Endogenous hub formations in international trade*

Nicholas-Joseph Lazarou†

University of Southampton

Abstract

Paul Krugman in 1993 explored the role of transportation hubs in the pattern of trade. He asked when and why will a transportation hub emerge. In this paper, I incorporate network theory in a simple trade model of monopolistic competition with representative firms in order to yield conditions governing the optimal network formation - a hub or a direct links structure. Firms take into consideration the optimal network for their trading strategy. Reductions in variable and fixed costs affect trade flows in the predictive way. But trade volumes also have an impact on the optimal transportation network structure because of increasing returns in the transport sector, herein modelled via the manufacturing firm.

I provide empirical evidence for the existence of hub formations and direct connections in international trade flows by employing HS2 bilateral trade flows for a panel of 63 countries between 2003 and 2007. I compare two measures of distance, direct capital and indirect capital distance, under two different regimes: countries trading directly, and countries trading indirectly. Indirect distance under hub networks reduces exports by about 5 per cent less relative to direct distance. The latter reduces export flows 24 per cent less than indirect distance when direct trading is allowed. I show that the network formation is driven through sectoral heterogeneity. The decision of the firm to choose network formations results in the endogeneity of distance. Consequently sector fixed effects are utilised or time difference as an instrument to identify the impact of the network structure in trade flows.

Preliminary Draft, August 2012. - Please do not cite without prior permission from the author. First version September 2011.

Keywords: International Trade, Network Formations, Transport Costs.

JEL Classification Numbers: F12, L91, C71.

*I am grateful to my Supervisors Dr. Hector Calvo-Pardo and Dr. Jackline Wahba for their faithful support and advice. I thank Dr. Luca David Opromolla for his constructive suggestions.

†Ph.D. Student in Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom. Email: N.Lazarou@soton.ac.uk

1 Introduction

Krugman (1993) explored the role of transportation hubs in the pattern of interregional and international trade. He tackled a problem that is widely ignored in international trade theory: when and why will a transportation hub emerge. He showed that production of commodities will be concentrated in a location from which all arrivals and departures have the lowest transportation costs. The interaction between increasing returns to scale in manufacturing and transport leads to the emergence of a hub region.

Transportation hubs are not a fundamental element of international trade study because not enough attention is cast upon the structure of the transport industry. For transportation costs are just perceived as a subset of trade costs and specifically variable trade costs. They are modelled using the iceberg assumption first developed 60 years ago. This simple modelling concept allows for a wedge that serves to inhibit allocation of production and consumption.

In this paper I perturb Krugman's model of transportation hubs in order to raise awareness of another dimension of the distortive impact of transportation costs on trade flows: That of the network organisation of the transport sector whereby its pervasive effects on trade flows are studied. Specifically I show that when transportation costs are high, manufacturing firms instead of lowering their output or exiting the market, can choose to trade via a central hub provided there is a cost saving incentive, or benefit, of connecting indirectly to their trading partner. When transportation costs are low for a given benefit, these firms will always trade directly with foreign countries or otherwise face losses. Empirically the existence of hub formations using sectoral aggregate export flows is confirmed and supporting evidence for such structures is then presented: Manufacturing firms tend to trade indirectly via a central hub as well as directly. Firms producing agricultural goods only to trade directly to their destinations. But network selection is defined endogenously by the firm as a form of a hedge against adverse transport costs. Such a decision induces simultaneous causality between export flows, distance and fixed costs and is attributed to the network effect. Appropriate instrumentation to identify the effect created by the network is then introduced and the true magnitude of trade costs on trade flows is recovered whilst controlling for network induced information.

According to Hummels (2007), 99% of the world's trade by weight is carried by sea. The combination of scale economies in maritime transport and in port infrastructure made possible by the advent of containerisation induced the creation of hub and spoke configurations[1] whereby export distribution is primarily defined by the cost saving size of the vessel and not necessarily via the closest distance. Hence, larger ships operate amongst the denser trade routes, whilst smaller ships deliver the quantity demanded to and from the major in-

[1] Henceforth I will be referring to hub and spoke configurations whenever I invoke the term "hub formation".

tersection ports (Hummels 2007). This organisation of trade distribution is a simple form of a network. In line with Hummels who yields this inference for the maritime transportation sector, Hendricks et al. (1995) are of the opinion that ingredients such as exercise of market power together with the freedom of setting prices and routes, leads also airlines to setting their optimal network formation. The optimal formation becomes the one which minimises total costs constrained by import demand.

The hub ports that emerged as a product of containerisation were not necessarily associated with facilitating supply of exports to cater for domestic import demand, but rather could be outcomes of geographical advantage and concentration of production that facilitate transit towards a final destination (Krugman 1993). Hence historical incidence, interregional trade and globalisation all play a role in the development of these formations leading to hub ports achieving large sizes as suggested in Levinson (2008). Table 1 presents the major containerised ports of the world arranged by the number of containers handled in the last year of measurement. An interesting finding is that the total number of containers handled by these ports alone in 1991 is approximately equal to the global containerised trade volume, whilst for 2008 it is 54 per cent higher than the volume of global containerised trade standing at 137 million TEU's. This converts to about 1.3 billion tons of traded goods which is nearly 25 per cent of the world's non-liquid traded goods. Assuming that the weight/TEU ratio remains constant then total throughput of these ports alone can be calculated to 35 per cent of the global trade in dry goods.

A hub formation is a particular form of a network which, in this study, is incorporated in the standard trade model of monopolistic competition with representative firms. I abstain from modelling a separate transportation sector[2] but keep the network structure as the product of the transportation sector's industrial organisation: The theoretical model presented herein is qualitatively equivalent of modelling a transportation sector operating under increasing returns. Under the assumption that fixed costs associated with transportation can vary, then hub formations are an outcome of increasing returns in transportation.[3] Then the impact of the network on trade flows through decisions of manufacturing firms can very simply studied.

In this framework the notion of iceberg costs is substituted with benefits and costs associated with network formations. This is done by merging the symmetric connections framework of Jackson & Wolinsky (1996) in a simple Krugman monopolistic competition model of trade for three countries. Using sufficiently less assumptions than Krugman (1993), firms first decide whether to export or not. If they do, I study which conditions will determine their organisation under a hub formation or direct links. The reason for this approach is to have an alternative methodological tool than the Krugman (1993) model. In the latter

[2] I am concurrently extending this model to cater for: a) network structure, b) increasing returns in the transportation sector and c) firm heterogeneity in manufacturing.

[3] See also Appendix B for a proof of this claim.

Table 1: The World's Largest Containerports, by container throughput.

Port	1990	2008
Singapore	5.21	29.91
Shanghai	0.5	27.99
Hong Kong	5.1	24.49
Shenzen	0	21.41
Busan	2.3	13.45
Dubai	1.1	11.82
Rotterdam	3.7	10.8
Qindgao	0.1	10.32
Hamburg	2	9.73
Kaohsiung	3.5	9.67
Antwerp	1.6	8.66
Port Klang	0.5	7.97
Los Angeles	2.6	7.85
Long Beach	1.6	6.48
Tanjung Pelepas	0	5.6
New York	1.9	5.26

Source: Levinson (2008) which has been augmented to include the latest figures based on the UNCTAD Review of Maritime Transport, 2011. Numbers are in Millions of Twenty Equivalent Units (TEU).

one needs to postulate concentration of production, centrality together with a defecting firm as sufficient conditions to yield equilibria. Further, it is not possible to develop a testable prediction for the existence of hub formations.

In this simple extension of the Krugman (1993) model of trade, locational advantage and the level of transportation costs are the only prerequisites after controlling for country size. The economic intuition behind this framework is related to the very nature of maritime trade. Whilst concentration of production jointly with centrality may be of importance under air transport, the same need not apply for maritime transport. Some of the less developed regions of the world, with examples present in Table 1, obtain hub status conditional only on centrality amongst the distribution of trade routes. Concentration of production may then take place but is certainly not a condition precedent.

Empirically I show, under different distance and fixed cost estimations, when hub networks appear to be more beneficial for trading partners than direct trading. When allowing for indirect trading, distances and fixed costs associated with hub formations appear to affect exporters less. Distances and fixed costs associated with direct trading appear to affect exporters more. The reverse holds true when direct trading is permitted. This application holds for manufacturing firms, but is violated for firms producing agricultural goods. The aforementioned estimators of distance take into consideration the network effect and are proved to be endogenous. This is because the firm always decides the formation that will allow it to increase its trade volume. I show that this effect is driven by sectoral heterogeneity and appropriate controls are subsequently applied. Another way of identifying the effect is by use of instrumentation: time difference becomes an instrument for distance. This assists in recovering the true impact of distance costs on trade flows.

The implications of this paper are directed towards developing countries. Exporters there have to absorb higher transport costs so as to be in the position to penetrate markets abroad. This situation prevents export-led development, reducing workers' wages and inducing a welfare impact. Higher transport costs are attributed to geographical 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 & Venables 2004, p.77). Therefore improvement of own and transit country infrastructure together with hub formations could make possible the amelioration of excessive transport costs (Limao & Venables 2001, p.471).

This study balances between two strands of theoretical and empirical trade literature. It also incorporates a branch of the separate field of social and economic networks. Predictive inference is undertaken through the standard trade model of representative firms developed in Krugman (1980) and Krugman (1991) by distorting the iceberg assumption. This is done after examining more closely the nature of transportation costs and the industrial organisation of the transport sector. The importance of transportation costs and maritime costs has been highlighted in the works of Hummels (2007) and Hummels et al. (2009). The relative importance of the validity of the iceberg cost assumption for transportation costs is contested in Irarrazabal et al. (2011) and Hummels & Skiba (2004). Insight about the structure of the transportation industry is provided in Hummels et al. (2009) and Hendricks et al. (1995) whilst the pervasive impacts of the modes of transport on trade flows are presented in Lux (2011). The importance of incorporating the network literature in international trade is currently being highlighted by Chaney (2011).

The remainder of this paper is organised as follows: Section 2 provides some basic notions of economic and social networks that will be used in the model. Section 3 incorporates network structure in a simple trade model of monopolistic competition. Equilibria for two countries and three countries are studied in section 4. Subsequently the conditions for endogenous hub formations are developed. Section 5 derives the gravity equation for aggregate trade flows and yields the empirical prediction for the existence of hub formations. Calibration of and endogeneity in the gravity equation are discussed in section 6. The evidence of hub network formations are subsequently presented and treatment for the endogeneity of distance is applied. 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* or *directly linked*, 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 linked to each other. Each link 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 link* will be denoted as $g_{ij} = 1$, and $g_{ij} = 0$ if countries i and j are *not directly connected*.

Each *link* is associated with costs and benefits. These costs and benefits affect firms that choose to enter the export market in each particular country. Two types of links can exist, a *direct* and an *indirect* link. If a direct link is formed by a firm then it must incur a cost c. There is also a benefit $0 \leq \delta \leq 1$ associated with proximity or distance between i and j in the sense that the firm will prefer to trade to closer trading partners rather than more distant ones. The firm has the additional option to form an indirect link. Implicitly, there must already exist a direct link to another country for the indirect link to be feasible. The indirect link is formed without cost and receives only a pure benefit $\delta^{t_{ij}}$, where $t_{ij} \geq 0$ is the integer number of links between countries i,j. This construction allows a firm in country i entering the export market to take into consideration the distance δ and the cost c associated with this distance. 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 proximity decreases. The difference between the benefit of forming a link and the cost of a type of link is thus defined as:

$$v_{ij} = \delta^{t_{ij}} - c_{ij|ij \in G} \tag{1}$$

By convention we have $g_{ii} = 0 \Rightarrow c_{ii} = 0$ since $g_{ii} = 0$ is not a link 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 as follows. 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 links between all participants. Then the network is defined as $G = \{12, 21, 13, 31, 23, 32\}$. The second formation is an indirect link between 1 and 2 and direct links 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 links $v_{12} = \delta - c = v_{21}$. For the case of an indirect link between 1 and 2 we have: $v_{12} = \delta^2 - 0 = v_{21}$. The latter indirect link implies the existence of two direct links forming this particular connection: the link between 1 and 3 and the link between 3 and 2.

Countries active in the network are also characterised by their participation share in the network depending on the types of links 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 in lieu of the unmeasurable benefit of forming an indirect link. In the empirical part the parameter δ will be substituted with the fixed costs associated with the network for each country. 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 $F_i = \frac{n_i(G)}{n(G)}$. To provide an example, consider the direct links 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 links between 1 and 2 we have $F_1 = 1/4$, since country 3 in the middle is burdened by the additional share $F_3 = 1/2$.

For the remainder of this paper, countries are symmetrically spaced: $ij = ji$ therefore $c_{ij} = c_{ji} = c, t_{ij} = t_{ji} = t$. This network structure will now be incorporated in a simple trade model of monopolistic competition with representative firms. The general equilibria between two trading countries and three trading countries are characterised and it is shown how the impact of the network formation on trade flows for three trading countries becomes an endogenous decision for the firm in each country.

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

$$U = q_0^{1-\mu} \left[\int_{\omega \in \Omega} q_{ij}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}\mu} \tag{2}$$

where $\sigma > 1$ is the elasticity of substitution between pairs of varieties and Ω 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:

$$q_{ij} = \frac{\mu L_j p_{ij}^{-\sigma}}{\sum_{j,i=1}^{N} \int_{\Omega} p_{ij}^{1-\sigma}(\omega) d\omega} \tag{3}$$

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

Production and Trade Costs — Good 0 is the numeraire homogeneous good. One unit of labour produces w and 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. The fixed cost for the remainder of the theoretical model will be constant. In the aggregate trade (gravity) equation this assumption will be relaxed by incorporating the share of each country's participation in the network. 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 link to another country. In this way I introduce network structure in the trade model. For domestic consumption the net benefit becomes by construction equal to the value of unity.

To produce and sell a variety ω either domestically or abroad, the firm in country i employs Labour input:

$$L(q) = \gamma \frac{q_{ij}}{v_{ij}} + F_i = \gamma \frac{q_{ij}}{\delta^{t_{ij}} - c_{ij|ij \in G}} + F_i \tag{4}$$

As such, if a direct link ij is formed the firm receives a benefit δ in the sense that $\frac{\partial \Pi}{\partial \delta} > 0$. In addition it incurs a transport cost c as $\frac{\partial \Pi}{\partial c} < 0$. In case of an indirect link, 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 require firms to enter the sector exhausting any profit margin. When profits are zero, 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 ω produced in each country, the total labour input must equal the labour share in the increasing returns sector: $\int_{\omega \in \Omega} L_{q,i}(\omega) d\omega = \mu L_i$. Since each firm produces one variety, the number of firms becomes finite and equal to $n_i = \frac{\mu L_i}{\sigma F_i}$. 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 Endogenous Formations

In this section I characterise equilibria based on network structure, symmetric geographical placement of countries, optimal prices given trade costs and traded quantities. Equilibria for a two country case and a three country case are considered. The two country case does not differ from the Krugman (1980) trade model of monopolistic competition. It yields the home market effect including also a network. In order to proceed to a higher number of trading countries a strong assumption must be introduced: that the home market effect is normalised to 1. In this respect, all countries and number of firms become identical. Under this condition firms choose their network formation for pre-determined values of costs given a constant benefit, and unique equilibrium network formations will arise. Otherwise firms will need to exit the export market due to negative profits. There exist also borderline cases when firms will become indifferent between formations or indifferent between remaining autarkic or choosing a particular network. Other network-specific notions such as stability and efficiency are relegated to the appendix since they are not intrinsic to the inferential properties of trade costs on trade flows.

Two Country Equilibrium — 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 links 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:

$$\sum_{1,j=1}^{2} q_{1j}(p_{1j} - \frac{\gamma}{v_{1j}}) = \sum_{2,j=1}^{2} q_{2j}(p_{2j} - \frac{\gamma}{v_{2j}}) \Rightarrow Q_1 = Q_2 \tag{5}$$

Domestic prices are equal across countries as γ 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:

$$\frac{n_1}{n_2} = \frac{\frac{L_1}{L_2} - (\delta - c)^{\sigma-1}}{1 - \frac{L_1}{L_2}(\delta - c)^{\sigma-1}} > 0 \tag{6}$$

Introducing a network leaves things unchanged in the standard model of trade with two countries. 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 link.

The link 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. The equilibrium outcome is 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 link. The condition holds as long as $(\delta - c)^{\sigma-1} < \frac{L_1}{L_2} < (\delta - c)^{1-\sigma}$ and n_1 and n_2 are non-zero.

Three Country Equilibrium — Similarly to Krugman (1993), in order to proceed to the three country example a strong simplifying assumption that will enables construction of the equilibrium is introduced: That is, all countries 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 this implies that the home market effect between any two trading partners in a three country world is normalised to 1. In relationship (6) this would be equivalent to stating that the ratio $L_1/L_2 = 1$. Then it must be that $n_1/n_2 = 1$.

The impact of the network structure becomes apparent when a firm has to consider whether it will form a direct link or an indirect link to a trading partner. In order to form an indirect link with one partner it must have implicitly formed a direct link with another. Exploiting the symmetry assumption any decision that a firm in country 1 may make is an equivalent decision for a firm that could be in the complement countries. Therefore I 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 links formed. The equilibrium condition is expressed for any country pair as:

$$\sum_{1,j=1}^{3} \frac{q_{1j}}{v_{1j}} = \sum_{2,j=1}^{3} \frac{q_{2j}}{v_{2j}} = \sum_{3,j=1}^{3} \frac{q_{3j}}{v_{3j}} \Rightarrow Q_1 = Q_2 = Q_3 \tag{7}$$

While this equilibrium may not be of particular interest, it helps to extract inference between the 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 links.

Direct Links Network — The net benefit term becomes $v_{ij} = \delta - c$ for all pairs $\{i, j\}$ in the network. The zero profit condition and the assumption of no home market effect naturally 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 links established. For country 2 for example it can be expressed as:

$$P_2 = n \left(\frac{\sigma}{\sigma - 1} \gamma \right)^{1-\sigma} \left(1 + 2(\delta - c)^{\sigma-1} \right) \tag{8}$$

Indirect Links Network — In the case of indirect links between countries 1 and 2 and direct links with country 3, the same equilibrium condition must hold, $Q_1 = Q_2 = Q_3$. The price index with one indirect connection and one direct is written:

$$P_2 = n\left(\frac{\sigma}{\sigma-1}\gamma\right)^{1-\sigma}\left(1+(\delta-c)^{\sigma-1}+(\delta^2)^{\sigma-1}\right) \tag{9}$$

where the term δ^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 δ 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 δ that admits the equilibrium condition:

$$\delta - c = \delta^2 \tag{10}$$

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 δ the values of transportation costs admit an equilibrium where only direct or indirect links 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 links network):

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

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 link):

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

The endogenous decision of the firm to change network formation arises by minimising firm losses given a constant benefit δ and a variable cost c. The indirect network formation will prevail if the cost from forming a direct link is very high. Then the firm may decide to sever the direct link 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. By setting the equilibrium condition to $0 = \pi_1^I > \pi_1^D$ we can test when will the indirect links network formation prevail. Solving this inequality yields the simple relationship $\delta - c < \delta^2$. Then denote as $c = \bar{c}$ the infimum of high transport 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 will be the indirect 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 minimises each country's exposure to transport costs: the indirect links network has the minimal number of links that connects all countries. The average path length is minimised given the minimal number of links.

Alternatively, when there is a very low transportation cost $c < \bar{c}$, the direct network formation will prevail and a firm will suffer losses if it is trading indirectly. Set then the equilibrium condition to $0 = \pi_1^D > \pi_1^I$ to test when will the direct links network formation arises. 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 link costs are low it becomes beneficial to form all direct links. The cost of adding a link is less than the benefit the firm gains from shortening the link of length two (δ^2) into a link of length one.

When the transportation cost is extremely high, none of these formations should prevail. The countries become autarkic. The autarkic equilibrium requires that trading 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 will take the at least 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 & Wolinsky (1996). I summarise them as follows. For unique values of c in the set $\mathbf{C}$ and holding constant the benefit term δ, the network formation decisions for a representative firm in the symmetric trade model with increasing returns are:

i. A direct links 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. While the concepts of network efficiency and stability are relegated to Appendix A, I analyse here some simple outcomes of this proposition. 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 links would give lower profits for the firm and lower consumption. For sufficiently low transportation costs, the cost of adding an extra link is less than the firm's gain from replacing an indirect link to a direct link. And so it will always prefer to have a direct link 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 link is severed since utility for consumers there decreases. A firm in country 1 similarly is adversely affected by this choice. The indirect link is severed and the varieties traded decrease. Profits for the firm decrease. Therefore a firm will never choose to sever the link with country 3. Suppose also that a firm in country 1 forms a direct link at this cost level with country 2 instead of the indirect link 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 link 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 link. Therefore any two countries which are not directly connected benefit from forming a link.[4]

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 geographical placement and as a form of a 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 δ 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 links 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 δ 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

[4]See Jackson (2003) for the intuition behind the definitions of efficiency and stability.

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.[5]

5 Aggregate Trade and the Gravity Equation

In this section I derive the general case for aggregate trade in the manufacturing sector of a country with representative firms. However, in the place of iceberg transportation costs that the model commands, the net benefit term and implicitly the network structure is introduced. The presence of the net benefit term implies the existence of a trade off: a shorter distance dictates a lower cost and higher benefit and indirect distance is associated with evermore discounted benefits. In order to yield a testable prediction however, the firm's unmeasurable benefit of forming a link needs to be replaced with each country's network participation share. In this way, we can postulate changes in fixed costs associated with the network and distance that act as the benefit of forming a link. Subsequently the trade off between the two network types based on increased distance and decreased fixed costs will be presented. As a robustness check to this operation, I 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. If a hub network prevails, fixed costs are reduced because the network participation share for a trading partner is decreased. But distance, the proxy for transportation costs, increases. This setting will be preferred against a network formation with direct links if exports under a hub are higher. Otherwise the opposite should hold: For direct links distance is less but fixed costs are higher.

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

$$x_{ij} = L_i \mu_j L_j \frac{v_{ij}^{\sigma-1}}{\theta_j} \frac{1}{\sigma F_i}, \text{ where } \theta_j = \sum_{j,l=1}^{N} n_l v_{lj}^{\sigma-1} \tag{13}$$

and θ_j is an aggregate index of network costs in j, derived from the price index. Since I cannot measure the net benefit from forming a link, I 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 $c_{ij} = d_{ij}^{\beta} * exp(\beta_0)$. Second for every network formation there are fixed costs associated with the network. These fixed costs burden all firms in a specific country. The fixed costs are proportional to the share of each country's participation in the network.

Therefore, the gravity equation of exports between two countries involving also a hub will

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

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 link. 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 with the benefit of forming direct links to all partners. If there is a hub formation present for a bilateral trading relationship then it must be because aggregate exports are higher. Otherwise bilateral trade must be conducted directly, without the presence of a hub.

If there is a hub involved between two countries trading, it must also imply that transportation costs were very high thus preventing direct connections. If there is no hub involved it must be that transportation costs were very low between these trading partners. Suppose that the hub country is k. Country i trades with j via country k. Denote x_{ij}^{I} as the aggregate exports using indirect links, or the hub country k. If country k is not involved, aggregate exports using direct links are denoted x_{ij}^{D} . Taking logs, sectoral export values become:

$$\ln x_{ij}^{D} = \beta_0 + \ln L_i + \ln(\mu L_j) - \beta_1(\sigma - 1)\ln d_{ij} - \beta_2 \ln F_i^{D} - \ln(\sigma\theta_j) \tag{14}$$

and,

$$\ln x_{ij}^{I} = \beta_0 + \ln L_i + \ln(\mu \ln L_j) - \tilde{\beta}_1(\sigma - 1)\ln(d_{ik} + d_{kj}) - \tilde{\beta}_2 \ln F_i^{I} - \ln(\sigma\theta_j) \tag{15}$$

We can control for country size and the impact of relative prices, herein the aggregate network cost indices θ_j, by using country fixed effects as in Chaney (2005). Subsequently, a hub formation will exist when $\ln x_{ij}^{I} \geq \ln x_{ij}^{D}$. Two conditions must then hold:

i. Ceteris paribus, $-\tilde{\beta}_1 \ln(d_{ik} + d_{kj}) \geq -\beta_1 \ln d_{ij}$ for the variable trade costs across formations.

ii. Ceteris paribus, $-\tilde{\beta}_2 \ln F_i^{I} \geq -\beta_2 \ln F_i^{D}$ for the fixed costs associated with the network.

The first condition is derived given that the elasticity of substitution is constant and greater than 1 for a particular sector. The condition can be reorganised as:

$$\frac{\beta_1}{\tilde{\beta}_1} \geq \frac{\ln(d_{ik} + d_{ij})}{\ln d_{ij}} \tag{16}$$

The right hand side can be at least 1. Thus the coefficient of indirect distance must be at least less than the coefficient of direct distance, $\tilde{\beta}_1 \leq \beta_1$.

The reduction of fixed costs ceteris paribus, must also yield similar results:

$$\frac{\beta_2}{\tilde{\beta}_2} \geq \frac{\ln F_i^{I}}{\ln F_i^{D}} \tag{17}$$

The right hand side of this relationship can be at most 1 as there is a percentage *reduction* in the network participation share for the country trading under a hub. This implies that the coefficient of fixed costs under a network involving a hub must be less than the coefficient under direct links and constrained by the right hand percentage difference can obtain some values greater than β_2.

Conversely, a direct connection will exist if $\ln x_{ij}^I < \ln x_{ij}^D$. Subject to a constant and greater than unity elasticity of substitution, the two conditions are hence reversed:

i. Ceteris paribus, $-\tilde{\beta}_1 \ln(d_{ik} + d_{kj}) < -\beta_1 \ln d_{ij}$ for the variable trade costs across formations.

ii. Ceteris paribus, $-\tilde{\beta}_2 \ln F_i^I < -\beta_2 \ln F_i^D$ for the fixed costs associated with the network.

Reorganising the first condition yields:

$$\frac{\beta_1}{\tilde{\beta}_1} < \frac{\ln(d_{ik} + d_{kj})}{\ln d_{ij}} \tag{18}$$

As the right hand side is greater than 1, the coefficient of indirect distance must be greater than the coefficient of direct distance, $\tilde{\beta}_1 > \beta_1$.

Straightforwardly, the same applies for fixed costs, where the right hand side is less than unity and so $\tilde{\beta}_2$ must be sufficiently larger than β_2:

$$\frac{\beta_2}{\tilde{\beta}_2} < \frac{\ln F_i^I}{\ln F_i^D} \tag{19}$$

The decision to trade via hub country k can be governed by a binary indicator function.

$$H(k) = \begin{cases} 1 & \text{trade via country } k \\ 0 & \text{trade directly} \end{cases} \tag{20}$$

Therefore trading via a hub implies the indicator is one and subsequently it must be that $\ln x_{ij}^I \geq \ln x_{ij}^D$ and vice versa. I can then summarise the conditions that must hold in order to deduce evidence of the existence of hub formations:

$$\text{if } H(k) = \begin{cases} k = 1 & \text{then } \dfrac{\beta_1}{\tilde{\beta}_1} \geq \dfrac{\ln(d_{ik} + d_{kj})}{\ln d_{ij}}, \ \dfrac{\beta_2}{\tilde{\beta}_2} \geq \dfrac{\ln F_i^I}{\ln F_i^D} \\ \\ k = 0 & \text{then } \dfrac{\beta_1}{\tilde{\beta}_1} < \dfrac{\ln(d_{ik} + d_{kj})}{\ln d_{ij}}, \ \dfrac{\beta_2}{\tilde{\beta}_2} < \dfrac{\ln F_i^I}{\ln F_i^D} \end{cases} \tag{21}$$

These two conditions constitute the specification for empirically testing whether hub formations exist in international trade. If a hub formation exists then the coefficient on indirect distance via the hub is sufficiently smaller than the coefficient of direct distance. In addition, the coefficient of fixed costs associated with a hub network must be smaller (and for a small range range greater) than the coefficient of fixed costs associated with a network with direct links. If both these conditions hold, then I conjecture that the finding is attributed to the existence of a network structure in international trade flows.

6 Empirical Strategy and Endogeneity

To apply the model to the data, I calibrate the two gravity equations to yield the empirical prediction. The aim, dictated by the model, is to show that the ratio of the direct to indirect distance coefficients is weakly greater than the increase in distance when the hub indicator becomes one. For fixed costs, the ratio of direct to indirect fixed costs coefficients needs to be greater than the reduction in fixed costs induced by a hub network formation. Conversely, the ratio of direct to indirect distance coefficients is less than the increase in distance when the hub indicator is zero, with the same applying for reductions in fixed costs. If this prediction holds, one can make inference about the existence of hub formations, or network formations in international trade flows.

However, a subtle point of the analysis so far is that distance and fixed costs are in fact, endogenous. The causal relationship runs from the regressors to the independent variable in the sense that longer distance, higher fixed costs dampen aggregate trade flows. This is what standard trade theory predicts. Except for this channel of causality the theoretical treatise has yielded an additional channel. That the firm decides in the interest of increasing export levels to switch the network formation and thus distance and fixed costs. If $\ln x_{ij}^{I} > \ln x_{ij}^{D}$ then it is because a hub network formation was preferred and vice versa. If all firms in the sector decide to trade via a hub, they choose implicitly to change the trading route and fixed costs to export. Firms always choose the route that will guarantee higher export flows. So the causality in this case runs from the dependent variable to the regressors of distance and fixed costs. This is a case of simultaneous causality.

The reverse causal relationship is solely attributed to the network. So the initial estimators for distance and fixed costs will be biased and inconsistent and thus require appropriate instrumentation. For this reason I proceed as follows. I define the empirical prediction and test the hypothesis for the existence of hub network formations in international trade flows. I then isolate the effect of the network by instrumenting distance and fixed costs and re-test the hypothesis. If the instruments are identifying the network effect, the hypothesis should fail.

The gravity equation for aggregated goods in a particular sector is:

$$\ln x_{ij}^{D} = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln(\mu_j L_j) + \beta_3(1-\sigma)\ln d_{ij} + \beta_4 \ln F_i^{D} + \beta_5 \ln(\sigma\theta_j) + \zeta_{ij} \quad (22)$$

and,

$$\ln x_{ij}^{I} = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln(\mu_j L_j) + \tilde{\beta}_3(1-\sigma)\ln(d_{ik} + d_{kj}) + \tilde{\beta}_4 \ln F_i^{I} + \beta_5 \ln(\sigma\theta_j) + \xi_{ij} \quad (23)$$

Both these specifications are estimated with OLS. For the moment we can assume that ζ_{ij}, ξ_{ij} are both orthogonal to the independent variables and normally distributed. Later this assumption is relaxed due to the presence of endogeneity. The assumption that the shocks ζ_{ij}, ξ_{ij} affect trade flows within each country pair is retained throughout and so all observations are clustered at this level.

In the data only one value of export volumes can be observed. This value may correspond to trade flows using direct connections and/or some network structure involving a hub location. Hence one cannot observe the differences between $\ln x_{ij}^{D}$ and $\ln x_{ij}^{I}$. So the dependent variable is the same in both estimations. Therefore we can test how the change in the indicator affects the ratio of coefficients of distance and fixed costs, checking also if the fit deteriorates or improves in each estimation. If a hub network formation prevails then the coefficient of indirect distance should be decreased compared to the coefficient of direct distance. The same applies for fixed costs with the exception that the coefficient of indirect fixed costs cannot be much larger than the coefficient of direct fixed costs. And the reverse must hold true. Given that the negative signs of trade costs coefficients, the prediction for existence of hub formations becomes:

Empirical Prediction

$$\text{if } H(k) = \begin{cases} k = 1 & \text{then } \dfrac{\beta_3}{\tilde{\beta}_3} \geq \dfrac{\ln(d_{ik} + d_{kj})}{\ln d_{ij}}, \; \dfrac{\beta_4}{\tilde{\beta}_4} \geq \dfrac{\ln F_i^{I}}{\ln F_i^{D}} \\ \\ k = 0 & \text{then } \dfrac{\beta_3}{\tilde{\beta}_3} < \dfrac{\ln(d_{ik} + d_{kj})}{\ln d_{ij}}, \; \dfrac{\beta_4}{\tilde{\beta}_4} < \dfrac{\ln F_i^{I}}{\ln F_i^{D}} \end{cases} \quad (24)$$

For the distance coefficients, it should be $\beta_3, \tilde{\beta}_3 < 0$. When $h = 1$, it is expected that $\tilde{\beta}_3 < \beta_3$. Conversely, when $h = 0$, then $\tilde{\beta}_3 > \beta_3$. The same application should hold for fixed costs. When $h = 1$, we can anticipate $\tilde{\beta}_4 < \beta_4$ or $\tilde{\beta}_4 \geq \beta_4$ by a small amount. When the hub indicator is zero the observation should become $\tilde{\beta}_4 >> \beta_4$.

If both these findings occur, it is inferred that the dependent variable, being export flows of a particular sector, corresponds to trading using a hub formation or direct trading and then this is proof of a network structure in international trade flows. It can also be inferred that distance and fixed costs are endogenous due to the decision of the firms to export in that sector using a particular network structure.

Lastly, I control for heterogeneity and for the price index with fixed effects by introducing country dummies, year dummies, industry, sector and transport mode dummies. A robustness check to the empirical prediction using alternative specifications and samples is available in Appendix D; the findings appear to be inline with the findings presented in the following sections.

7 Construction of the Dataset

The dataset is constructed by merging data from various available sources. Exports from 2003 to 2007 are obtained through the World Bank WITS interface and the classification level is 2 digit HS 1988/1992 for all possible trading partners. I perceive that goods under the umbrella of each 2 digit aggregation forms a distinct sector. These observations are matched with corresponding data on ad-valorem maritime transport costs of the same classification level and year. Observations for maritime transport costs as well as the mode of transport are available under subscription to the OECD Maritime Transport Costs Database. The justification for such an operation is twofold. First, according to Hummels (2007), 99% of the world's trade by weight is carried by sea and based on the OECD (2008) 90% of the volume of world trade is transported by sea. Second, with the existence of hub and spoke networks induced by the advent of containerisation it is not strict to assume that the observed price of shipping services, i.e. transport costs, is a function of the network organisation of the transportation sector. 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 transport costs are removed. Each surviving trade partnership is assigned nominal GDP values, two measures of distance, border and language characteristics. GDP values are obtained through the World Bank Databank. Information on capital distances, border and languages come from the Trade, Production, and Protection Database compiled by Nicita and Olarreaga and the CEPII GeoDist dataset compiled by Mayer and Zignago.

I construct a measure of indirect distance in addition to capital distance. For a subset of trading partners observations by measuring the distance from the capital of the exporter to the closest major exporting port are measured. The exporting port is located within the country of export and for the cases of landlocked countries the closest major foreign port is chosen, through which it may proceed to export. Then trading partner port distances are measured using the U.S. National Imagery and Mapping Agency Distances Between Ports publication as well as the online resource Port World Distance Calculator. For each particular partnership, listed in table E-II, the main country ports of origin and destination are assessed and the shortest shipping route is calculated. If the shipping route requires passage through any of the below exogenously imposed *hub* areas, the route is assigned an indicator value equal to one or otherwise zero. No such distinction is made for direct capital distance. These areas are the Panama Canal, the port of Gibraltar, Port Said, Singapore, Cape Town, Istanbul, Paranagua and the port of Arica. Finally the distance from the major importing port to the capital is measured. To provide an example of this construction, the direct distance between Beijing and Brasilia is 16,937 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 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.

The data come with a weakness in the sense that transport costs for individual E.U. 15 countries are not observed. For this reason, and when otherwise not available, all other units are aggregated to provide approximations at the 15 country level. The capital distance of the E.U. 15 entity with the rest of the world is then measured from Brussels and its main export/import port becomes Rotterdam.

8 Existence of Hub networks in international trade

The theoretical approach herein asserts that geographical characteristics now become endogenous elements as countries open to trade: in order to protect themselves from adverse transport costs they create a hub network and can determine hub location either by concentration of production and historical precedence or simply by geographical advantage. This helps to explain why neighbouring countries have more integrated networks than countries which are not, and infrastructure crucially determines country access and transit facilitation (Limao & Venables (2001); Krugman (1993)).

The empirical specifications are estimated robust to heteroscedastic clustered standard errors in the presence of country, year and sectoral, industrial and transport mode fixed effects. I find that equation (24) holds for the distance variables, when controlling for all effects except sectoral. When controlling for heterogeneity across sectors the effect is reversed and equation (24) fails for distance.

I cannot make any conclusive inference about the coefficients for fixed costs as the border and language variables act in the same manner as the religion variable in Helpman et al. (2008). They assume that common religion affects the fixed cost of export but not the variable cost of export. This assumption must therefore hold for common language and border, however Anderson (2010) mentions the potential identification issue that this approach may entail.

Most importantly, the reason why these are not good proxies to the theoretical model is that they do not capture the change in costs due to the change in the network. Suitable variables to extract inference concerning fixed costs would be customs data such as exporting costs per good type per destination but these are widely unavailable. While country specific export and import costs are available from the World Bank, they are absorbed by the country fixed effects. The lack of suitable proxies for fixed costs remains a challenge for the validation of the second half of the empirical prediction (24) and may necessitate structural estimation. Hence, for the remainder of this version of the paper only the variable cost prediction is validated and treated for endogeneity: half of the empirical prediction is proved due to lack of suitable data.

Table 2: Exports: Direct and Indirect Distance

Variable	(1) Exports	(2) Exports
Direct Distance	-0.957***	
	(0.143)	
Indirect Distance		-1.081***
		(0.198)
Common Language	1.130***	1.040***
	(0.199)	(0.217)
Common Border	0.958**	1.185***
	(0.461)	(0.388)
Exporter GDP	1.058***	1.093***
	(0.0341)	(0.0358)
Importer GDP	0.836***	0.905***
	(0.0366)	(0.0366)
Fixed Effects		
Exporter	-	-
Importer	-	-
Sector	YES	YES
Year	YES	YES
Mode	YES	YES
Industry	YES	YES
Observations	73,237	73,237
R-squared	0.521	0.517

Pooled OLS estimation for years 2003 – 2007, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Constants included but not reported.

The first set of results in Table 2 is the outcome of equations (22) and (23) which are represented in columns (1) and (2) respectively. There are controls for year, sectors within an industry, the industry itself and the transport mode, and apply no control on country size. An industry is defined in the OECD dataset as an aggregate of sectors that fall under the umbrella of Agriculture, Manufacturing, Crude Oils and Raw Materials[6]. Given these controls, aggregate prices are incorporated in the orthogonal error term (Anderson & van Wincoop 2004). For these estimations there is no specific assumption about the presence of a hub indicator. The results exhibit the typical adversity caused by direct distance. They also exhibit an increased adversity caused by indirect distance. Since on average indirect distance is 3,773 km longer than direct distance, we observe a higher coefficient for indirect distance than direct distance for the same sample. This result is straightforward: given two measures of distance, one indirect and one direct, with the former being on average greater than the latter, an exporter will choose the shorter distance. Both coefficients of distance fall within the range of the observed estimates in the trade literature (Overman et al. 2001) and the same applies for the elasticities of GDP being ~ 1. The common border variable enters also with a typical observed elasticity (see Limao & Venables (2001) and Helpman et al. (2008)) but the language coefficient is perceived as very high.

For these two specifications I now proceed to incorporate the hub indicator. This will test the behaviour of relationship (22) characterised by direct distance, when the indicator is zero and one. The same exercise is repeated for relationship (23) which employs indirect distance. In this case controls are applied for country size and the relative prices by including country dummies (Chaney 2005). These are also controls for the transport mode and the industry but sectoral variation is left to influence.

Tables 3 and 4 outline the results for this experiment. Let us assume that we are in a world with direct connections. That is the hub indicator is zero. Therefore if everybody is trading directly it is because transportation costs are lower than indirect trading according to our theoretical treatise. The first two columns of Table 3 indicate this effect. Direct distance impacts trade flows 24% less than trading via a hub country. That is for trade flows in the data whose indirect distance does not involve transit via a hub, exporters should (*and appear to*), utilise the shortest possible distance to the importer. The situation is reversed when the hub indicator is turned to one. Now we are assuming that partners are trading indirectly because trade costs are less in this regime compared to the alternative. Columns (3) and (4) reveal the reversed impact of the two measures of distance. For observed trade flows that pass through a hub location, utilising direct routing is considered 5% more costly than trading indirectly. This holds despite the fact that indirect distance is greater than direct distance by approximately 36%. This somewhat counter intuitive result seems to confirm the theoretical prediction of this study. Whenever a geographically advantageous country exists that helps to link two relatively remoter countries, partners should (*and appear to*)

[6] For the detailed industrial composition of sectors see Appendix Table E-I.

avoid direct trading due to the higher costs involved.

Table 3: Exports: Impact of the Hub Indicator

Variable	(1) Exports Hub=0	(2) Exports Hub=0	(3) Exports Hub=1	(4) Exports Hub=1
Direct Distance	-1.430***		-0.925***	
	(0.113)		(0.268)	
Indirect Distance		-1.877***		-0.881***
		(0.249)		(0.278)
Common Language	0.971***	0.815***	0.696***	0.671***
	(0.198)	(0.276)	(0.147)	(0.139)
Common Border	1.329***	1.154*	-1.104**	-0.874*
	(0.469)	(0.690)	(0.513)	(0.512)
Fixed Effects				
Exporter	YES	YES	YES	YES
Importer	YES	YES	YES	YES
Sector	-	-	-	-
Year	YES	YES	YES	YES
Mode	YES	YES	YES	YES
Industry	YES	YES	YES	YES
Observations	41,481	41,481	32,055	32,055
R-squared	0.413	0.402	0.465	0.465

Pooled OLS estimation for years 2003 – 2007, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Constants included but not reported.

Table 4: Empirical Prediction

	$\frac{\beta_3}{\tilde{\beta}_3}$		$\frac{\ln(d_{ik}+d_{kj})}{\ln d_{ij}}$
h=1	1.05	$>$	1.03 (mean)
h=0	0.76	$<$	1.03 (mean)

Taking the ratios of the distance coefficients and comparing them to the mean of the ratio of indirect to direct distance the empirical prediction (24) is confirmed as shown in Table 4. This result seems to hold when allowing for sectoral variation in the data. However when sectors are controlled for by incorporating a dummy for each of the 96+1 HS 2 digit aggregates the result breaks down as is shown in Table 5.

When adding sectoral fixed effects, indirect distance always seems to be costlier to exporters compared to direct distance irrelevant of the value of the hub indicator. This interesting finding implies that the impact of the hub indicator is driven through heterogeneity existing across sectors, given the mode of transport for each commodity produced. It is equivalent to stating that when controlling for sectoral differences by virtue of the fixed effect, any unobserved information implying that some sectors within each country pair trade by using hub formations is subsumed by the fixed effect.

Table 5: Exports: Impact of the Hub Indicator

Variable	(1) Exports Hub=0	(2) Exports Hub=0	(3) Exports Hub=1	(4) Exports Hub=1
Direct Distance	-1.482***		-1.031***	
	(0.117)		(0.241)	
Indirect Distance		-2.080***		-1.101***
		(0.242)		(0.229)
Common Language	1.329***	1.106***	0.478***	0.454***
	(0.229)	(0.278)	(0.132)	(0.114)
Common Border	1.403***	1.123	-1.017**	-0.855*
	(0.506)	(0.700)	(0.456)	(0.461)
Fixed Effects				
Exporter	YES	YES	YES	YES
Importer	YES	YES	YES	YES
Sector	YES	YES	YES	YES
Year	YES	YES	YES	YES
Mode	YES	YES	YES	YES
Industry	YES	YES	YES	YES
Observations	41,481	41,481	32,055	32,055
R-squared	0.615	0.607	0.627	0.627

Pooled OLS estimation for years 2003 – 2007, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Constants included but not reported.

Grouping observations within each industrial aggregate, we can observe how sectoral differences affect the impact of the hub indicator from the viewpoint of agricultural/perishable goods, or differentiated manufactured products. While the former goods would accommodate modes of transport or transport networks to be dispatched to the destination directly, or in the least possible time, the same may not apply for the latter. Relaxing the sector fixed

Table 6: Empirical Prediction

	$\frac{\beta_3}{\tilde{\beta}_3}$		$\frac{\ln(d_{ik}+d_{kj})}{\ln d_{ij}}$
h=1	0.93	$<$	1.03 (mean)
h=0	0.71	$<$	1.03 (mean)

effects and repeating the estimation at the industry level for agriculture and manufactures, the empirical prediction is observed in manufactures but not in agriculture. The summaries of the empirical prediction for these two broad aggregates are presented below with output tables together with the sectors contained within each industry contained in Appendix C.

	$\frac{\beta_3}{\tilde{\beta}_3}$		$\frac{\ln(d_{ik}+d_{kj})}{\ln d_{ij}}$
h=1	0.93	$<$	1.03 (mean)
h=0	0.69	$<$	1.03 (mean)

(a) Agriculture - Empirical Prediction

	$\frac{\beta_3}{\tilde{\beta}_3}$		$\frac{\ln(d_{ik}+d_{kj})}{\ln d_{ij}}$
h=1	1.14	$>$	1.03 (mean)
h=0	0.79	$<$	1.03 (mean)

(b) Manufactures - Empirical Prediction

Summarising these results, I confirm, at least for variable trade costs, the existence of hub network formations in international trade flows. These formations appear to be driven mainly by the characteristic nature of the types of goods produced by each sector. The reason for existence of these formations according to the theoretical framework is solely attributed to cost reductions associated with network formations. When transportation costs are high (longer distances), partners are more inclined to form hub network formations in order to increase their export volume rather than trade directly. When transportation costs are low (shorter distances) partners are increasing their export volume by trading directly rather than involving a beneficially located trading partner.

When controlling for differences across sectors the result is treated; information on the network appears to be subsumed by the fixed effect. To validate this result further, more reliable measures of fixed costs associated with export flows are required. Specifically measurements of fixed costs related to the network structure of transportation be it direct trading or transit costs at hub locations are desirable, however these costs are unavailable at present.

8.1 Addressing Endogeneity

The decision to trade using a hub network or a direct connection is made by the exporting firm based on its observations of transportation costs. It will prefer the formation that allows for the greatest exporting volume to be produced. For this reason exists a reverse channel of causality which dictates that the estimates for the two distance variables are endogenous.

The endogeneity is caused by the effect of the network formation. And so if a candidate instrument exists such that it identifies the network effect but retains the standard causal relationship, it will also invalidate straightforwardly the empirical prediction (24). The instrument should be impervious to the fluctuations caused by the hub indicator but should retain the information of the impact of variable trade costs on trade flows.

A candidate instrument satisfying relevance and exogeneity criteria is time difference. In that sense firms in a country export to other countries which have small time differences rather than large time differences. This has not so much to do with the actual distance that goods have to travel to reach the destination, but is related with the relative ease of building trade partnerships through channels of communication.[7] The time barrier of trade should contain information that is relevant to distance but then it should not contain information pertaining to the network structure of the transportation sector. The current model is parsimonious to explain the effect of the network through the construction of a separate transportation sector. In reality, the firm would most probably produce specific export volumes of a good by observing a transportation cost. It would not be able to deduce or exert any power upon the route that the good will take other than commanding an arrival date to the destination. This reasoning leads to the satisfaction of the relevance and exogeneity criteria for the instrument.

Yet the coefficient of time difference would be inefficient as it would fail to capture variation on a vertical trade axis which it interprets as minimal. The notion of time difference can then be exploited on a two axis assumption. Since time difference measured on a scale from zero to 24 hours can be captured at 15 degree intervals of longitudinal lines, the same can be applied at 7.5 degree intervals of latitudinal lines. Measures of time difference are constructed by subtracting the longitude of the importer's capital from the longitude of the exporter's capital divisible by 15 degrees to obtain the 24 hour measure of standard time difference. Then by subtracting the latitude of the importer's capital from the latitude of exporter's capital divisible by 7.5 degrees we can obtain the 24 hour measure of vertical time difference. Subsequently the absolute value of each difference is taken in order to obtain a measure of time difference expressed on a decimal scale rather than hours and minutes. This will generate sufficient variation across country pairs. I then estimate the gravity equations (22) and (23) using one overidentifying restriction for both direct and indirect distances. If both instruments are valid, tests for relevance and exogeneity are carried out. If exogeneity fails, the 2SLS estimate is reported otherwise the OLS equivalent of Table 3 are replicated. If one instrument is valid then I the vertical measure of time difference and repeat relevance and exogeneity tests using standard time difference. If exogeneity fails, the 2SLS equivalent

[7] An interesting validation of this thought experiment was realised on 29 December 2011, when Samoa and Tokelau skipped a day in order to ease trade relationships with Australia and New Zealand. According to the Prime Minister of Samoa, "While it's Friday here, it's Saturday in New Zealand, and when we're at church on Sunday, they're already conducting business in Sydney and Brisbane." Source: (BBC News, 30 December 2011).

is presented or otherwise the OLS distance equivalent of Table 3.

Table 7: Exports: Impact of the Hub Indicator

Variable	(1) 2SLS Exports Hub=0	(2) OLS Exports Hub=0	(3) 2SLS Exports Hub=1	(4) 2SLS Exports Hub=1
Direct Distance	-1.175***		-1.164***	
	(0.156)		(0.259)	
Indirect Distance		-1.877***		-1.179***
		(0.249)		(0.277)
Common Language	1.104***	0.815***	0.710***	0.681***
	(0.224)	(0.276)	(0.152)	(0.152)
Common Border	1.543***	1.154*	-1.344**	-1.112**
	(0.496)	(0.690)	(0.518)	(0.533)
Fixed Effects				
Exporter	YES	YES	YES	YES
Importer	YES	YES	YES	YES
Sector	-	-	-	-
Year	YES	YES	YES	YES
Mode	YES	YES	YES	YES
Industry	YES	YES	YES	YES
Observations	41,481	41,481	32,055	32,055
Overidentifying Restrictions	0	0	1	0
Robust Hausman Statistic	9.63***	0.02	4.95**	4.04**
R-squared	0.412	0.402	0.464	0.446

Pooled 2SLS/OLS estimations for years 2003 – 2007, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Constants included but not reported.

Table 8: Empirical Prediction

	$\frac{\beta_3}{\tilde{\beta}_3}$		$\frac{\ln(d_{ik}+d_{kj})}{\ln d_{ij}}$
h=1	0.98	$<$	1.03 (mean)
h=0	0.62	$<$	1.03 (mean)

The results of this experiment are contained in Tables 7 and 8. Two interesting outcomes are observed. First, all estimations except (3) are exactly identified and (2) has no trace of endogeneity. The desired network configuration, which is the realisation of the firms endogenous decision, contains the endogenous variable of distance. In that sense when the hub indicator is zero, firms according to the theoretical expansion prefer direct trading compared to indirect trading. Then column (1) contains the treated coefficient for direct distance as the null of exogeneity was rejected. This does not seem to apply for indirect distance as the Hausman statistic supports exogeneity and so column (2) listing the alternative decision of the firm is unchanged. When the hub indicator is one it is because firms prefer indirect trading to direct trading. Then column (4) lists the treated coefficient for indirect distance since the latter was rejected under the null of exogeneity. Column (3) however shows the alternative decision for direct trading when the hub is one, and lists the overidentified treated coefficient of direct distance. The endogenous decisions of firms in columns (1) and (4) are rejected under the null of exogeneity when allowing for variation across sectors. The alternative decisions of firms are somewhat misleading: when the hub indicator is zero, the alternative decision is exogenous; when the hub indicator is one, the alternative decision is endogenous. One possible reasoning for this finding is that perhaps the instrument is also identifying sectoral variation (since all other variation according to the model is controlled for) that is not related to distance in either or both of the two samples. Another reasoning could be identification, in the sense that the exogenous variables in the first stage may not be identified correctly or are very strongly correlated with each other.

Second, by instrumenting distance with time difference we can invalidate the empirical prediction (24). In Table 8, ratios of distance coefficients under the one and zero outcomes of the hub indicator are less than the ratios of indirect to direct distance. The implication is that the endogenous variable, when identified and treated, should contain no information about the network structure, be it direct connections or a hub network since it violates the testable prediction. This leads to the conclusion that time difference is a valid instrument for harnessing the impact of network structures that might be present, but unobserved, in variable trade costs suggesting caution on the exogeneity assumptions for trade costs.

I conclude that the endogenous decision of firms to choose the network formation is evident in international trade flows. This inference comes after validating the empirical prediction of the theoretical model: firms choose the formation that is minimising their exposure to transportation costs, herein proxied by distances that correspond to actual measurements of transport costs. The endogeneity can be harnessed using two tools. One can either incorporate sector fixed effects as it appears that the information on the network structure based on the types of goods produced is alleviated by the fixed effect. As an alternative, one can choose to instrument variable trade costs but only if these variable costs are suspect of containing network effects.

9 Conclusion

I study the reasons for formation and the impacts of transportation hub networks in international trade by merging the symmetric connections model of Jackson & Wolinsky (1996) in a trade model of monopolistic competition with representative firms. Under reasonable assumptions, firms begin to export and choose the network formation that is minimising their exposure to transport costs. As such when transport costs are extremely high countries remain closed. Upon their gradual reduction, firms commence exporting and create hub networks that are associated with higher levels of costs. The equilibrium is attained given a fixed benefit value for which a unique cost leads to satisfaction of the zero profit condition. Further reduction of variable trade costs causes firms to favour direct trading and the equilibrium is admitted if the exogenous benefit of connecting directly exactly offsets the marginal cost of reaching the destination.

Through the equilibrium gravity equation, the empirical prediction is formed by which firms choose the formation that maximises the volume of output given two measures of distance, one direct and one indirect. I find that when direct connections are preferred the coefficient for direct distance is less than the coefficient of indirect distance to the same destination. On the other hand when indirect connections are preferred the coefficient of indirect distance is lower than the coefficient of direct distance. These findings carry the justification of endogeneity insofar as firms, depending on the type of goods they produce, prefer a direct or an indirect route. As such, agricultural goods seem to be always directly traded. Manufactured goods tend to be indirectly traded. This endogeneity can be treated in two ways, either by incorporating sectoral fixed effects or by instrumentation when there is suspicion that the variable trade costs may contain information on the unobserved or unknown network formation.

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, irrespective of distance tend to be low (assuming that transportation markups are not variable as in the case of Hummels et al. (2009)). This does not affect much the productivity or number of firms. Transportation costs on low-volume trading routes, tend to be higher and distance here plays a crucial role. This should affect the productivity of exporting firms and the number of firms more severely. Additional factors are 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 transport costs achieve a more beneficial trading position when forming a transportation network by utilising a hub. The link with at least one proximate geographically advantaged partner improves market access, ameliorates exposures to these costs and leads to improvements in own and transit infrastructure.

A Proof of Efficiency and Pairwise Stability of network formations

A.1 Efficiency

Based on the definitions of Jackson & Wolinsky (1996), each graph g has a value $y : \{g|g \subset g^N\} \to \mathbb{R}$, where $y \in Y$ is the set of functions. The value is an aggregate of individual exporting firm values $y_j(g) = \sum_{i:ij \in g} v_{ij}(g)$, where $v_{ij} : \{g|g \subset g^N\} \to \mathbb{R}$ is a net benefit value. The graph $g \subset g^N$ is *efficient* if $y_j(g) \geq y_j(g')$ for all $g' \subset g^N$.

Case 1, Direct Trading vs Autarky: $G = \{g_{ij} = \{12, 21\}, g'_{ij} = \{\emptyset\}\}$— The simplest case shows how forming a direct link is more efficient than not forming a link. Firms profits and utility increases as long as the benefit of forming a link exceeds its cost for countries 1 and 2. Due to symmetry, global utility and firm profits (until the trade equilibrium restores profits to zero) increase:

$$y_1(g) = (\delta - c_{12}) = (\delta - c_{21}) > 0, \text{when } \delta > c_{ij} \tag{A-1}$$

Case 2, Direct Trading vs Indirect Trading: $G = \{g_{ij} = \{12, 21, 13, 31, 23, 32\}, g'_{ij} = \{13, 31, 23, 32\}\}$— In the case of the direct links network with three countries trading, direct connections are more efficient than indirect connections and no trading at all. Denote the graphs originating from direct trading between any integer pair $\{i, j\} \in [1, 3]$ as g^D. For pairs that are indirectly trading denote their graph as g^I. Therefore in case (i) of the Proposition, country 1 obtains from trading to 2 and 3:

$$y_1(g^D) = (\delta - c_{13}) + (\delta - c_{12}) > \delta - c_{13} + \delta^2 = y_1(g^I) \tag{A-2}$$

Country 3 similarly yields:

$$y_3(g^D) = (\delta - c_{31}) + (\delta - c_{32}) = (\delta - c_{31}) + (\delta - c_{32}) = y_3(g^I) \tag{A-3}$$

because in the two alternate formations country 3, being in the middle, must always form direct connections. The symmetric case for country 1 applies for country 2.

Hence, total profits (or utilities in the receiving countries) have increased by
$y = 6(\delta - c) > 0$.

Case 3: Indirect Trading vs Direct trading: $G = \{g_{ij} = \{13, 31, 23, 32\}, g'_{ij} = \{12, 21, 13, 31, 23, 32\}\}$— Countries 1 and 2 are symmetric, therefore the following is obtained also for country 2.

$$y_1(g^I) = (\delta^2 + \delta - c_{13}) > (\delta - c_{13}) + (\delta - c_{12}) = y_1(g^D) \tag{A-4}$$

For country 3:

$$y_3(g^D) = (\delta - c_{31}) + (\delta - c_{32}) = (\delta - c_{31}) + (\delta - c_{32}) = y_3(g^I) \tag{A-5}$$

Hence, total profits (or utilities in the receiving countries) have increased by $y = 2\delta^2 + 4(\delta - c) > 0$.

A.2 Pairwise Stability

There exists an allocation rule $K : \{g|g \subset g^N\} \times Y \rightarrow \mathbb{R}^{\mathbb{N}}$ and $K_j(g, y_j)$ is the distribution of each network value to individual firms or representative agents. The graph is pairwise stable w.r.t. y and K if:

a. For all $ij \in g, K_j(g, y_j) \geq K_j(g - ij, y_j)$ and $K_i(g, y_i) \geq K_i(g - ij, y_i)$.

b. For all $ij \notin g, K_j(g, y_j) < K_j(g + ij, y_j)$ then $K_i(g, y_i) > K_i(g + ij, y_i)$.

The implication is that if j strictly prefers to form link ij and i is indifferent, the link is formed.

Case 1, Direct Trading $c < \delta - \delta^2$***:*** —

a. For $\{ij\} \in [1, 3]$: If any one link is severed, utility (profits) decrease for the trading pair: $\delta - c_{ij} > 0$ as long as $\delta > c_{ij}$.

b. For $\{ij\},\ i, j \in \{1, 3\}$ and country 2 remaining autarkic: utility (profits) of 2 connecting to 1 or 3 are lowered: $\delta - c_{2j} < 0$. But if $\delta - c_{2j} > 0$, then utility (profits) increase and all links are formed.

Case 2, Indirect Trading $\delta - \delta^2 < c$***:*** —

a. For $\{ij\} \in [1, 3]$: If any one link is severed, utility (profits) decrease for the trading pair: $\delta^2 + \delta - c_{ij} > \delta - c_{ij} > 0$ for countries, 1 and 2. For country 3 connected to 1 and 2, $2(\delta - c_{3j}) > (\delta - c_{3j}) > 0$ as long as $\delta > c_{3j}$.

b. For $\{ij\},\ i, j \in \{1, 3\}$ trading directly and country 2 trading directly to 3. It considers connecting directly or indirectly to 1. Utility (profits) of 2 connecting directly to 1 are lowered: $2(\delta - c_{2j}) < 0 < (\delta - c_{23}) + \delta^2$. But if it actually connected to 1 directly, country 1 has utility (profits) decreased: $2(\delta - c_{1j}) < 0 < (\delta - c_{12}) + \delta^2$ and therefore country 1 will break the link to country 3 and start trading indirectly.

It is deduced that when costs are low, $c < \delta - \delta^2$, the direct links network is uniquely efficient and pairwise stable. When costs are consistent with the range $\delta - \delta^2 < c$, the indirect links network is uniquely efficient. It is also pairwise stable but not necessarily unique as the system can rotate between partners.

B A Model of IRS in Transportation and Hub Formations

I build a three country model with increasing returns in each country's transportation sector in order to carry the domestically produced output of homogeneous manufacturing firms that engage in trading. This exercise is undertaken in order to show 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 expansion. 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 and produced under constant returns that will be defined as a 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: — Agents in country $i \in K$ notwithstanding their sector of occupation, consume differentiated varieties of agricultural and manufacturing goods under the same utility function,

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

where q_0 is consumption of the aggregate agricultural, $1 - \mu$ is the share of income expenditure on 0, $\sigma > 1$ is the elasticity of substitution between pairs of varieties and Ω 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:

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

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

I then rewrite the demand function as $q_{ij} = \frac{\mu L_j p_{ij}^{\frac{1}{\theta-1}}}{P}, j \in K$, where P will represent the price index.

Manufacturing Production and Trade Costs — Good 0, the agricultural good, is the numeraire homogeneous good. One unit of labour produces w and 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. *Contrary to transportation costs, the fixed costs of manufacturing for the remainder of model will be constant.*

To produce and sell a variety ω 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 labour share in manufacturing. Lastly all firms must produce goods that are consumed domestically and abroad after exporting.

Pricing Regime: — C.i.f. prices of imported goods consist of a multiplicative iceberg cost $\tau_{ij} \geq 1$ and an additive transport price that is the optimal price set by the transportation firm:

$$p_{ij}^{cif} = p_{ij}^{fob} \tau_{ij} + f_{ij} \tag{B-1}$$

Transportation Production: — 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[8] Each country uses transport services produced domestically. Each specific variety is produced by a single transportation

[8]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.

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 variable overhead/fixed cost,* $d_{ij} > 0$ *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

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

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 maximising price becomes $\frac{p_{fob}}{w} = \frac{\gamma}{\theta}$ 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 $\frac{p_{fob}}{w} = \gamma + \frac{F_i^m}{x_i}$, where x_i is the total output produced by each firm.

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

$$x_i = \sum_{i,j=1}^{3} q_{ij}t_{ij} = \frac{F_i^m}{\gamma}\frac{\theta}{1-\theta} \tag{B-2}$$

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

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

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 through the c.i.f. price of the manufacturing good and the intermediate input assumption. Provided $i \neq j \in K$, their profit function is

$$\Pi_i^t = \sum_{i,j=1}^{3} f_{ij} q_{ij} - \sum_{i,j=1}^{3} w d_{ij} q_{ij} - w F_i^t$$

I am required to 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.

Given this assumption, transport firms proceed to profit maximisation and yield transport prices that are a function of the f.o.b. price and a markup over transport marginal cost due to the transport elasticity of import demand [9]:

$$\frac{f_{ij}}{w} = \frac{d_{ij}}{\theta} + \frac{\gamma \tau_{ij}}{\theta} \frac{1-\theta}{\theta} \tag{B-3}$$

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 characterisation of the transport price allows us then to characterise the c.i.f. price, reminding that $w = 1$ for all countries:

$$p_{ij}^{cif} = \frac{1}{\theta}(p_{ii}^{fob} \tau_{ij} + d_{ij}) = \frac{1}{\theta}(\frac{\gamma}{\theta} \tau_{ij} + d_{ij}) \tag{B-4}$$

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

Consumption Ratios: — 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):

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

For simplicity let us 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 have to face are:

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

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

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, let us make one last normalisation since the symmetry of the iceberg trade cost τ and the common marginal cost γ are identical across countries. Hence, impose $\tau = \gamma = 1$.

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

$$f_{31} = f_{32} = d + F_3^t (q_{31} + q_{32})^{-1} \Longrightarrow q_{31} = \frac{1}{2} \frac{F_3^t}{\frac{1}{\theta} + d} \frac{\theta}{1-\theta} \tag{B-5}$$

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: — 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 transport price. The relationship can be rearranged to write:

$$q_{13} = q_{31} = \frac{F_1^t}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \frac{\theta}{1-\theta} \tag{B-6}$$

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

Hub formations: — The left hand sides of equations (B-5) and (B-6) are necessarily the same as it is the expression of the common unit of exports. We have assumed that the overhead costs of transportation F_i^t in any country can be variable. Equating the two terms then yields a ratio of the fixed cost of transport in the two countries:

$$\frac{F_3^t}{F_1^t} = \frac{2\left(\frac{1}{\theta} + d\right)}{\frac{1}{\theta} + d + \left(\frac{1}{\theta} + d'\right)^{\frac{\theta}{\theta-1}} \left(\frac{1}{\theta} + d\right)^{\frac{1}{1-\theta}}} \tag{B-7}$$

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 (B-7) has $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} > 0$:

$$\begin{aligned} \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= -\frac{\left[2\left(\frac{1}{\theta}+d\right)\right]\frac{\theta}{\theta-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}-1}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}{\left[\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\right]^2} \iff \\ \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d'} &= \frac{F_3^t}{F_1^t}\frac{\frac{\theta}{1-\theta}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}-1}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}} > 0 \end{aligned} \tag{B-8}$$

Since both fractions are positive.

(B-7) has also $\frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} < 0$:

$$\begin{aligned} \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{2}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}} \\ &- \frac{\left[2\left(\frac{1}{\theta}+d\right)\right]\left(1+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\left[\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\right]^2} \iff \\ &\text{multiplying and dividing the first term with } \left(\tfrac{1}{\theta}+d\right) \text{ yields} \\ \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left[\left(\frac{1}{\theta}+d\right)^{-1} - \frac{\left(1+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}-1}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}}\right] \iff \\ \frac{\partial \frac{F_3^t}{F_1^t}}{\partial d} &= \frac{F_3^t}{F_1^t}\left(\frac{1}{\theta}+d\right)^{-1}\left[1-\frac{\left(\frac{1}{\theta}+d+\frac{1}{1-\theta}\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}\right)}{\frac{1}{\theta}+d+\left(\frac{1}{\theta}+d\right)^{\frac{1}{1-\theta}}\left(\frac{1}{\theta}+d'\right)^{\frac{\theta}{\theta-1}}}\right] < 0 \end{aligned} \tag{B-9}$$

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 transportation 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 transportation 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 link 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 variable fixed costs in transportation, the benefit of forming a link 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 links.

B.5 The number of transport firms

For concreteness, I close the model by characterising the number of transportation 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:

$$n_1^t = \frac{L^t}{F_1^t}(1-\theta) \times \left[(1-\theta) + K\frac{1}{2}\frac{F_3^t}{F_1^t}\right]^{-1} \tag{B-10}$$

where $K(d, d')$[10] 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 δ 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

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

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 crucially 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:

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

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 δ'. 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 Results for broadly defined Agricultural and Manufacturing Industries

Table C-I: Agricultural Exports

Variable	(1) Exports Hub=0	(2) Exports Hub=0	(3) Exports Hub=1	(4) Exports Hub=1
Direct Distance	-1.333***		-0.993**	
	(0.0981)		(0.436)	
Indirect Distance		-1.909***		-1.060**
		(0.235)		(0.444)
Common Language	0.713***	0.491*	0.769**	0.713**
	(0.225)	(0.276)	(0.360)	(0.341)
Common Border	1.191***	0.913*	0.0493	0.253
	(0.387)	(0.547)	(0.677)	(0.622)
Fixed Effects				
Exporter	YES	YES	YES	YES
Importer	YES	YES	YES	YES
Sector	-	-	-	-
Year	YES	YES	YES	YES
Mode	YES	YES	YES	YES
Observations	11,148	11,148	6,985	6,985
R-squared	0.463	0.455	0.531	0.531

Pooled OLS estimations for years 2003 − 2007, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Constants included but not reported.

Table C-II: Manufactured Exports

Variable	(1) Exports Hub=0	(2) Exports Hub=0	(3) Exports Hub=1	(4) Exports Hub=1
Direct Distance	-1.479***		-0.992***	
	(0.130)		(0.201)	
Indirect Distance		-1.857***		-0.869***
		(0.280)		(0.243)
Common Language	1.227***	1.097***	0.710***	0.697***
	(0.207)	(0.303)	(0.182)	(0.196)
Common Border	1.416***	1.261*	-1.610***	-1.308***
	(0.485)	(0.745)	(0.451)	(0.477)
Fixed Effects				
Exporter	YES	YES	YES	YES
Importer	YES	YES	YES	YES
Sector	-	-	-	-
Year	YES	YES	YES	YES
Mode	YES	YES	YES	YES
Observations	28,238	28,238	23,875	23,875
R-squared	0.415	0.403	0.473	0.473

Pooled OLS estimations for years $2003 - 2007$, monetary units are real U.S. Dollars. Standard errors are allowed to be correlated within country pair clusters. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Constants included but not reported.

D Robustness checks to the Empirical Prediction

Consider the gravity equation derived from the theoretical model:

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

Consider ignoring the trade off between benefits and costs of forming links that lead to the construct of v_{ij}. Suppose that, like any other form of trade costs in the gravity equation literature (except for the recent cases in the literature of additive trade costs), the benefit term enters as an element of the set that constitutes the iceberg cost, with other elements being distance, common border and language.[11] In this case we have:

$$(1-\sigma)\ln \tau_{ij} = (1-\sigma)\left(\beta_1 \ln d_{ij} + \beta_2 \text{Border}_{ij} + \beta_3 \text{Language}_{ij} + \beta_4 \text{Number of hubs}_{ij}\right)$$

and the fraction $\frac{v_{ij}^{\sigma-1}}{\theta_j}$ of equation (D-1) is inversed to $\frac{\theta_j}{\tau_{ij}^{\sigma-1}}$.

We can then estimate this equation and try to extract evidence on the sign of β_4. A negative coefficient of β_4, given $\sigma > 1$ leads to a positive sign on the coefficient of the number of hubs and vice versa. This simple estimation forms the first part of what I will refer to as Specification 1.

The second part of Specification 1 is to estimate an interaction between the two types of distances and the number of hubs. Therefore the iceberg is perturbed to:

$$(1-\sigma)\ln \tau_{ij} = (1-\sigma)\left(\beta_1 \ln d_{ij} + \beta_2 \text{Border}_{ij} + \beta_3 \text{Language}_{ij} + \beta_4 \text{Number of hubs}_{ij} \times \ln d_{ij}\right)$$

Given that the sign of $(1-\sigma)\beta_1$ will always be negative, the presence of hubs along the route will be acting beneficially if it is reducing the trade cost, ie $(1-\sigma)\beta_4$ is positive. Then we can infer whether there is an ameliorating impact on distance due to an increase in the number of hubs on the route.

In summary, when referring to Specification 1 the following will be estimated:

Specification 1: $(1-\sigma)\beta_k \text{Number of hubs}_{ij}$ **OR** $(1-\sigma)\beta_l \ln d_{ij} \times \text{Number of hubs}_{ij}$

In an alternative take, I estimate the interaction in the presence of both the number of hubs and the two distance variables. This allows for the intercept to change, but it does not always yield a rejection of a joint hypothesis test on zero coefficients of the number of hubs and the interaction.

[11] See Feenstra (2004) for further details concerning elements of the set of iceberg costs.

Analytically,

$$(1-\sigma)\ln\tau_{ij} = (1-\sigma)\left(\beta_1 \ln d_{ij} + \beta_2 Border_{ij} + \beta_3 Language_{ij} + \beta_4 Number\ of\ hubs_{ij} \right.$$
$$\left. + \beta_5 Number\ of\ hubs_{ij} \times \ln d_{ij}\right)$$

Where the sign of $(1-\sigma)\beta_1$ will be negative and the ameliorating effect of our proposition should dictate that $(1-\sigma)\beta_5$ and $(1-\sigma)\beta_4$ are both positive.

The above construct will be referred to as Specification 2 and is presented in summary below:

Specification 2: $(1-\sigma)\beta_k \text{Number of hubs}_{ij}$ **AND** $(1-\sigma)\beta_l \ln d_{ij} \times \text{Number of hubs}_{ij}$

The third attempt is to model the benefit term of v_{ij}, that is $\delta^{t_{ij}}$. I then incorporate it in the iceberg cost in order to estimate:

$$(1-\sigma)\ln\tau_{ij} = (1-\sigma)\left(\beta_1 \ln d_{ij} + \beta_2 Border_{ij} + \beta_3 Language_{ij} + \beta_4 \ln\left(\delta_{ij}^{t_{ij}}\right)\right)$$

The equation is estimated first using only the term δ_{ij} without considering the number of links t_{ij}, and then including the number of links.

This will form Specification 3 which is summarised below:

Specification 3: $(1-\sigma)\beta_m \delta_{ij}$ **OR** $(1-\sigma)\beta_n \ln\left(\delta_{ij}^{t_{ij}}\right)$

where δ_{ij} =*Adjusted Container Connectivity Index* and t_{ij} =the number of links between i, j.

The term δ_{ij} needs to be suppressed between zero and one and needs to represent a benefit term from forming a link. A positively correlated to exports proxy is the UNCTAD container connectivity index. According to the World Bank "The Liner Shipping Connectivity Index captures how well countries are connected to global shipping networks. It is computed by the United Nations Conference on Trade and Development (UNCTAD) based on five components of the maritime transport sector: number of ships, their container-carrying capacity, maximum vessel size, number of services, and number of companies that deploy container ships in a country's ports. For each component a country's value is divided by the maximum value of each component in 2004, the five components are averaged for each country, and the average is divided by the maximum average for 2004 and multiplied by 100. The index generates a value of 100 for the country with the highest average index in 2004."

This index is also strongly positively correlated with Country GDP. In order to reduce the correlation and also to extract a measure of benefit I rescale the index to have a maximum of 100 instead of 124 that was the case for the data. Then assign to each exporter, importer and country that serves as the hub in between each exporter and importer its specific index for each year. The variable is normalised and the simple average using the number of the total links in each trade flow (exporter, importer and number of hubs) is obtained. Added to each flow average is the absolute value of the minimum observation and restricted to limit zero instead of attaining zero. The resulting variable yields values that range between 0.001 and 0.74. The index is then raised to the power of the number of links of each trade flow and the result is a benefit term that should serve our theoretical projection. This construction is based loosely on the construction of infrastructure variables in Limao & Venables (2001).

D.1 Three alternative samples for robustness

The above three specifications are estimated using three different samples to try to reduce any bias that I may have incorporated in the data due to the exogenous selection of hub areas.

The first sample, contains the original selection that is used in the paper, which is based on the actual geographical distribution of the trade flows.

In the second sample, all countries that have a hub area within their border, in case they coincide to be the exporter or the importer, have their hub counted in the total number hubs that separates them with the destination or the origin respectively. For example, if Singapore was the exporter to EU15, then the total number of hubs would be 3, including the origin and destination and including also Port Said in Egypt. Lastly, I have removed two hub areas which are of lesser trade importance to see if there was any qualitative impact compared to the original first sample. These two areas are Arica, which serves as a hub for Bolivia only and Paranagua which serves as a hub for Paraguay only.

The third sample, has the hub areas within each country removed from the count and does not consider the above-mentioned two areas. Hence the number of hubs between Singapore and the EU15 area becomes one, that of Port Said in Egypt.

D.2 Estimation Procedure

The three specifications are estimated using OLS and sector (two digit), transport mode, year and industry (manufacturing, agriculture, raw Materials, crude Oils) Fixed Effects (FE) and alternate between adding or excluding origin or destination fixed effects. All standard errors are clustered within each country pair and are treated for heteroscedasticity.

To avoid clutter, the tables pertaining to each estimation are not included, and these can be provided upon request from the author. Instead, the results are summarised in three tables, one for each specification. Each table contains the results from each of the three samples, in the presence or not of country FE. Instead of listing the values of coefficients, their signs and their significance or not are considered. S stands for a significant coefficient and I stands for an insignificant coefficient. Concerning the results of joint hypothesis tests for the interaction, R will stand for Rejected and NR for the complement. The new variables and interactions are added to the baseline specification of Table 2 of the main body of the paper

D.3 Summary of Results

D.3.1 Specification 1

The first apparent result is that the coefficient on the number of hubs is positive and significant for most of the experiments involving indirect capital distance. For direct capital distance most experiments yield insignificant coefficients for this variable. This is a simple validation of the proposition of the paper: When increasing the distance between origin and destination the impact to trade is more severe, based on Table 2 of the baseline specification. However when there is an increasing presence of hub ports, ceteris paribus, this has a beneficial impact on trade flows in a regime involving only increased distance (indirect distance), since the latter is on average longer than direct distance.

Table D-I: Specification 1

		Distance		Distance		Distance
		Direct	Indirect	Direct	Indirect	Direct and Indirect
	Country FE	Number of Hubs		Interaction		Ho: See A)
Sample 1	NO	+S(10%)	+S	+I	+S	R
	YES	-I	+S	+S	+S	R
Sample 2	NO	+I	+S	-I	+S	R
	YES	-I	+S	+S	+S	R
Sample 3	NO	+I	+I	-I	+I	R
	YES	-I	+S	+S	+S	R
A) Ho: Coefficient of Distance=Coefficient of Interaction=0						

The second part of Specification 1 incorporates the interaction. For all experiments the Ho described in A) is rejected. Therefore one can observe that under country fixed effects, all experiments yield a positive interaction term confirming further the proposition. The case of indirect distance has positive interaction coefficients without country fixed effects.

In conclusion although increasing distance has a dampening effect on trade flows, if there is an increase in the number of hubs on the route, it lessens the impact. When switching between distance regimes, the impact of the number of hubs or of the interaction in a direct distance regime is less compared to that of indirect distance. The absolute value of coefficients of direct distance in all experiments are less in magnitude compared to those of indirect distance.

D.3.2 Specification 2

The second specification includes both the number of hubs and the interaction as dependent variables. These are then tested for their joint significance.

The outcomes of the experiments are presented in Table D-II. For the direct distance case without country fixed effects, we cannot reject the null hypothesis B) therefore there is no joint impact of the number of hubs and the interaction. However the joint impact of the interaction and direct capital distance is significant, thus supporting Specification 1. In the presence of country fixed effects the results become mixed. The coefficient of number of hubs becomes significant and negative and the coefficient of the interaction is positive and significant thus leading to rejection of their zero joint impact. Concluding, an increase in number of hubs along a route has a negative impact but the impact is lessening as distance is increasing. The same applies for distance, an increase in distance has a negative impact on trade flows but the impact is lessened by increasing the number of hubs.

For the indirect capital distance case without country fixed effects the impact of the number of hubs is insignificant in all experiments. The interaction is positively significant in experiments with country fixed effects whilst positively insignificant in experiments without country fixed effects. The null Hypothesis B) is not always rejected and cases of rejection at the 5% and 10% critical values give rise to ambiguity as to whether to exclude or not the number of hubs as an independent variable from the specification. For cases where the interaction is significant, the same inference applies as to that of the direct distance regime or the outcomes of Specification 1.

Making inference across distance regimes, there might be sufficient evidence against Specification 2 and in favour of Specification 1. The occurrences of insignificant coefficients of hubs in the presence of the interaction term and the frequency of non rejection of the null in the three samples may provide an indication that Specification 1 is more informative.

Table D-II: Specification 2

		Distance				Distance			
		Direct		Interaction		Indirect		Interaction	
	Country FE	Number of Hubs	Interaction	Ho: See A)	Ho: See B)	Number of Hubs	Interaction	Ho: See A)	Ho: See B)
Sample 1	NO	+S (10%)	-I	R	NR	+I	+I	R	R(5%)
	YES	-S	+S	R	R	+I	+S	R	R(5%)
Sample 2	NO	+I	-I	R	NR	+I	+I	R	R(10%)
	YES	-S	+S	R	R	+I	+S	R	R
Sample 3	NO	+I	-I	R	NR	-I	+I	R	NR
	YES	-S	+S	R	R	+I	+S	R	R

A) Ho: Coefficient of Distance=Coefficient of Interaction=0
B) Ho: Coefficient of Number of Hubs=Coefficient of Interaction=0

D.4 Specification 3

The following table includes the adjusted index in the two distance regimes, direct and indirect capital distance. The index is raised to the power of the number of connections in order to yield a measure of benefit from forming a link between i, j as the theoretical part of the paper asserts.

Table D-III: Specification 3

		Distance		Distance		Distance		Distance	
		Direct	Indirect	Direct	Indirect	Direct		Indirect	
	Country FE	Index		Benefit		Index	Benefit	Index	Benefit
Sample 1	NO	+S	+S	+I	-I	+S	-S (10%)	+S	-S
	YES	-I	+I	-I	-S	-I	-I	+I	-S
Sample 2	NO	+S	+S	+I	+I	+S	-I	+S	-S(10%)
	YES	-I	+I	-I	-S	-I	-I	+I	-S
Sample 3	NO	+S	+S	+S	+S (10%)	+S	-I	+S	-I
	YES	-I	+I	-I	-S	-I	-I	+I	-S

Consider the outcomes of the experiments for the index first. For the case of no country fixed effects the coefficients are positive and significant, yet when suppressing the heterogeneity the measurement of the index becomes insignificant. The reason for this occurrence may be attributed to the fact that the index is constructed by country specific observations. The correlation between country size and index may not have become obsolete despite it being blended by the destination's value and the values of the index assigned to any hub country interspersed between the origin and destination: the overall impact may be absorbed by the fixed effect. Provided that this is not the case, we can then conclude that the index appears to have a moderately positive effect on trade flows for both distance regimes.

Consider now the outcomes of the benefit term. The results here become even more ambiguous but for the majority of experiments they are insignificant: For the direct distance case, all cases are insignificant. For the indirect distance case, without country fixed effects, all cases bar one are insignificant. In the presence of country fixed effects, the impact is significantly negative compared to the case of the index, which leads me to the conclusion that the benefit term does not act as the theoretical model states. The reasons for this behaviour may be absorption by the country fixed effects, or mispecification in the sense that the index does not capture the beneficial measure of forming a link or the model.

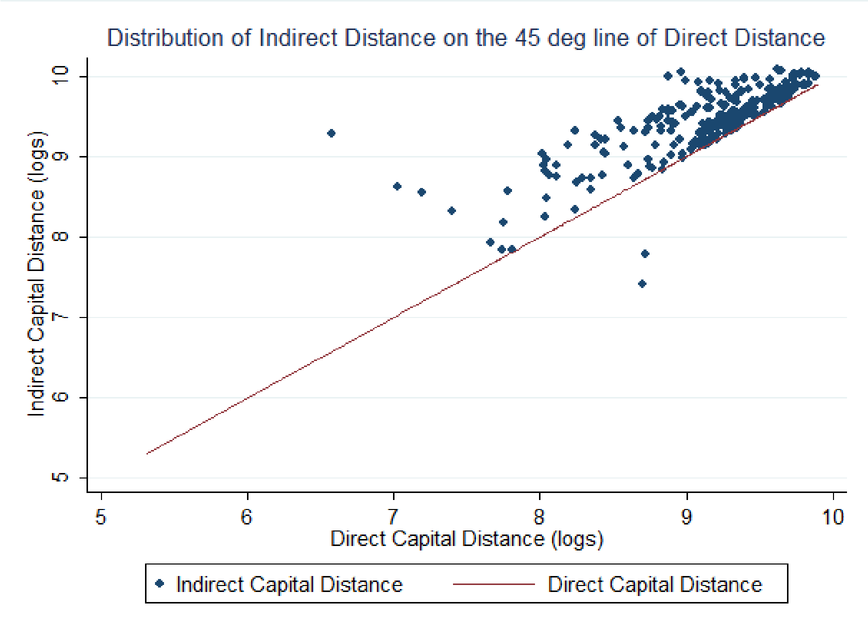
However, in the event that the benefit term is actually picking up the effect, it becomes plausible that it is capturing the decreasing impact of the benefit term when there is an increase in the number of links. The term d_{ij} ranges between 0 and 1 and an increase in the number of links decreases the beneficial impact. Then the number of links acts like an interaction with the index. This is loosely captured in the last columns of Table D-III, where without country fixed effects, there is a discounted benefit as the number of links increases that reduce the positive impact of connectivity, i.e. the index. Yet under country fixed effects, this argument fails.

Relative to Specifications 1 and 2, there is evidence in favour of Specification 1. Then it is proved that an increase in distance has a negative impactive to trade, which can be lessened by increasing the number of hubs, or indirect connections, to the destination. This applies to the indirect distance regime severely, and with lesser magnitude to the direct distance case, as one should expect.

Relative to Specification 3, there is very weak evidence, if any, that the theoretical term $\delta_{ij}^{t_{ij}}$ can be estimated using the selected data. This can be attributed to measurement error, mispecification or that the fixed effects are picking up the impact and thus cannot yield conclusive inference.

Lastly, there is a positive relation between the number of hubs on a route and the volume of trade. If two trading partners choose to trade indirectly, their costs increase however the presence of the hub seems to play a dampening role. Hence the proposition is confirmed; the first part of the empirical prediction is also confirmed due to the equivalent behaviour of indirect distance in the presence of the number of hubs and the hub indicator.

E Data Details



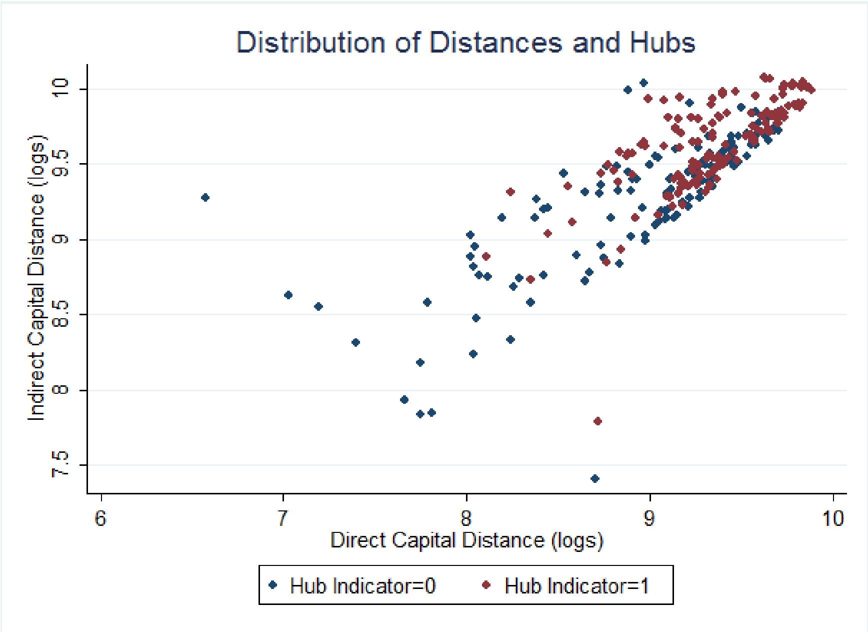


Table E-I: Industrial Composition of Commodities.

Agriculture		Crude oils		Raw Materials		Manufactures	
Commodity	Transport Mode	Commodity	Transport Mode	Commodity	Transport Mode	Commodity	Transport Mode
01: Live animals	Containers	27: Mineral fuels, oils, distillation products, etc	Dirty bulk	25: Salt, sulphur, earth, stone, plaster, lime and cement	Dirty bulk	15: Animal,vegetable fats and oils, cleavage products, et	Dirty bulk
02: Meat and edible meat offal	Containers	27: Mineral fuels, oils, distillation products, etc	Tankers	26: Ores, slag and ash	Dirty bulk	15: Animal,vegetable fats and oils, cleavage products, et	Containers
03: Fish, crustaceans, molluscs, aquatic invertebrates ne	Containers			72: Iron and steel	Dirty bulk	28: Inorganic chemicals, precious metal compound, isotope	Dirty bulk
04: Dairy products, eggs, honey, edible animal product ne	Containers					29: Organic chemicals	Dirty bulk
05: Products of animal origin, nes	Containers					30: Pharmaceutical products	Containers
06: Live trees, plants, bulbs, roots, cut flowers etc	Containers					31: Fertilizers	Dirty bulk
07: Edible vegetables and certain roots and tubers	Containers					32: Tanning, dyeing extracts, tannins, derivs,pigments et	Containers
08: Edible fruit, nuts, peel of citrus fruit, melons	Containers					33: Essential oils, perfumes, cosmetics, toiletries	Containers
09: Coffee, tea, mate and spices	Containers					34: Soaps, lubricants, waxes, candles, modelling pastes	Containers
10: Cereals	Clean bulk					35: Albuminoids, modified starches, glues, enzymes	Containers
11: Milling products, malt, starches, inulin, wheat glute	Containers					36: Explosives, pyrotechnics, matches, pyrophorics, etc	Containers
12: Oil seed, oleagic fruits, grain, seed, fruit, etc, ne	Clean bulk					37: Photographic or cinematographic goods	Containers
12: Oil seed, oleagic fruits, grain, seed, fruit, etc, ne	Containers					38: Miscellaneous chemical products	Containers
13: Lac, gums, resins, vegetable saps and extracts nes	Containers					39: Plastics and articles thereof	Containers
14: Vegetable plaiting materials, vegetable products nes	Containers					40: Rubber and articles thereof	Containers
15: Animal,vegetable fats and oils, cleavage products, et	Tankers					41: Raw hides and skins (other than furskins) and leather	Containers
15: Animal,vegetable fats and oils, cleavage products, et	Containers					42: Articles of leather, animal gut, harness, travel good	Containers
16: Meat, fish and seafood food preparations nes	Containers					43: Furskins and artificial fur, manufactures thereof	Containers
17: Sugars and sugar confectionery	Containers					44: Wood and articles of wood, wood charcoal	Containers
18: Cocoa and cocoa preparations	Containers					45: Cork and articles of cork	Containers
19: Cereal, flour, starch, milk preparations and products	Containers					46: Manufactures of plaiting material, basketwork, etc.	Containers
20: Vegetable, fruit, nut, etc food preparations	Containers					47: Pulp of wood, fibrous cellulosic material, waste etc	Containers
21: Miscellaneous edible preparations	Containers					48: Paper & paperboard, articles of pulp, paper and board	Containers
22: Beverages, spirits and vinegar	Containers					49: Printed books, newspapers, pictures etc	Containers
23: Residues, wastes of food industry, animal fodder	Containers					50: Silk	Containers
24: Tobacco and manufactured tobacco substitutes	Containers					51: Wool, animal hair, horsehair yarn and fabric thereof	Containers
						52: Cotton	Containers
						53: Vegetable textile fibres nes, paper yarn, woven fabri	Containers
						54: manmade filaments	Containers
						55: manmade staple fibres	Containers
						56: Wadding, felt, nonwovens, yarns, twine, cordage, etc	Containers
						57: Carpets and other textile floor coverings	Containers
						58: Special woven or tufted fabric, lace, tapestry etc	Containers
						59: Impregnated, coated or laminated textile fabric	Containers
						60: Knitted or crocheted fabric	Containers
						61: Articles of apparel, accessories, knit or crochet	Containers
						62: Articles of apparel, accessories, not knit or crochet	Containers
						63: Other made textile articles, sets, worn clothing etc	Containers
						64: Footwear, gaiters and the like, parts thereof	Containers
						65: Headgear and parts thereof	Containers
						66: Umbrellas, walking-sticks, seat-sticks, whips, etc	Containers
						67: Bird skin, feathers, artificial flowers, human hair	Containers
						68: Stone, plaster, cement, asbestos, mica, etc articles	Containers
						69: Ceramic products	Containers
						70: Glass and glassware	Containers
						71: Pearls, precious stones, metals, coins, etc	Containers
						73: Articles of iron or steel	Containers
						74: Copper and articles thereof	Containers
						75: Nickel and articles thereof	Containers
						76: Aluminium and articles thereof	Containers
						78: Lead and articles thereof	Containers
						79: Zinc and articles thereof	Containers
						80: Tin and articles thereof	Containers
						81: Other base metals, cermets, articles thereof	Containers
						82: Tools, implements, cutlery, etc of base metal	Containers
						83: Miscellaneous articles of base metal	Containers
						84: Nuclear reactors, boilers, machinery, etc	Containers
						85: Electrical, electronic equipment	Containers
						86: Railway, tramway locomotives, rolling stock, equipmen	Containers
						87: Vehicles other than railway, tramway	Containers
						88: Aircraft, spacecraft, and parts thereof	Containers
						89: Ships, boats and other floating structures	Containers
						90: Optical, photo, technical, medical, etc apparatus	Containers
						91: Clocks and watches and parts thereof	Containers
						92: Musical instruments, parts and accessories	Containers
						93: Arms and ammunition, parts and accessories thereof	Containers
						94: Furniture, lighting, signs, prefabricated buildings	Containers
						95: Toys, games, sports requisites	Containers
						96: Miscellaneous manufactured articles	Containers
						97: Works of art, collectors pieces and antiques	Containers

Table E-II: Trade Partnerships with finite port distance.

Exporter	Importer	Exporter	Importer	Exporter	Importer	Exporter	Importer	Exporter	Importer	Exporter	Importer
ARE	AUS	CAN	MYS	ECU	USA	ISL	ARG	MYS	USA	THA	USA
ARE	EU15	CAN	NZL	EGY	AUS	ISL	AUS	NGA	AUS	TUN	AUS
ARE	IND	CAN	PAK	EGY	USA	ISL	BRA	NZL	ARG	TUN	USA
ARE	NZL	CAN	PER	EU15	ARE	ISL	COL	NZL	AUS	TUR	ARG
ARE	USA	CAN	RUS	EU15	ARG	ISL	ECU	NZL	BRA	TUR	AUS
ARG	AUS	CAN	SAU	EU15	AUS	ISL	NZL	NZL	COL	TUR	BRA
ARG	COL	CAN	TUN	EU15	BGD	ISL	PER	NZL	ECU	TUR	COL
ARG	ECU	CAN	TWN	EU15	BRA	ISL	URY	NZL	PER	TUR	ECU
ARG	EU15	CAN	URY	EU15	CHN	ISL	USA	NZL	URY	TUR	NZL
ARG	MEX	CAN	VEN	EU15	COL	ISR	ARG	NZL	USA	TUR	PER
ARG	NZL	CAN	ZAF	EU15	ECU	ISR	AUS	PER	AUS	TUR	URY
ARG	PER	CHE	ARG	EU15	EGY	ISR	BRA	PER	NZL	TUR	USA
ARG	USA	CHE	AUS	EU15	IDN	ISR	COL	PER	USA	TWN	AUS
AUS	ARG	CHE	BRA	EU15	IND	ISR	ECU	PHL	AUS	UKR	AUS
AUS	BGD	CHE	COL	EU15	JOR	ISR	NZL	PHL	EU15	UKR	USA
AUS	BRA	CHE	ECU	EU15	JPN	ISR	PER	PHL	USA	URY	AUS
AUS	CHN	CHE	NZL	EU15	KOR	ISR	URY	POL	ARG	URY	NZL
AUS	COL	CHE	PER	EU15	LBY	ISR	USA	POL	AUS	URY	USA
AUS	ECU	CHE	URY	EU15	MAR	JPN	ARG	POL	BRA	USA	ARG
AUS	EGY	CHE	USA	EU15	MYS	JPN	AUS	POL	COL	USA	AUS
AUS	IDN	CHL	AUS	EU15	NZL	JPN	BRA	POL	ECU	USA	BRA
AUS	JPN	CHL	BRA	EU15	PER	JPN	COL	POL	NZL	USA	CHN
AUS	KOR	CHL	COL	EU15	PHL	JPN	ECU	POL	PER	USA	COL
AUS	LKA	CHL	ECU	EU15	SAU	JPN	EU15	POL	URY	USA	ECU
AUS	NZL	CHL	NZL	EU15	SDN	JPN	NZL	POL	USA	USA	EGY
AUS	PAK	CHL	URY	EU15	SGP	JPN	PER	RUS	AUS	USA	EU15
AUS	PER	CHL	USA	EU15	THA	JPN	URY	RUS	USA	USA	IDN
AUS	SAU	CHN	ARG	EU15	TUN	JPN	USA	SAU	AUS	USA	IND
AUS	TWN	CHN	AUS	EU15	URY	KAZ	AUS	SAU	EU15	USA	JPN
AUS	URY	CHN	BRA	EU15	USA	KOR	ARG	SAU	USA	USA	KOR
AUS	USA	CHN	COL	EU15	VNM	KOR	AUS	SGP	AUS	USA	LKA
AUS	ZAF	CHN	ECU	EU15	YEM	KOR	BRA	SGP	EU15	USA	MAR
BGD	AUS	CHN	EU15	GHA	AUS	KOR	COL	SGP	NZL	USA	MYS
BGD	USA	CHN	IND	HUN	ARG	KOR	ECU	SGP	USA	USA	NZL
BGR	USA	CHN	NZL	HUN	AUS	KOR	NZL	SVK	ARG	USA	PAK
BHR	AUS	CHN	PER	HUN	BRA	KOR	PER	SVK	AUS	USA	PER
BOL	AUS	CHN	URY	HUN	COL	KOR	URY	SVK	BRA	USA	PHL
BOL	NZL	CHN	USA	HUN	ECU	KOR	USA	SVK	COL	USA	RUS
BOL	USA	COL	AUS	HUN	NZL	KWT	AUS	SVK	ECU	USA	SAU
BRA	EU15	COL	NZL	HUN	PER	LBN	AUS	SVK	NZL	USA	SGP
CAN	ARG	COL	USA	HUN	URY	LKA	AUS	SVK	PER	USA	THA
CAN	AUS	CRI	AUS	HUN	USA	LKA	USA	SVK	URY	USA	TUN
CAN	BRA	CRI	COL	IDN	AUS	LTU	AUS	SVK	USA	USA	URY
CAN	CHN	CRI	ECU	IDN	EU15	MAR	AUS	SVN	ARG	USA	VEN
CAN	COL	CZE	ARG	IDN	USA	MAR	USA	SVN	AUS	USA	VNM
CAN	DZA	CZE	AUS	IND	AUS	MEX	ARG	SVN	BRA	USA	YEM
CAN	ECU	CZE	BRA	IND	BRA	MEX	AUS	SVN	COL	USA	ZAF
CAN	EGY	CZE	COL	IND	CHN	MEX	BRA	SVN	ECU	VEN	AUS
CAN	EU15	CZE	ECU	IND	COL	MEX	COL	SVN	NZL	VEN	USA
CAN	IDN	CZE	NZL	IND	ECU	MEX	ECU	SVN	PER	VNM	AUS
CAN	IRN	CZE	PER	IND	EU15	MEX	NZL	SVN	URY	VNM	EU15
CAN	JOR	CZE	URY	IND	NZL	MEX	PER	SVN	USA	VNM	USA
CAN	JPN	CZE	USA	IND	PER	MEX	URY	THA	ARG	YEM	USA
CAN	KOR	DOM	USA	IND	SAU	MYS	AUS	THA	AUS	ZAF	ARG
CAN	LKA	DZA	USA	IND	URY	MYS	BRA	THA	BRA	ZAF	AUS
CAN	MAR	ECU	AUS	IND	USA	MYS	EU15	THA	EU15	ZAF	NZL
CAN	MEX	ECU	NZL	IRN	AUS	MYS	NZL	THA	NZL	ZAF	USA

References

Anderson, J. E. (2010), The gravity model, NBER Working Papers 16576, National Bureau of Economic Research.

Anderson, J. E. & van Wincoop, E. (2004), 'Trade costs', *Journal of Economic Literature* **42**(3), 691–751.

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

Chaney, T. (2011), The network structure of international trade, NBER Working Papers 16753, National Bureau of Economic Research.

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

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

Helpman, E., Melitz, M. & Rubinstein, Y. (2008), 'Estimating trade flows: Trading partners and trading volumes', *The Quarterly Journal of Economics* **123**(2), 441–487.

Hendricks, K., Piccione, M. & Tan, G. (1995), 'The economics of hubs: The case of monopoly', *The Review of Economic Studies* **62**(No. 1), 83–99.

Hummels, D. (2007), 'Transportation costs and international trade in the second era of globalization', *Journal of Economic Perspectives* **21**(3), 131–154.

Hummels, D., Lugovskyy, V. & Skiba, A. (2009), 'The trade reducing effects of market power in international shipping', *Journal of Development Economics* **89**(1), 84–97.

Hummels, D. & Skiba, A. (2004), 'Shipping the good apples out? an empirical confirmation of the alchian-allen conjecture', *Journal of Political Economy* **112**(6), 1384–1402.

Irarrazabal, A., Moxnes, A. & Opromolla, L. (2011), The tip of the iceberg: A quantitative framework for estimating trade costs. Banco de Portugal, Economics and Research Department, Working Paper, No. 25/2011.

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. & Wolinsky, A. (1996), 'A strategic model of social and economic networks', *Journal of Economic Theory* **71**(1), 44–74.

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

Krugman, P. (1991), 'Increasing returns and economic geography', *Journal of Political Economy* **99**(3), 483–499.

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.

Krugman, P. R. (1979), 'Increasing returns, monopolistic competition, and international trade', *Journal of International Economics* **9**(4), 469–479.

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

Limao, N. & Venables, A. (2001), 'Infrastructure, geographical disadvantage and transport costs.', *The World Bank Economic Review* **15**, 451–479.

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

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

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

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

UNCTAD (2011), 'Review of maritime transport'.