by the exporting firm in order to express exports across all countries in common units. Define hence the ratio of consumption for exports to country $j$ relative to domestic consumption (which is identical in all countries due to similar technology):

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

For simplicity assume that other trade costs $\tau_{ij} = \tau$ are symmetric and the distortion is only created by the asymmetry $d_{13} = d_{23} = d < d_{12} = d'$. Consumption ratios that exporting firms of country 3 have to face are:

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

For countries 1 and 2 equivalently we have for trading between them:

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

and for trading with country 3 being the most proximal to both:

$$\frac{q_{13}}{q_{11}} = \frac{q_{23}}{q_{22}} = \left(\frac{\frac{\gamma}{\theta}\tau + d}{\gamma}\right)^{\frac{1}{\theta-1}}$$

Prior to deriving the result, one last normalisation is made since the symmetry of the iceberg trade cost $\tau$ and the common marginal cost $\gamma$ are identical across countries. Hence, impose $\tau = \gamma = 1$.

***Country 3, Zero Profit Condition in transportation*** $\triangleright$ Free entry and exit of transport firms results in zero long term profits satisfying the P=AC condition:

$$f_{31} = f_{32} = d + F_3^t(q_{31} + q_{32})^{-1} \Longrightarrow q_{31} = \frac{1}{2}\frac{F_3^t}{\frac{1}{\theta} + d}\frac{\theta}{1-\theta} \tag{7}$$

It is straightforward to see that due to symmetry, the total exports of country 3 are split equally between countries 1 and 2.

***Country 1, Zero Profit Condition in transportation*** $\triangleright$ Free entry and exit of transport firms results in zero long term profits:

$$f_{12}q_{12} + f_{13}q_{13} = d'q_{12} + dq_{13} + F_1^t$$

Using the consumption ratios we can express $q_{12}$ in units of $q_{13}$ and replacing the transportation price. The relationship can be rearranged to write:

$$q_{13} = q_{31} = \frac{F_1^t}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}}\frac{\theta}{1-\theta} \tag{8}$$

Exports from country 1 (and 2 by symmetry) are clearly less than what country 3 can achieve due to its beneficial location.

***Hub formations*** $\triangleright$ The left hand sides of equations 7 and 8 are necessarily the same as it is the expression of the common unit of exports. We have assumed that the overhead costs of transport $F_i^t$ in any country can vary. Equating the two terms then yields a ratio of the fixed cost of transport in the two countries:

$$\frac{F_3^t}{F_1^t} = \frac{2\left(\frac{1}{\theta} + d\right)}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \tag{9}$$

The ratio of fixed costs of transport and the assumption of their variability are crucial in identifying the type of formation between the trading countries. The term 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.* The term (9) has $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} > 0$:

$$\begin{aligned}
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= -\frac{\left[2\left(\frac{1}{\theta}+d\right)\right]\frac{\theta}{\theta-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}-1}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}{\left[\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\right]^2} \iff \\
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= \frac{F_3^t}{F_1^t}\frac{\frac{\theta}{1-\theta}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}-1}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}} > 0
\end{aligned} \tag{10}$$

Since both fractions are positive.

(9) has also $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} < 0$:

$$\begin{aligned}
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{2}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}} \\
&- \frac{\left[2\left(\frac{1}{\theta}+d\right)\right]\left(1+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\left[\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\right]^2} \iff \\
&\text{multiplying and dividing the first term with } \left(\tfrac{1}{\theta}+d\right) \text{ yields} \\
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left[\left(\frac{1}{\theta}+d\right)^{-1} - \frac{\left(1+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}\right] \iff \\
\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left(\frac{1}{\theta}+d\right)^{-1}\left[1-\frac{\left(\frac{1}{\theta}+d+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}}\right] < 0
\end{aligned} \tag{11}$$

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

These properties allow the following statements:

1. *If country 1 decides to trade using country 3 as a hub, it will have to increase its trading distance to $d'' = 2d > d'$. This will necessitate an increase in the ratio of*

*the fixed costs of transport between the hub country 3 and country 1. Hence there will either need to be an increase in the fixed costs of transport of the hub country or a decrease in the fixed costs of transport for the connecting country or any increasing combination of both.*

2. *If country 3, the hub country, is ever more distant from the connecting country 1, the ratio of fixed costs needs to be decreased. This implies either a decrease in the fixed costs of the hub country or an increase in the fixed costs of the connecting country or any decreasing combination of both.*

A change in country 1's trading decisions will however not enforce a change in country 3's level of fixed costs as the profit functions of transport firms are independent of each other. Hence all the changes in the ratio are driven by adjustments in the fixed costs of transport for the connecting country 1. By symmetry of the distance $d'$ the same observations hold for country 2.

The above two statements are equivalent with the operation of replacing the unmeasurable benefit of forming a connection with a change in the fixed costs associated with the network, subsequently leading to the empirical prediction. By assuming existence of increasing returns to scale in the transportation sector and varying fixed costs in transportation, the benefit of forming a connection can be represented by changes in fixed costs stemming from the decision of the firm to trade directly or indirectly. This alternative setting confirms qualitatively the theoretical exposition where the cumbersome approach of assuming the presence of a transportation sector can be avoided by simply assuming the existence of benefits and costs associated with connections.

### B.5 The number of transport firms

For completeness, we close the model by characterising the number of transport firms. For country 3, the partial equilibrium price and output can then be utilised to extract the number of firms. The former satisfies the full employment condition and is shown to be:

$$n_3^t = \frac{L^t}{F_3^t}(1-\theta)\frac{d\theta+1}{d\theta+1-\theta}$$

For country 1, expressing output in common units of $q_{31}$ we have:

$$n_1^t = \frac{L^t}{F_1^t}(1-\theta) \times \left[(1-\theta) + K\frac{1}{2}\frac{F_3^t}{F_1^t}\right]^{-1}$$

where $K(d,d')$[20] is a function of the distances between trading partners. The term $K$ is increasing in $d$ and decreasing in $d'$. The number of transport firms as shown below is decreasing in $\delta$ which is what one should expect since by virtue of the second statement more labour is required to be allocated to accommodate an increase in $F_1^t$ and an increase in $d$. The change in the number of transport firms is ambiguous wrt to changes in $d'$. It will be determined by the level of the ratio of fixed costs. If the level of fixed costs is substantially large implying the level of fixed costs of the connecting country is small then the number of transport firms is decreasing in $d'$. If the level of fixed costs of the connecting country is large then the ratio becomes small implying an increase in the number of transport firms. This arises because there is an increase in labour input due to the increase occurring in $d'$ and a decrease in labour input as a result of a reduction of fixed costs of the connecting country. Hence the number of firms will depend on the level of the ratio of fixed costs.

*Proof.* $\frac{\partial n_1^t}{\partial d} = -\frac{L^t}{F_1^t}(1-\theta)\left[(1-\theta)+K\frac{1}{2}\frac{F_3^t}{F_1^t}\right]^{-2}\frac{1}{2}\left(\frac{\partial K}{\partial d}\frac{F_3^t}{F_1^t} + K\frac{\partial\frac{F_3^t}{F_1^t}}{\partial d}\right)$

The last term in brackets can be rearranged to write:

$M = \theta\frac{F_3^t}{F_1^t}\left[\frac{\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{\theta}{\theta-1}}} + \frac{(d+d')\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{1-2\theta}{1-\theta}}}\left(1+\frac{1}{1-\theta}\times\Lambda\right)\right] > 0$

where $\Lambda = 1 - \frac{(1-\theta)(\frac{1}{\theta}+d)+\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}} > 0$ since the fraction is clearly less than unity.

$$\frac{\partial n_1^t}{\partial d'} = -\frac{L^t}{F_1^t}(1-\theta)\left[(1-\theta)+K\frac{1}{2}\frac{F_3^t}{F_1^t}\right]^{-2}\frac{1}{2}\left(\frac{\partial K}{\partial d'}\frac{F_3^t}{F_1^t} + K\frac{\partial\frac{F_3^t}{F_1^t}}{\partial d'}\right)$$

The last term in brackets can be rewritten as:

---

[20] $K = \frac{\theta(d+d')\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{\theta}{\theta-1}}}$, $\quad \frac{\partial K}{\partial d} = \theta\left[\frac{\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{\theta}{\theta-1}}} + \frac{1}{1-\theta}\frac{(d+d')\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{1-2\theta}{1-\theta}}}\right] > 0$,
$\frac{\partial K}{\partial d'} = \frac{\theta\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{\theta}{\theta-1}}}\left[1 - \frac{1}{1-\theta}(d+d')\left(\frac{1}{\theta}+d'\right)^{\frac{1+\theta}{1-\theta}}\right] < 0$

$$M' = \frac{\theta\left(\frac{1}{\theta}+d'\right)^{\frac{1}{\theta-1}}}{\left(\frac{1}{\theta}+d\right)^{\frac{\theta}{\theta-1}}}\left[1-(d+d')\left(\frac{1}{1-\theta}\left(\frac{1}{\theta}+d'\right)^{\frac{1+\theta}{1-\theta}}-\frac{\partial\frac{F_3^t}{F_1^t}}{\partial d'}\right)\right]$$

Expanding the partial derivative of fixed costs wrt $d'$ and grouping terms the last term in the brackets can be expressed as:

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

where the magnitude of the ratio of fixed costs will determine whether the term in brackets is positive or negative since all other terms are less than unity.

The ratio of fixed costs is greater than unity since $d' > d$:

$\frac{F_3^t}{F_1^t} = 2\frac{\left(\frac{1}{\theta}+d\right)}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}} \geq 1 \Leftrightarrow \frac{1}{\theta}+d \geq \left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}} \Leftrightarrow$ $\left(\frac{1}{\theta}+d\right)^{-\frac{\theta}{1-\theta}} \geq \left(\frac{1}{\theta}+d'\right)^{-\frac{\theta}{1-\theta}}$ which can only hold if $d' \geq d$, which is true.

Therefore if the magnitude is such that $\Lambda' < 0$ then $M' > 0$ and hence $\frac{\partial n_1^t}{\partial d'} < 0$: the number of transport firms are decreasing in $\delta'$. If the magnitude of the ratio of fixed costs is such that $\Lambda' > 0$ then $M' < 0$ and the number of transport firms are increasing in $d'$. □

# C Robustness Checks

This section addresses concerns regarding sample selection associated with the selection of the particular hub locations. It does so by replicating outcomes of section 7 for a dataset that has been constructed by utilising an alternative distance matrix. All other control variables and the time period of estimation remain unchanged.

## C.1 Use of an alternative distance matrix

A grid[21] containing all possible distances from/to country capitals to/from their principal ports and intervening port to port distance is compiled. An indicator is activated if a particular route passes within a few kilometres to every possible location that is a canal, passage or a transshipment hub. We exclude landlocked exporting/importing countries from the sample. These locations are illustrated in figure 6. 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 thickness at the HS 2 and 6 levels, the distance matrix reveals that 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 6 suggests that the coefficients of indirect capital distance are lower than those of direct capital distance and that this difference is significant in mose cases of the HS 2 and 6 levels. The interaction term, when introduced has a negative sign opposite of that found in section 7. This combination of findings leads to either a reduced or negative discount from utilising a hub location that is at least geographically advantageous, relative to trading directly and seems to contradict the main findings of section 7.

[21] I would like to thank Michael Traut for providing these observations which are 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.

Figure 6: Hub Locations, (alternative distance matrix)

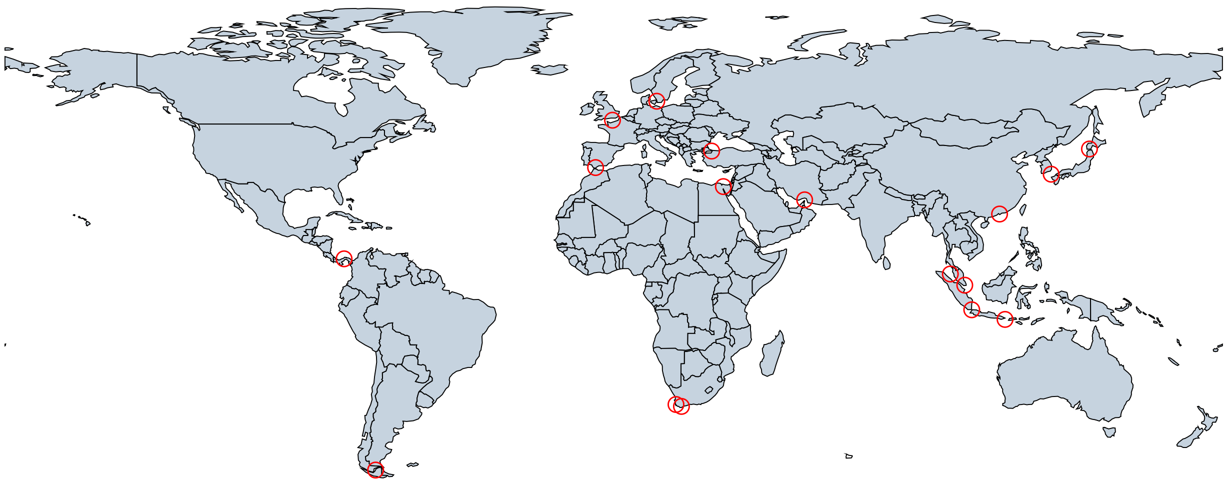


Table 6: Exports: Coefficients of direct distance and indirect distance.

| | (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** | **Raw** | | | | | | **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 Online Appendix.

In order to address the routing problem, we repeat the two stage least squares estimation using time difference for the exclusion restriction to hold and recover the true impact of frictions associated with remoteness and transfer of information. The findings in table 8 are supportive of those in section 7 as in most cases the difference between the two distance

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** | | | **Raw** | | | | | | **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 Online Appendix.

variables becomes indistinguishable with exceptions being columns 11 and 12 where the null hypothesis is rejected under the t-test but not under the Paternoster test.

Finally we depart from the specification of the theoretical model by checking the marginal effect of distance when none or a finite number of hub locations intervene between a country pair. Table 9 reveals that pairs which do not have a hub location intervening have a lower impact of distance on exports relative to the complement. When both the number of hubs and the interaction are introduced, either of the two variables is significant in a few cases at the HS 2 level and carry a positive sign. At the HS 6 level both variables are jointly significant and the interaction term is positive but any additional hub reduces exports.

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 | -1.523*** | -1.551*** | | | | | -1.312*** | -1.341*** | | | | |
| $R^2_{Direct}$ | 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 | -1.219*** | -1.234*** | | | | | -1.036*** | -1.050*** | | | | |
| $R^2_{Indirect}$ | 0.869 | 0.887 | | | | | 0.838 | 0.866 | | | | |
| Underidentification Test | 36.20 | 34.54 | | | | | 32.96 | 31.26 | | | | |
| Cragg-Donald | 1857 | 1596 | | | | | 1882 | 1576 | | | | |
| p-value Ho A | 0.192 | 0.218 | | | | | 0.258 | 0.283 | | | | |
| p-value Ho A (P. T.) | 0.521 | 0.550 | | | | | 0.576 | 0.600 | | | | |
| **HS 2 Level** | | | | | | | | | | | | |
| Direct Capital Distance | -0.844*** | -0.876*** | -1.095*** | -1.131*** | -1.821*** | -2.063*** | -0.691** | -0.750** | -0.828** | -0.890** | -1.461*** | -1.718** |
| $R^2_{Direct}$ | 0.366 | 0.377 | 0.499 | 0.511 | 0.760 | 0.812 | 0.261 | 0.278 | 0.412 | 0.431 | 0.695 | 0.779 |
| Underidentification Test | 9.693 | 9.782 | 9.565 | 9.659 | 8.732 | 9.752 | 9.350 | 9.272 | 9.251 | 9.179 | 8.397 | 9.208 |
| Cragg-Donald | 18462 | 17537 | 18267 | 17359 | 15822 | 9685 | 16450 | 15507 | 16351 | 15425 | 14527 | 9312 |
| Indirect Capital Distance | -0.598*** | -0.610*** | -0.775*** | -0.786*** | -1.244*** | -1.379*** | -0.486*** | -0.521*** | -0.581*** | -0.617*** | -0.989*** | -1.147*** |
| $R^2_{Indirect}$ | 0.361 | 0.372 | 0.493 | 0.505 | 0.754 | 0.811 | 0.257 | 0.273 | 0.407 | 0.425 | 0.688 | 0.774 |
| Underidentification Test | 21.13 | 21.65 | 20.99 | 21.52 | 20.35 | 22.82 | 20.18 | 20.11 | 20.07 | 20 | 19.15 | 20.62 |
| Cragg-Donald | 41233 | 39839 | 40949 | 39573 | 37541 | 27933 | 38820 | 36877 | 38686 | 36761 | 36361 | 25789 |
| p-value Ho A | 0.246 | 0.228 | 0.144 | 0.132 | 0.0290 | 0.0294 | 0.263 | 0.242 | 0.212 | 0.195 | 0.0590 | 0.0607 |
| p-value Ho A (P. T.) | 0.554 | 0.541 | 0.522 | 0.509 | 0.477 | 0.492 | 0.612 | 0.598 | 0.609 | 0.597 | 0.543 | 0.553 |
| **HS 6 Level** | | | | | | | | | | | | |
| Direct Capital Distance | -0.0268 | -0.0944 | -0.308 | -0.367 | -1.368*** | -1.781*** | 0.0402 | -0.0144 | -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 | 10.02 | 9.266 | 9.926 | 9.166 | 11.93 | 17.70 | 8.412 | 7.799 | 8.153 | 7.546 | 7.812 | 12.13 |
| Cragg-Donald | 140632 | 131738 | 136264 | 127487 | 102803 | 38102 | 104723 | 96411 | 99794 | 91793 | 64170 | 19243 |
| Indirect Capital Distance | -0.0212 | -0.0736 | -0.242 | -0.285 | -1.043*** | -1.408*** | 0.0302 | -0.0106 | -0.134 | -0.164 | -0.730*** | -1.025*** |
| $R^2_{Indirect}$ | 0.211 | 0.220 | 0.360 | 0.368 | 0.693 | 0.816 | 0.100 | 0.114 | 0.301 | 0.312 | 0.644 | 0.803 |
| Underidentification Test | 22.29 | 21.38 | 22.16 | 21.24 | 27.72 | 38.74 | 20.26 | 19.17 | 20.08 | 18.94 | 23.31 | 33.96 |
| Cragg-Donald | 264046 | 245998 | 256136 | 238625 | 182084 | 85825 | 179777 | 164446 | 172550 | 157817 | 111100 | 49095 |
| p-value Ho A | 0.982 | 0.937 | 0.779 | 0.737 | 0.185 | 0.241 | 0.956 | 0.984 | 0.753 | 0.693 | 0.0595 | 0.0797 |
| p-value Ho A (P. T.) | 0.990 | 0.964 | 0.862 | 0.834 | 0.498 | 0.544 | 0.977 | 0.992 | 0.860 | 0.826 | 0.478 | 0.477 |
| **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** | | | **Raw** | | | | | **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 Online Appendix.

Table 9: Number of hubs.

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **Total Trade** | | | | | | |
| | Indirect Capital Distance, #Hubs=0 | -1.408*** | -1.376*** | | | | | -1.139*** | -1.152*** | | | | |
| | $R^2$ | 0.880 | 0.913 | | | | | 0.880 | 0.914 | | | | |
| A | Indirect Capital Distance,#Hubs>0 | -1.855*** | -1.830*** | | | | | -1.729*** | -1.810*** | | | | |
| A | Interaction | 0.278 | 0.188 | | | | | -0.0129 | -0.0127 | | | | |
| A | Number of Hubs | -0.0452 | 0.0118 | | | | | 0.0592 | 0.0704 | | | | |
| A | $R^2_A$ | 0.880 | 0.902 | | | | | 0.833 | 0.865 | | | | |
| B | Indirect Capital Distance,#Hubs>0 | -1.795*** | -1.845*** | | | | | -1.804*** | -1.901*** | | | | |
| B | Interaction | 0.211* | 0.206 | | | | | 0.0760 | 0.0953 | | | | |
| B | $R^2_B$ | 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** | | | | | | |
| | 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*** |
| A | 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 |
| A | 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 |
| A | $R^2_A$ | 0.372 | 0.384 | 0.506 | 0.520 | 0.775 | 0.847 | 0.234 | 0.253 | 0.395 | 0.416 | 0.697 | 0.814 |
| B | 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*** |
| B | 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 |
| B | $R^2_B$ | 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** | | | | | | |
| | 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 |
| A | 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*** |
| A | 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 |
| A | 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 |
| A | $R^2_A$ | 0.214 | 0.224 | 0.366 | 0.375 | 0.728 | 0.871 | 0.088 | 0.104 | 0.310 | 0.323 | 0.667 | 0.853 |
| B | 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*** |
| B | 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 |
| B | $R^2_B$ | 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** | | | **Raw** | | | | | | **No USA-EU15 Bilateral Trade** | | | |

Reported coefficients 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. *A* indicates a set of regressions including indirect capital distance and the interaction with the number of hubs on a route as independent variables. *B* indicates a set of regressions s 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 Online Appendix.