Roadway Density, Input Sourcing, and Patterns of Specialisation

Esteban Jaimovich*

Preliminary and incomplete

Abstract

This paper argues that the density of the internal transport network represents a key factor in shaping comparative advantage and specialisation. We propose a model where a denser transport network facilitates the sourcing of local inputs from different locations. Hence, efficient internal transportation becomes instrumental to the development of industries whose production processes rely on a wide input base. This grants countries with denser transport networks a comparative advantage in industries that rely on large variety of inputs. Evidence based on industry-level trade data grants support to the main prediction of the model: countries with denser road networks export relatively more in industries that exhibit wider input bases. We show that this correlation is robust to several possible confounding effects proposed by the literature, such as the impact of institutions on specialisation in complex goods. Furthermore, we show that a similar correlation arises for poorer economies when the density of the local transport network is measured by the density of their waterways, rather than by roadway density.

Keywords: International Trade, Comparative Advantage, Internal Transportation Costs.

JEL Classifications: F11, O18, R12

*University of Surrey. Mailing address: School of Economics, Guildford, GU2 7XH, United Kingdom. Email: e.jaimovich@surrey.ac.uk

1 Introduction

The spatial distribution of economic activities means that transport cost are a major factor influencing countries' output, trade flows and specialisation. Apart from few exceptions, the vast majority of the past trade literature has centred their attention on the cost of shipping goods internationally.[1] However, the evidence at hand suggests that internal transport costs are far from being a secondary component that can be disregarded when confronted with transboundary costs.[2] Furthermore, the effects of internal transport costs on specialisation get magnified by the fact that the local infrastructure differs quite substantially between countries, especially when comparing economies at different stages of development.

Being able to efficiently transport commodities is crucial to keep total costs low. Yet, this does not benefit all goods and sectors to the same degree. Owing to specificities of their physical characteristics and of their production processes, some commodities turn out to be inherently more transport-intensive than others. This means that the efficiency of the local transportation infrastructure may unevenly impact the development of different industries. This paper studies a specific channel by which the internal transport network may shape countries' comparative advantages and specialisation. We argue that one key role of the internal transportation network is that it facilitates the sourcing of intermediate inputs from different locations. This means that industries that demand a large variety of intermediate inputs tend to make more intense use of the network. Efficient commodity transportation becomes thus especially critical for sectors whose production processes combine a wide set of intermediate inputs.

To illustrate this idea, we introduce a simple model with two intermediate inputs and a continuum of final good producers. A larger road network allows cheaper transportation of the intermediate inputs to the location site of final good producers. A key feature of the model is that industries producing final goods differ in terms of the breadth of their intermediate input requirements. In particular, some industries have production functions that are very intensive in only one intermediate input, while others require a more balanced use of the two intermediate inputs. Since transportation of inputs is costly, those industries that require a relatively balanced combination of the intermediate inputs will benefit relatively more (in terms of cost reduction) from a denser road network.

This simple mechanism yields a very clear prediction in terms of specialisation when it is

[1] For a few papers that have incorporated internal transport costs into trade models, see Coşar and Fajgelbaum (2016), Ramondo, Rodriguez-Clare and Saborio-Rodriguez (2016), Redding (2016), Matsuyama (2017).

[2] See, e.g., Limao and Venables (2001), Anderson and Van Wincoop (2004), Hillberry and Hummels (2008), Agnosteva, Anderson and Yotov (2014), Atkin and Donaldson (2015), Donaldson (forthcoming).

incorporated into an international trade model. Countries that enjoy a denser local transport network tend to display a comparative advantage in the goods whose production process requires a relatively balanced mix of the intermediate inputs. This is because these are the industries that tend to make heavier use of the local network to source their inputs. Conversely, countries with underdeveloped transport networks tend to specialise in industries with narrow input bases, which allows them to economise on input sourcing.

After presenting the model we provide evidence consistent with its main prediction. To do so, we proceed as follows. Firstly, we index industries by their degree of input breadth using the information contained in the US input-output matrix. Secondly, we measure the density of the transport networks of countries by the length of their roadway per square kilometer. Finally, we correlate countries specialisation by industries (measured by their total exports at the industry level) with an interaction term between industries' input breadth and countries' roadways density. We find that countries with denser road networks export relatively more in industries that exhibit a wider input base.

The correlation between road density and specialisation in industries with wider input bases may obviously be driven by other mechanisms to the one suggested by our model. We show however that this correlation is robust to the inclusion of a large set of possible confounding covariates. In particular, one important channel related to ours works through institutions, as industries that rely on a wide set of inputs tend to be more dependent on contract enforcement [Levchenko (2007) and Nunn (2007)]. We show that the correlation predicted by our model is still present once we also include the effect of institutions. In that respect, our findings complement the previous studies that have interpreted the degree of input variety as a sign of product complexity, showing that industries with a wide input bases seem also to be strongly reliant on the internal transport network.

One additional question is whether the found correlation can be interpreted at all as evidence of *causation* from road density to specialisation in transport-intensive industries. Roadways are the result of investment choices. Hence, road infrastructure may positively respond to transport needs resulting from patterns of specialisation, reversing thus the direction of causation. Interestingly, we show that an analogous correlation to that one found with road density arises when using *waterways* density as an alternative measure of the depth of the local transport network. Moreover, this correlation is especially strong and significant in the case lower-income countries, which are exactly the types of economies that tend to suffer from thinner road networks. Admittedly, the evidence based on waterways density does not directly address the reverse causation concern. However, since waterways cannot be as easily molded and expanded as a roadway network, they are less sensitive to issues of reverse causation than roadways are.

There is a growing literature studying the impact of the local transport infrastructure on international and intra-regional trade and specialisation. Most of the literature has exploited variation across regions within a specific country. For example, Volpe Martincus and Blyde (2013) study access to foreign markets and international trade in different regions of Chile, Coşar and Demir (2016) for Turkey, and Volpe Martincus, Carballo and Cusolito (2017) for Peru. Donaldson (forthcoming) looked at reductions of price and output distortions across Indian regions after expansions of the local railroad network. Fajgelbaum and Redding (2014) and Coşar and Fajgelbaum (2016) investigate the regional location of export-oriented activities given the local infrastructure in the cases of Argentina and China, respectively. Closer to our main focus of interest, Duranton, Morrow and Turner (2014) and Coşar and Demir (2016) have tried to capture whether there is some effect of road infrastructure on specialisation in transport-intensive activities. Duranton *et al* (2014) show that US cities with more highways tend to produce goods of higher weight per physical unit, while Coşar and Demir (2016) find a similar effect for Turkey after a national road expansion policy. We focus on specific channel whereby the density of the local transport infrastructure may impact specialisation, which rests on the notion that the need to source large variety of inputs makes an industry transport-intensive. Moreover, looking at exports by countries in different industries, we provide evidence consistent with this mechanism at the country level.

Our paper also relates to two recent strands of literature have expanded upon the traditional models of international trade, following Ricardian and Heckscher-Ohlin models. One set of papers have looked at institutions and contract enforcement as a source of comparative advantage in industries producing complex goods that require a large variety in input-specific relationships [Antràs (2005), Acemoglu, Antràs and Helpman (2007), Levchenko (2007), Nunn (2007), Costinot (2009)]. The other one, has delved into the role of financial markets fostering exports in industries that are heavy users of external finance [Beck (2002), Svaleryd and Vlachos (2005), Becker, Chen and Greenberg (2012), Manova (2013)]. Our paper seeks to highlight the importance of the local transport network for industries that need sourcing a large variety of intermediate inputs, and how this can influence trade flows and specialisation.

The rest of the paper is organised as follows. Section 2 introduces the main features of the model in the case of a closed economy. Section 3 extends the model to a two-country setup, and derives the main predictions in terms of comparative advantage and trade flows. Section 4 contrasts the main predictions of the model with the data. Section 5 discusses some endogeneity issues and alternative interpretations of the empirical results. Section 6 concludes.

2 General Setup in a Closed Economy Model

This section presents the environment and main features of our model in the specific case of a closed economy. Starting first by introducing a closed economy proves helpful in two aspects. First, it allows an easier description of the main building blocks of the model. Second, it facilitates the task of providing the main intuition for how the density the transport network may heterogeneously affect the cost of production of goods in different sectors.

2.1 Intermediate and Final Goods Sector

There exists a unit continuum of final goods, indexed by the letter $j \in [0,1]$. All final good markets are perfectly competitive. Final goods are purchased by consumers with preferences given by

$$U = \int_0^1 \ln(y_j)\, dy_j, \tag{1}$$

where y_j denotes the consumed amount of j. There is a mass L of consumers, and that each of them is endowed with one unit of labour which is supplied inelastically for a wage w.[3]

In addition to the set of final goods, there exist two intermediate goods, indexed by $i = 0,1$, which are used as inputs by final good producers. The markets for both intermediate goods are perfectly competitive. Each intermediate good is produced with labour, according to the following linear production functions:

$$X_i = \frac{L_i}{1+\varepsilon_i}, \qquad i = 0,1. \tag{2}$$

In (2), X_i denotes the total amount of intermediate good i produced in the economy, L_i is the total amount of labour used in producing i, and $\varepsilon_i \geq 0$ is a technological parameter determining labour productivity in sector i.

Final goods are produced by combining the two intermediate goods within Cobb-Douglas production functions. Total output of final good $j \in [0,1]$ is given by:

$$Y_j = \frac{1}{\alpha_j^{\alpha_j}(1-\alpha_j)^{1-\alpha_j}}\, X_{0,j}^{1-\alpha_j}\, X_{1,j}^{\alpha_j}, \qquad \text{where } \alpha_j \in [0,1], \tag{3}$$

and $X_{0,j}$ and $X_{1,j}$ denote the amount of intermediate good 0 and 1 used in the production of final good j, respectively.

[3] None of the derivations of this section will actually make explicit use of the assumed utility function (1). We will return to make active use of consumer preferences in the next section, when extend the model to a two-country setup with international trade between them.

The Cobb-Douglas production functions (3) differ across final good sectors in terms of the intensity requirements of each intermediate good. Sectors with a small (resp. large) α_j use input 0 (resp. input 1) relatively more intensively. On the other hand, sectors whose α_j lies in the vicinity of 0.5 tend to use a relatively balanced bundle of both inputs. For the rest of the paper, we will assume that, when considering the whole set of final good producers, the values of α_j are uniformly distributed within the unit interval. Abusing a bit the notation, we can thus henceforth index final goods by their value of $\alpha_j \in [0,1]$.

Perfect competition in final good markets implies that, in equilibrium, each final good j will be sold at a price equal to its marginal cost. Using (3), we can obtain the expression for the marginal cost, which we denote by c_j. Namely,

$$c_j = p_{0,j}^{1-\alpha_j}\, p_{1,j}^{\alpha_j}, \tag{4}$$

where $p_{0,j}$ and $p_{1,j}$ are the prices at which the producer of final good j can purchase each unit of input 0 and 1, respectively.[4]

2.2 Geographic Structure of the Economy

We assume that each intermediate good is produced in a different site, which we refer to as location 0 (for input 0) and location 1 (for input 1). Labour is perfectly mobile across locations at zero cost. Intermediate goods must, however, incur in an iceberg transport cost to be moved around. In particular, we assume that, when the distance between the location of j and that of i is $d_{j,i} \geq 0$, the intermediate good producer i must ship $1 + t\, d_{j,i}$ units of input i in order for the final good producer j to receive one unit of i.

We assume there exists an infrastructure network connecting location 0 and 1. There network comprises two types of pathways. One is a semi-circular path of total length $\pi/2$, which represents the *least* direct path between the two locations. The other one is a road of length $r \in [0,1]$, which shortens the distance between the two locations given by the semi-circular path.[5]

[4] Although we are assuming that intermediate goods are sold in competitive markets, in principle, our model will not always lead to the same price paid by each final good producer j for each of the inputs. The reason for such price disparities will be that both $p_{0,j}$ and $p_{1,j}$ will also incorporate internal transport costs, and these costs may well differ across final good producers given their location and the locations of intermediate goods.

[5] Conceptually, we intend to represent the notion that the economy has a road network of length r which allows faster transportation of inputs across location 0 and 1, relative to the semi-circular path of length $\pi/2$. Of course, nothing precludes the fact that the road network could comprise several segments, whose lengths sum up to r. In Appendix #, we show that this would not be optimal. More precisely, given a total length of

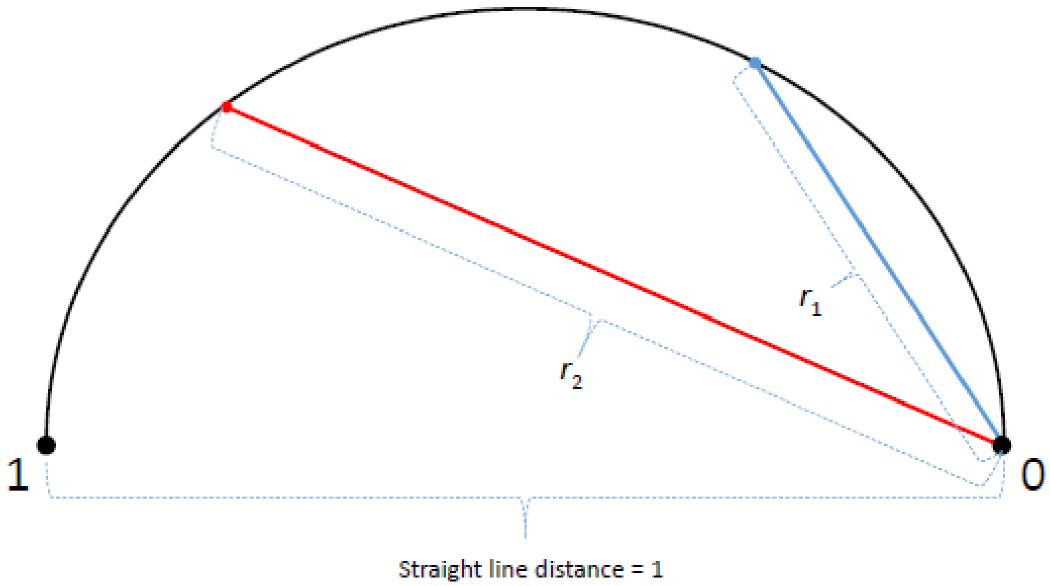


Figure 1: Geographic Structure of the Economy

Figure 1 plots the geographic structure of the economy for two different roads lengths, namely $0 < r_1 < r_2 < 1$. Although not plotted in the graph, the two extreme cases $r = 0$ and $r = 1$ would correspond, respectively, to the case where the only path available from location 0 to location 1 is via the semi-circular arch of length $\pi/2$, and the case where the two locations are connected by a straight horizontal line of length one.[6]

Henceforth, we denote by $\varphi(r)$ the *shortest* distance between location 0 and 1, given a road network of length $r \in [0, 1]$. Appendix A shows that $\varphi(r)$ is given by the following expression

$$\varphi(r) \equiv \frac{\pi}{2} + r - \arcsin(r). \tag{5}$$

From (5), it is straightforward to observe that $\varphi'(r) < 0$ and $\varphi''(r) < 0$.

road network equal to $r \in [0, 1]$, the shortest distance to connect inputs locations is achieved by building one single straight line of length r. Note, however, that this segment could, in principle, connect *any* two points within the semi-circular path. Without any loss of generality, we will arbitrarily place the segment as starting from location 0.

[6]Note that the iceberg cost $1 + td_{j,i}$ implicitly assumes that the transport cost per unit of distance is equal to t, regardless of whether intermediate goods are shipped via the semi-circular path, or via a combination between such path and the straight road. This assumption is posed just in the sake of algebraic simplicity, and none of the results would be altered if we instead assumed that the cost per distance is smaller when using the road than when using the semi-circular path.

2.3 Location Choice by Final Good Producers

The previous subsection assumed that each intermediate good is produced in a specific and exogenously given location. With regards to final goods producers, we assume that they can freely choose a location in any point of the network depicted by Figure 1. Given that shipping inputs across production sites entails a transport cost, final good producers will choose their own location so as to minimise their marginal costs (c_j).

Recall that, given a road network of length $r \in [0,1]$, the minimum distance between location 0 and 1 is given by $\varphi(r)$. Let $l_j\varphi(r)$ denote the (minimum) distance between the location chosen by producer j and location 0, where $l_j \in [0,1]$. (Naturally, $l_j = 0$ means that j selects location 0, while $l_j = 1$ means that j chooses location 1.) Producer j must thus pay

$$p_{0,j} = [1 + l_j\varphi(r)t]\,(1+\varepsilon_0)\,w$$

for each unit of input 0 that he purchases, while he must pay

$$p_{1,j} = [1 + (1-l_j)\,\varphi(r)t]\,(1+\varepsilon_1)\,w$$

for each purchased unit of input 1.

Bearing in mind (4), producer j will choose his location by solving the following minimisation problem:

$$\min_{l_j \in [0,1]} : c_j(l_j) = \left[(1 + l_j\varphi(r)t)\,(1+\varepsilon_0)\,w\right]^{1-\alpha_j} \left[(1 + (1-l_j)\,\varphi(r)t)\,(1+\varepsilon_1)\,w\right]^{\alpha_j}. \tag{6}$$

The above problem yields corner solutions (note the second derivative of the right-hand side of (6) is positive). Comparing thus $c_j(0)$ vis-a-vis $c_j(1)$, we obtain

$$l_j^* = \begin{cases} 0 & \text{if } \alpha_j \leq 0.5 \\ 1 & \text{if } \alpha_j \geq 0.5 \end{cases} \tag{7}$$

The result in (7) is quite intuitive: final producers choose to locate their firm in the same place where the input they use more intensively is being produced.[7]

Finally, plugging (7) back into the expression in the right-hand side of (6) we can obtain the marginal cost of final good j, namely:

$$c_j^* = \begin{cases} (1+\varphi(r)t)^{\alpha_j}\,(1+\varepsilon_0)^{1-\alpha_j}\,(1+\varepsilon_1)^{\alpha_j}\,w & \text{if} \quad \alpha_j \leq 0.5 \\ (1+\varphi(r)t)^{1-\alpha_j}\,(1+\varepsilon_0)^{1-\alpha_j}\,(1+\varepsilon_1)^{\alpha_j}\,w & \text{if} \quad \alpha_j \geq 0.5 \end{cases} \tag{8}$$

[7] In the case where $\alpha_j = 0.5$ (that is, in the case in which both inputs are used with identical intensity), producer j is indifferent between location 0 and 1.

The expression in (8) shows that the marginal cost of final good j is determined by the labour cost of producing the required inputs (via the wage w, and the parameters ε_0 and ε_1), and *also* by the transport cost of involved in sourcing those inputs. Importantly, recall that final good producers will optimally choose to set up their firms in the *same* location where the input they use more intensively is being produced. As a result, the transport cost ends up *only* being applied to the input whose Cobb-Douglas weight in (3) is smaller than 0.5. In turn, this implies that internal transport costs tend to affect more severely the marginal cost of those final goods whose α_j lies near 0.5. In other words, internal transport costs tend to particularly hurt sectors which use a relatively even combination of inputs. On the other hand, this also implies that while improvements in transport infrastructure will lower the cost of production of all final goods (except for the extreme cases where either $\alpha_j = 0$ or $\alpha_j = 1$), such improvements will end up lowering the marginal cost of goods whose α_j is closer to 0.5 by relatively more.

Figure 2 plots the marginal cost of production for the whole set of final goods, under two different values of r, namely $r_1 < r_2$. To depict more cleanly the (differential) effects of transport cost and r on c_j^*, the figure implicitly assumes that $\varepsilon_0 = \varepsilon_1 = 0$ (i.e., we remove any differential effect of sectoral labour productivities on the marginal cost of each final good j). As it can be readily observed from Figure 2, c_j^* is increasing in α_j for $\alpha_j < 0.5$, while it is decreasing in it for $\alpha_j > 0.5$. Also, the figure shows that $c_j^*(r_2) < c_j^*(r_1)$ for all $\alpha_j \in (0, 1)$. More importantly, the figure shows that the difference $c_j^*(r_1) - c_j^*(r_2)$ is increasing in α_j for $\alpha_j < 0.5$ and decreasing in it for $\alpha_j > 0.5$, while it is largest when α_j equals one half.

3 Two-Country Model

We consider now a world economy a lá Dornbusch-Fischer-Samuelson (1977) with two countries: H and F. Both countries are populated by a mass L of individuals. Each individual is endowed with one unit of labour that is supplied inelastically in the local labour market. We let w_H and w_F denote the wage in H and F, respectively. Henceforth, we set $w_F = 1$ (i.e., we set w_F as the *numeraire*), and use $\omega \equiv w_H/w_F$ to denote the relative wage. All individuals share the same preferences –given by (1)– over the unit continuum of final goods.

Each final good could in principle be produced by any of the two countries. The technologies to produce final goods are identical in both H and F, given by the Cobb-Douglas functions (3). All final goods markets are perfectly competitive. In addition, we assume that all final goods are internationally tradeable, subject to an iceberg cost $\tau > 0$ (that is, when $1 + \tau$ units of j are shipped internationally, only 1 unit of j will arrive at the destination country).

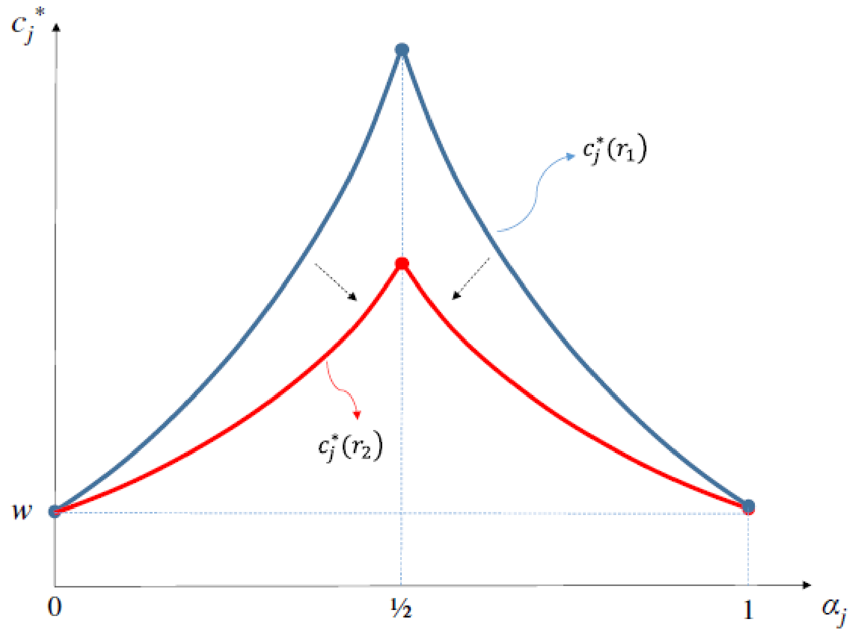


Figure 2: Marginal Cost of Final Goods

Unlike for final goods, we assume that intermediate goods are non-tradeable internationally. We also assume that the technologies to produce the intermediate goods differ between H and F. Letting $X_{i,c}$ denote the total amount of intermediate good i produced in country c, we assume that in H

$$X_{0,H} = L_{0,H} \quad \text{and} \quad X_{1,H} = \frac{L_{1,H}}{1+\varepsilon}, \tag{9}$$

while in F,

$$X_{0,F} = \frac{L_{0,F}}{1+\varepsilon} \quad \text{and} \quad X_{1,F} = L_{1,F}, \tag{10}$$

where $L_{i,c}$ is the total amount of labour used in producing input i in country c, and $\varepsilon > 0$. The intermediate goods markets are perfectly competitive both in H and in F.

Two features implied by (9) and (10), coupled with the final goods production functions (3), are worth stressing here. First, since they imply that H is relatively more productive than F in the intermediate sector 0, they tend to yield a comparative advantage by H on the final goods that rely more heavily on input 0 (that is, on those j whose α_j is small). Second, since (9) and (10) exactly mirror one another, they implicitly assume away any aggregate absolute advantage by one country over the other one stemming from the distribution of sectoral labour productivities.[8]

[8]None of the main results of the model crucially depend on this last feature. In fact, the model could be easily generalised to encompass intermediate production functions $X_{i,c} = L_{i,c}/(1+\varepsilon_{i,c})$, where, $i = 1, 2$, $c = H, F$ and

Analogously to the closed economy setup in Section 2.2, we assume that each input is produced in a specific location. We keep referring as location 0 to the production site of input 0, and as location 1 to that one of input 1. (In this case, there is one such location in each of the countries.) Also like in the closed economy setup, we assume that (both in H and F) location 0 and 1 are connected by a semi-circular path of length $\pi/2$. In addition to this path, each country has a road network, denoted by $r_H \in [0,1]$ and $r_F \in [0,1]$ for H and F, respectively, which allows shortening the distance between (their) location 0 and 1. Road networks are built such that they minimise internal travel cost; hence, they consist of single straight lines of length r_H and r_F, respectively. Without any loss of generality, we will assume that both r_H and r_F start from location 0 in each of the countries. We also assume that the iceberg cost t per unit of distance $d_{j,i}$ traveled by input i to reach producer j is identical in H and F.

Henceforth, we assume that H enjoys a longer road network than F.

Assumption 1 $\quad r_L < r_H$.

In our model, Assumption 1 will tend to convey a source of comparative advantage to H in the types of goods that depend on (internal) transport of inputs more strongly. In addition, $r_H > r_L$ also implies that H can, in general, ship inputs internally at lower cost than F. This fact will, in turn, grant a source of aggregate absolute advantage by H over F.

3.1 Pricing of Final Goods in *H* and *F*

The fact that all good markets in H and F are perfectly competitive implies again that final goods will be sold at their marginal costs. Notice that this will include both the incurred internal and international transport costs. In its general form, the price of final good $j \in [0,1]$ produced in country $c = H, F$ and sold in country $m = H, F$ will be given by

$$P_{j,c}^{m} = (1 + \tau \cdot 1\{m \neq c\}) \left[(1 + l_{j,c}\,\varphi(r_c)t)\,(1 + \varepsilon_{0,c})\right]^{1-\alpha_j} \left[(1 + (1 - l_{j,c})\,\varphi(r_c)t)\,(1 + \varepsilon_{1,c})\right]^{\alpha_j} w_c, \tag{11}$$

where: *i)* $1\{m \neq c\}$ is an index function that is equal to one when $m \neq c$, and zero otherwise; *ii)* $\varepsilon_{0,H} = \varepsilon_{1,F} = 0$ and $\varepsilon_{1,H} = \varepsilon_{0,F} = \varepsilon$; *iii)* $l_{j,c}\,\varphi(r_c)$, where $l_{j,c} \in [0,1]$, is the (minimum) distance between producer j in country c and location 0.

Final good producers will optimally seek to minimise their marginal costs. Analogously as done in Section 2.3, it can be proved that this is achieved by setting up firm j in location 0

$\varepsilon_{i,c} \geq 0$. We deliberately choose a symmetric distribution of heterogeneous labour productivities, as featured by (9) and (10), only because this allows depicting the influence of road networks on the patterns of comparative advantage across H and F more cleanly.

when $\alpha_j \leq 0.5$, and setting it up in location 1 when $\alpha_j \geq 0.5$. That is, condition (7) still holds true within the two-country model, with $l_{j,c} = l_j^*$ for $c = H, F$.

By using this result, together with (11), the price of good j when produced in country H and sold in $m = H, F$, denoted by $P_{j,H}^m$, can be written as

$$P_{j,H}^m = \begin{cases} (1+\tau \cdot 1\{m \neq H\})\ (1+\varphi(r_H)t)^{\alpha_j}\ (1+\varepsilon)^{\alpha_j}\ \omega & \text{if} \quad \alpha_j \leq 0.5 \\ (1+\tau \cdot 1\{m \neq H\})\ (1+\varphi(r_H)t)^{1-\alpha_j}\ (1+\varepsilon)^{\alpha_j}\ \omega & \text{if} \quad \alpha_j \geq 0.5 \end{cases}. \tag{12}$$

Analogously, $P_{j,F}^m$, with $m = H, F$, can be written as

$$P_{j,F}^m = \begin{cases} (1+\tau \cdot 1\{m \neq F\})\ (1+\varphi(r_F)t)^{\alpha_j}\ (1+\varepsilon)^{1-\alpha_j} & \text{if} \quad \alpha_j \leq 0.5 \\ (1+\tau \cdot 1\{m \neq F\})\ (1+\varphi(r_F)t)^{1-\alpha_j}\ (1+\varepsilon)^{1-\alpha_j} & \text{if} \quad \alpha_j \geq 0.5 \end{cases}. \tag{13}$$

To ease notation, it proves convenient to define

$$\gamma \equiv \frac{1+\varphi(r_H)t}{1+\varphi(r_L)t}. \tag{14}$$

Notice that $\gamma \in (0,1)$, since $r_H > r_L$. In the context of our model, γ^{-1} can be interpreted as a measure of the advantage of H over F in terms of length of road network.[9]

3.2 Traded (and Non-Traded) Goods

In equilibrium, consumers will buy each final good j from the producer who can offer j at the lowest price. In some cases this will mean that consumers will source good j locally, while in others they will choose to import it. Naturally, given that shipping final goods internationally entails an iceberg cost $\tau > 0$, if in equilibrium country c is an exporter of good j, then it must also be the case that individuals from c must be buying good j from local producers.

By comparing (12) vis-a-vis (13), we can observe that international trade of final goods takes place when the following conditions hold true (henceforth, without any loss of generality, we assume that when confronted with identical prices, consumers always buy from local producers).

- H will export final good j to F if and only if:

$$\begin{aligned} \omega &< \frac{1}{1+\tau}\frac{(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{\alpha_j}} \qquad \text{when } \alpha_j \leq 0.5, \\ \omega &< \frac{1}{1+\tau}\frac{(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{1-\alpha_j}} \qquad \text{when } \alpha_j \geq 0.5 \end{aligned} \tag{15}$$

[9] Smaller values of γ reflect larger disparities in terms of the road network length between H and F, while values of γ close to one are the result of r_H and r_L being very similar.

- H will import final good j from F if and only if:

$$\begin{aligned} \omega &> (1+\tau)\frac{(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{\alpha_j}} \qquad \text{when } \alpha_j \leq 0.5, \\ \omega &> (1+\tau)\frac{(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{1-\alpha_j}} \qquad \text{when } \alpha_j \geq 0.5 \end{aligned} \tag{16}$$

The presence of $\tau > 0$ in (15) and (16) implies that some final goods may end up *not* being traded internationally. In particular, if for some subset of final goods whose $0 \leq \alpha_j \leq 0.5$, the model yields $(1+\tau)^{-1} \leq \omega\gamma^{\alpha_j}(1+\varepsilon)^{2\alpha_j-1} \leq (1+\tau)$, then consumers from both H and F will end up sourcing these goods locally. Similarly, if for some subset of final goods whose $0.5 \leq \alpha_j \leq 1$, the model yields $(1+\tau)^{-1} \leq \omega\gamma^{1-\alpha_j}(1+\varepsilon)^{2\alpha_j-1} \leq (1+\tau)$, these goods will also be sourced in both H and F from local producers.

3.3 Equilibrium and Patterns of Specialisation

In equilibrium, the total (world) spending on final goods produced in each country must equal the total labour income of each country. In our two-country setup, this condition can be restated as a trade balance equilibrium for either H or F. Namely, in equilibrium, the total value of exports by H must be equal to the total value of imports by H.

The utility function (1) implies that consumers allocate identical expenditure shares across all final goods in the optimum.[10] Hence, in our model, the equilibrium condition in the world economy boils down to:

$$\begin{aligned} &\mu_{\alpha_j \leq 0.5}\left(\alpha_j \middle| \omega < \tfrac{(1+\varepsilon)^{1-2\alpha_j}}{(1+\tau)\gamma^{\alpha_j}}\right) + \mu_{\alpha_j > 0.5}\left(\alpha_j \middle| \omega < \tfrac{(1+\varepsilon)^{1-2\alpha_j}}{(1+\tau)\gamma^{1-\alpha_j}}\right) = \\ &\left[\mu_{\alpha_j \leq 0.5}\left(\alpha_j \middle| \omega > \tfrac{(1+\tau)(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{\alpha_j}}\right) + \mu_{\alpha_j > 0.5}\left(\alpha_j \middle| \omega > \tfrac{(1+\tau)(1+\varepsilon)^{1-2\alpha_j}}{\gamma^{1-\alpha_j}}\right)\right]\omega. \end{aligned} \tag{17}$$

In (17), the expressions $\mu_{\alpha_j \leq 0.5}(\alpha_j | \, \omega \lesseqgtr A)$ [resp. $\mu_{\alpha_j > 0.5}(\alpha_j | \, \omega \lesseqgtr A)$] denote the mass of final sectors j whose $\alpha_j \leq 0.5$ [resp. $\alpha_j > 0.5$] for which the condition $\omega \lesseqgtr A$ holds true. The left-hand side of (17) thus amounts to the total value of H's exports, whereas its right-hand side equals the total value of H's imports.

Henceforth, we impose an additional parametric restriction to the model. Namely,

Assumption 2 $\varepsilon > \tau$.

[10] All the results in this section can easily be extended to a general Cobb-Douglas utility function with constant (but non-equal) expenditure shares across goods. The specific choice of (1) is just for algebraic simplicity.

Assumption 2 ensures that our model will always feature positive trade in equilibrium. Intuitively, $\varepsilon > \tau$ implies that the source of comparative advantages linked to heterogeneities in sectoral labour productivities –i.e., those determined by (9) and (10)– are strong enough so as not to be completely overturned by international trade costs in all final sectors.[11]

From the trade balance equilibrium condition (17) we can obtain our first result concerning the equilibrium relative wage, ω^*.

Proposition 1 *In equilibrium, the wage in H is strictly greater than in F. That is, $\omega^* > 1$. Furthermore, ω^* is strictly decreasing in γ, and $\omega^* < \min\left\{(1+\tau)\gamma^{-\frac{1}{2}}, (1+\varepsilon)(1+\tau)^{-1}\right\}$ if $(1+\varepsilon)\gamma^{\frac{1}{2}} > 1$, whilst $\omega^* < (1+\tau)^{-1}\gamma^{-\frac{1}{2}}$ if $(1+\varepsilon)\gamma^{\frac{1}{2}} \leq 1$.*

The result $\omega^* > 1$ is a straightforward implication of the fact that, when considering the whole unit-continuum of final goods, Assumption 1 conveys an aggregate advantage of H over F. As a result, in equilibrium, ω must rise above one, in order to allow F to be able to export to H as much as H exports to F. Notice that since labour is the only non-reproducible input in our model, wages are also equal to income per head in each country. Thus, Proposition 1 is ultimately stating that H is richer than F.

For future reference it proves convenient to define now four different thresholds for the values of α_j. Namely,

$$\underline{\alpha}_H \equiv \frac{\ln(1+\varepsilon) - \ln(1+\tau) - \ln(\omega^*)}{2\ln(1+\varepsilon) + \ln(\gamma)} \tag{18}$$

$$\overline{\alpha}_H \equiv \frac{\ln(1+\varepsilon) - \ln(1+\tau) - \ln(\gamma) - \ln(\omega^*)}{2\ln(1+\varepsilon) - \ln(\gamma)} \tag{19}$$

$$\underline{\alpha}_F \equiv \frac{\ln(1+\varepsilon) + \ln(1+\tau) - \ln(\omega^*)}{2\ln(1+\varepsilon) + \ln(\gamma)} \tag{20}$$

$$\overline{\alpha}_F \equiv \frac{\ln(1+\varepsilon) + \ln(1+\tau) - \ln(\gamma) - \ln(\omega^*)}{2\ln(1+\varepsilon) - \ln(\gamma)}. \tag{21}$$

The above thresholds are obtained from the expressions in (12) and (13) in the following way: $\underline{\alpha}_H$ solves $P^F_{j,F}(\underline{\alpha}_H) = P^F_{j,H}(\underline{\alpha}_H)$ and $\underline{\alpha}_F$ solves $P^H_{j,H}(\underline{\alpha}_F) = P^H_{j,F}(\underline{\alpha}_F)$ when $\alpha_j \leq 0.5$, whereas $\overline{\alpha}_H$ solves $P^F_{j,F}(\overline{\alpha}_H) = P^F_{j,H}(\overline{\alpha}_H)$ and $\overline{\alpha}_F$ solves $P^H_{j,H}(\overline{\alpha}_F) = P^H_{j,F}(\overline{\alpha}_F)$ when $\alpha_j \geq 0.5$. Hence, the thresholds $\underline{\alpha}_H$ and $\overline{\alpha}_H$ (resp. $\underline{\alpha}_F$ and $\overline{\alpha}_F$) pin down the final goods such that, given the

[11] Note that Assumption 2 is a *sufficient* condition (but is not a *necessary* condition) to ensure that positive trade between H and F always takes place in equilibrium. Intuitively, Assumption 1 creates another source of comparative advantage in our model, in addition to heterogeneities in sectoral labour productivities. As a result, even when $\varepsilon \leq \tau$, our model may still deliver positive international trade in equilibrium, provided γ is sufficiently small.

value of ω^*, their market price when sold in F (resp. when sold in H) would be identical regardless of where it was originally produced. Notice that our parametric assumptions imply that $\overline{\alpha}_H < \overline{\alpha}_F$, while the equilibrium result $\omega^* > 1$ means that $\overline{\alpha}_F < 1$. Furthermore, $\underline{\alpha}_H > \underline{\alpha}_F$ when $(1+\varepsilon)\gamma^{\frac{1}{2}} > 1$, while $\underline{\alpha}_H < \underline{\alpha}_F$ holds when $(1+\varepsilon)\gamma^{\frac{1}{2}} < 1$. In addition, the results in Proposition 1 concerning the bounds for ω^* imply that $\underline{\alpha}_H > 0$ always hold.[12]

By using the thresholds (18)-(21), we can fully split the space of final goods according to their price in the destination country, given the country of origin of the good.[13]

Lemma 1 *From the expressions in (12) and (13), and the equilibrium relative wage ω^*, by using (18)-(21), we can derive the following a set of conditions for $P^F_{j,F}$ relative to $P^F_{j,H}$ and for $P^H_{j,H}$ relative to $P^H_{j,F}$:*

1. *Suppose $(1+\varepsilon)^2\gamma > 1$, then:*
   - *$P^F_{j,H} < P^F_{j,F}$ for $0 \leq \alpha_j < \alpha^*_H$, while $P^F_{j,H} > P^F_{j,F}$ for $\alpha^*_H < \alpha_j \leq 1$, where $\alpha^*_H = \underline{\alpha}_H$ if $\omega^* \geq (1+\tau)^{-1}\gamma^{-\frac{1}{2}}$ holds true in equilibrium, while $\alpha^*_H = \overline{\alpha}_H$ if instead in the equilibrium $\omega^* < (1+\tau)^{-1}\gamma^{-\frac{1}{2}}$.*
   - *$P^H_{j,H} < P^H_{j,F}$ for $0 \leq \alpha_j < \overline{\alpha}_F$, while $P^H_{j,H} > P^H_{j,F}$ for $\overline{\alpha}_F < \alpha_j \leq 1$.*
2. *Suppose $(1+\varepsilon)^2\gamma \leq 1$, then:*
   - *$P^F_{j,H} < P^F_{j,F}$ for $\underline{\alpha}_H < \alpha_j < \overline{\alpha}_H$, while $P^F_{j,H} > P^F_{j,F}$ for $0 \leq \alpha_j < \underline{\alpha}_H$ and for $\overline{\alpha}_H < \alpha_j \leq 1$.*
   - *$P^H_{j,H} < P^H_{j,F}$ for $\max\{0, \underline{\alpha}_F\} < \alpha_j < \overline{\alpha}_F$, while $P^H_{j,H} > P^H_{j,F}$ for $\overline{\alpha}_F < \alpha_j \leq 1$ and $0 \leq \alpha_j < \max\{0, \underline{\alpha}_F\}$ whenever $\underline{\alpha}_F > 0$, where $\underline{\alpha}_F > 0$ if and only if $\omega^* \geq (1+\tau)(1+\varepsilon)$ holds in the equilibrium.*

In equilibrium, consumers in both H and F will always buy good j from the producer who can sell it in each market at the lower price. Hence, from Lemma 1, we can straightforwardly derive the equilibrium patterns of trade and specialisation in our two-country world economy.

[12] The comparisons of $\underline{\alpha}_H$ vis-a-vis $\overline{\alpha}_H$ and $\underline{\alpha}_F$ vis-a-vis $\overline{\alpha}_F$ are somewhat more convoluted, as they involve several possible combinations of parametric configurations and feasible solutions for ω^* given those configurations. For example, whenever $(1+\varepsilon)\gamma^{\frac{1}{2}} < 1$ holds true, for any feasible values of ω^*, we have $0 < \underline{\alpha}_H < 0.5 < \overline{\alpha}_H < 1$ and $0 < \underline{\alpha}_F < 0.5 < \overline{\alpha}_F < 1$. Instead, when $(1+\varepsilon)\gamma^{\frac{1}{2}} > 1$, we have $\underline{\alpha}_H < \overline{\alpha}_H$ if and only if $\omega^* > (1+\tau)^{-1}\gamma^{-\frac{1}{2}}$, which will actually fail to hold for τ sufficiently close to zero. On the other hand, when $(1+\varepsilon)\gamma^{\frac{1}{2}} > 1$, it is always the case that $\underline{\alpha}_F > \overline{\alpha}_F$, since Proposition 1 shows that $\omega^* < (1+\tau)\gamma^{-\frac{1}{2}}$ holds in that range.

[13] Naturally not all these prices will actually materialise in equilibrium, as consumers in each country will only buy from the producers who can offer the good at the lowest price.

Proposition 2 *The patterns of specialisation and trade differ qualitative depending on whether* $\gamma^{-1} > (1+\varepsilon)^2$ *or* $\gamma^{-1} < (1+\varepsilon)^2$.

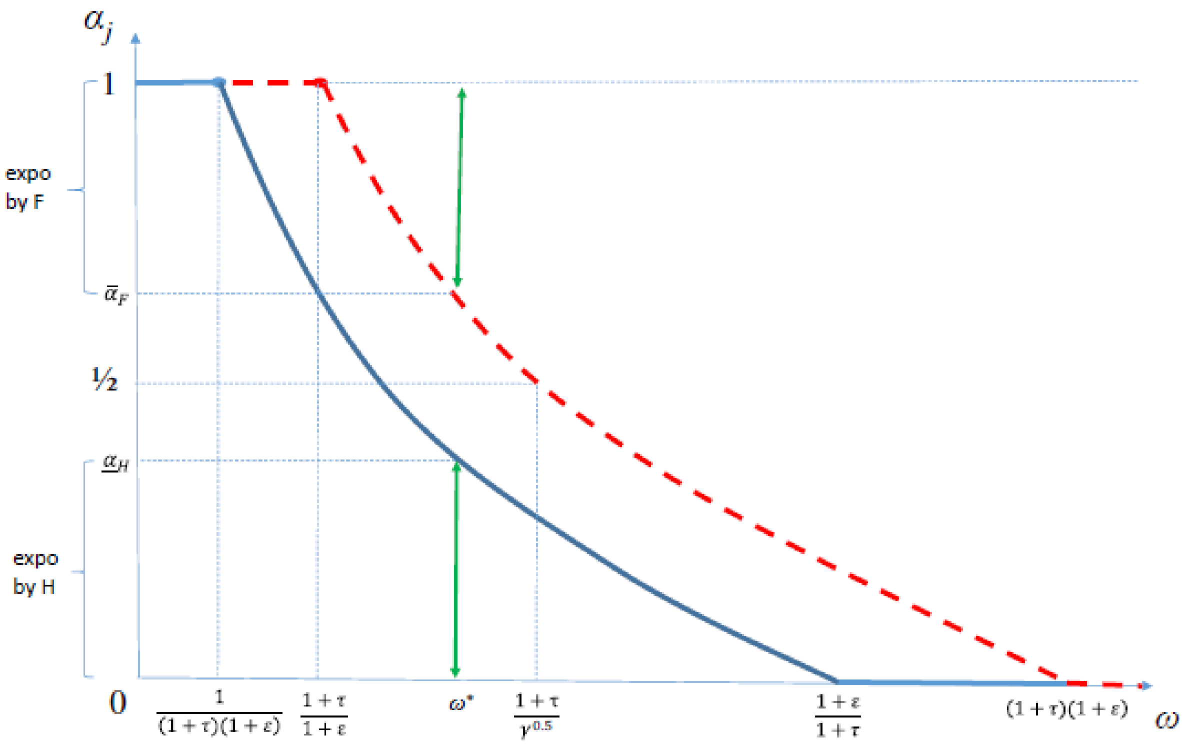
i) When $\gamma^{-1} < (1+\varepsilon)^2$*, trade patterns and specialisation are mainly determined by heterogeneities in labour productivities. Country H becomes an exporter of final goods whose* $\alpha_j \in [0, \alpha_H^*)$*, where* $\alpha_H^* = \underline{\alpha}_H$ *(resp.* $\alpha_H^* = \overline{\alpha}_H$*) if* $\omega^* \geq (1+\tau)^{-1} \gamma^{-\frac{1}{2}}$ *(resp.* $\omega^* < (1+\tau)^{-1} \gamma^{-\frac{1}{2}}$*) holds true. On the other hand, F becomes an exporter of the final goods whose* $\alpha_j \in (\overline{\alpha}_F, 1]$*. Final goods whose* $\alpha_j \in [\alpha_H^*, \overline{\alpha}_F]$ *are sourced locally by both H and F.*

ii) When $\gamma^{-1} > (1+\varepsilon)^2$*, trade patterns and specialisation are mainly determined by road network length differences between H and F. In those cases, H becomes an exporter of final goods whose* $\alpha_j \in (\underline{\alpha}_H, \overline{\alpha}_H)$*. On the other hand, F becomes an exporter of final good whose* $\alpha_j \in (\overline{\alpha}_F, 1]$ *if* $\omega^* > (1+\varepsilon)(1+\tau)$ *holds true, while it becomes an exporter of final goods whose* $\alpha_j \in [0, \underline{\alpha}_F) \cup (\overline{\alpha}_F, 1]$ *if instead* $\omega^* \leq (1+\varepsilon)(1+\tau)$ *holds true. When* $\omega^* > (1+\varepsilon)(1+\tau)$ *final goods whose* $\alpha_j \in [0, \underline{\alpha}_H] \cup [\overline{\alpha}_H, \overline{\alpha}_F]$ *are sourced locally by both H and F, while when* $\omega^* \leq (1+\varepsilon)(1+\tau)$ *this happens for those whose* $\alpha_j \in [\underline{\alpha}_F, \underline{\alpha}_H] \cup [\overline{\alpha}_H, \overline{\alpha}_F]$.

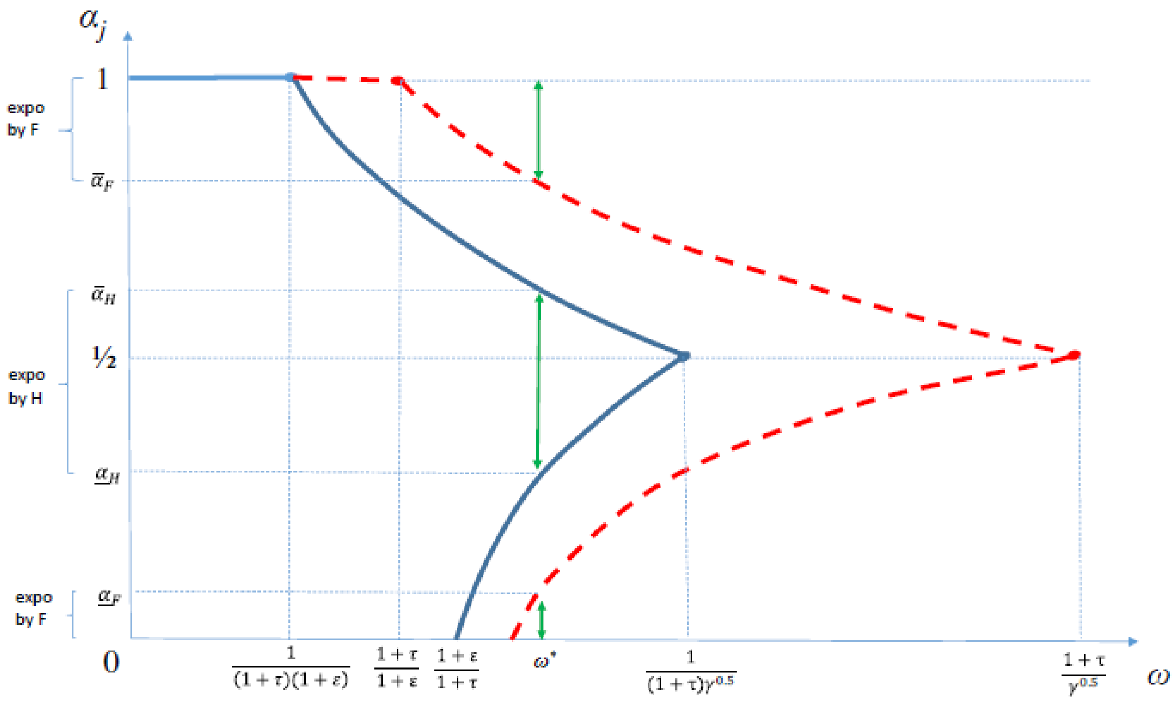
The patterns of trade and specialisation described by Proposition 2 are graphically depicted in Figure 3. The upper panel plots *case i)* of Proposition 2 –i.e., $\gamma^{-1} < (1+\varepsilon)^2$–, while and the lower panel shows *case ii)* –i.e., $\gamma^{-1} > (1+\varepsilon)^2$.[14] The vertical axis of Figure 3 orders final goods according to their specific $\alpha_j \in [0,1]$; the horizontal one measures the relative wage ω.[15]

[14]For brevity, the upper panel of Figure 3 shows the sub-case where $\omega^* > (1+\tau)^{-1} \gamma^{-\frac{1}{2}}$ –implying that H exports goods with $\alpha_j \in [0, \underline{\alpha}_H)$–, while its lower panel shows the sub-case where $\omega^* \leq (1+\varepsilon)(1+\tau)$ –entailing that F exports goods with $\alpha_j \in [0, \underline{\alpha}_F)$ and with $\alpha_j \in (\overline{\alpha}_F, 1]$–. In Appendix #, Figure 3 (bis) plots the other two sub-cases encompassed by Proposition 2.

[15]The solid line in Figure 3 is obtained by plotting $\underline{\alpha}_H$ and $\overline{\alpha}_H$, as given by (18) and (19) but replacing the specific equilibrium wage ω^* by a generic $\omega \geq 0$. Similarly, the dashed line is obtained by plotting $\underline{\alpha}_F$ and $\overline{\alpha}_F$, as given by (20) and (21) but using a generic $\omega \geq 0$. Notice that only the parts of the solid and dashed lines below $\alpha_j = 0.5$ follow the expressions in (18) and (20), while only the parts above $\alpha_j = 0.5$ are dictated by (19) and (21), respectively.



case 1: $\gamma^{-1} < (1+\varepsilon)^2$



case 2: $\gamma^{-1} > (1+\varepsilon)^2$

Figure 3: Patterns of Trade and Specialisation

Consider first the upper panel of Figure 3. Given a certain level of ω, all the goods that lie below the solid line would be exported by H, while all the goods lying above the dashed line would be exported by F. The gap in between the two curves represents the set of goods that would *not* be traded internationally. As it can be observed, the set of goods exported by H gets smaller as ω increases. Conversely, the set of goods exported by F expands with ω. At the extremes, when $\omega \leq 1/(1+\tau)(1+\varepsilon)$ all final goods would be produced in H, whereas for $\omega \geq (1+\tau)(1+\varepsilon)$ they would all be produced in F (naturally, such extreme values of ω could not possibly hold in equilibrium). At the equilibrium wage, ω^*, final goods with $\alpha_j > \overline{\alpha}_F$ are exported by F, and those with $\alpha_j < \underline{\alpha}_H$ are exported by H.[16] Hence, H becomes an exporter of the final goods that use input 0 more intensively (i.e., low-α_j goods), while F an exporter of those which use input 1 more intensively (i.e., high-α_j goods). Intuitively, the condition $\gamma^{-1} < (1+\varepsilon)^2$ means that differences in road network lengths between H and F are *small* relatively to their heterogeneities in sectoral labour productivities. As a result, the labour productivity differentials in the intermediate sectors –dictated by (9) and (10)– becomes the leading source of comparative advantage, regulating trade flows in the model.

Consider now the lower panel of Figure 3. For values of $\alpha_j > 0.5$, this graph exhibits the same qualitative features as the one in the upper panel. In fact, the interpretation of the curves within the range $\alpha_j > 0.5$ is analogous in both graphs: given a ω, the final goods that lie below the solid line would be exported by H and those lying above the dashed line would be exported by F. The main visual differences between the graphs arise when $\alpha_j < 0.5$. Within this range, the final goods located *above* the solid line would be exported by H, whereas those located *below* the dashed line would be exported by F.[17] In turn, this case leads to a pattern of specialisation that differs quite drastically from that one depicted in the upper panel of Figure 3. In particular, when, $\gamma^{-1} > (1+\varepsilon)^2$, we can observe that F ends up exporting the final goods located at the two (opposite) ends of the unit set –namely, $\alpha_j \in [0, \underline{\alpha}_F)$ and $\alpha_j \in (\overline{\alpha}_F, 1]$–, while H becomes an exporter of those in the intermediate range of α_j –namely, $\alpha_j \in (\underline{\alpha}_H, \overline{\alpha}_F)$–.

The pattern of specialisation depicted in the lower panel of Figure 3 represents the most important result of the model. The intuition for the result rests on the fact that when $\gamma^{-1} > (1+\varepsilon)^2$ the gap in the road network length is *large* relative to the heterogeneities in sectoral

[16] Notice that given the utility function (1), it must be that in equilibrium $(1-\overline{\alpha}_F) \times \omega^* = \underline{\alpha}_H$, where $(1-\overline{\alpha}_F) \times \omega^*$ equals total imports by H and $\underline{\alpha}_H$ equals total exports by H.

[17] Analogously to the upper panel of Figure 3, as ω increases, the set of goods exported by H shrinks and that one exported by F expands. The gap in between the curves represents the set of goods that are not traded internationally. Finally, for $\omega < 1/(1+\tau)(1+\varepsilon)$ all final goods would end up being produced by H, while for $\omega > (1+\tau)\gamma^{-0.5}$ they would all end up being produced by F.

labour productivities, and thus becomes the leading determinant of comparative advantages. Final goods with intermediate values of α_j require the use of both inputs in similar intensity. This means that a large share of their inputs will *necessarily* have to be transported along the road network of the economy. Instead, firms producing final goods with high and low values of α_j can source a relatively large share of their inputs from the *same* location where they are, thus *without* the need to rely for it on the internal transport network so heavily. In other words, sectors in the intermediate range of α_j require the use of the internal transport network more strongly than those whose α_j lies on the upper and lower spectrum of the unit interval. Accordingly, when H enjoys a much larger road network than F, the former specialises in the final goods with intermediate values of α_j, and the latter in those with more extreme values of α_j.

4 Empirical Predictions: From the Theory to the Data

In this section, we first describe how we attempt to bring to the data the main variables of interest present in the model. Next, we explain how we approach the data on bilateral trade flows to seek for evidence consistent with the main prediction of the model.

4.1 Main Variables of Interest

Input Narrowness

The first task in bringing some of the key aspects of the model to the data is coming up with a measure of the breadth of set of intermediates used in the production of each final good. Clearly, the model presented in Sections 2 and 3 is quite stylised to allow a direct match between its technological environment and real world data on inputs and outputs by sectors. In particular, in the real world, production functions depart from those postulated in (3) in two main dimensions. First, sectors tend to use more than only two intermediate goods. Second, the distinction between final goods and intermediate goods is not so clear-cut, as many goods satisfy both roles. Despite these shortcomings, we can nonetheless use the model as a guide to construct measures of narrowness of the intermediate inputs base for different industries.

In our model, sector j allocates a fraction α_j of their total spending in intermediate goods on input 0 and the remainder $(1-\alpha_j)$ on input 1. This means that sectors with very low or very high values of α_j source most of their inputs from only one intermediate sector. In other words, sectors on either end of the spectrum of $\alpha_j \in [0,1]$ exhibit a relatively *narrow* intermediate input base. Conversely, sectors with values of α_j around one half rely quite importantly on both inputs, and thus display a relatively *wide* intermediate input base.

We formally define the *narrowness* of the input base of sector j by the *Gini coefficient* of their expenditure shares across both inputs; henceforth denoted by $Gini_j$. The greater the value of $Gini_j$ is the narrower input base of sector j would be.[18] By using the fact that expenditure shares on input 0 and 1 are given, respectively, by α_j and $1-\alpha_j$, we can observe that:

$$Gini_j = \begin{cases} \frac{1}{2} - \alpha_j & \text{if } 0 \leq \alpha_j \leq \frac{1}{2} \\ \alpha_j - \frac{1}{2} & \text{if } \frac{1}{2} \leq \alpha_j \leq 1. \end{cases} \tag{22}$$

Hence, the above expression shows that $Gini_j = 0$ when $\alpha_j = 0.5$, while it grows symmetrically as α_j moves from its central value of 0.5 towards either extremes on 0 and 1.

To construct a measure input narrowness analogous to that one in (22), but based on the available real world data, we resort to the input-output (IO) matrix of the US in 2007 from the Bureau of Economic Analysis (BEA).[19] The IO matrix comprises 389 sectors/industries. Although the IO matrix exhibits the same number of sectors producing intermediate goods as those producing final output, we restrict the set of final goods to those also present in the international trade data (see description below). Thus, we index separately by $k = 1, 2, ..., K$ each of the sectors present in the IO matrix that produce final goods, and by $n = 1, 2, ..., N$ each of the sectors selling intermediate inputs.

We let $X_{k,n} \geq 0$ denote the total value of intermediate good n purchased by sector k. Defining $S_{k,n} \equiv X_{k,n} / \sum_{n=1}^{N} X_{k,n} \geq 0$ as the share of n over the total value of intermediates purchased by k, we can compute the Gini coefficients analogously to those in (22). Namely,

$$Gini_k = \frac{2 \times \sum_{n=1}^{N} n \times S_{k,n}}{N \times \sum_{n=1}^{N} S_{k,n}} - \frac{N+1}{N}, \tag{23}$$

where the argument $\sum_{n=1}^{N} n \times S_{k,l}$ in the numerator of $Gini_k$ is ordering intermediates in non-decreasing order (i.e., $S_{k,n} \leq S_{k,n+1}$).

In the empirical analysis conducted in Section 4.3, we use $Gini_k$ to measure the degree of narrowness of the input base of sector k. Large values of $Gini_k$ are the result of sector k sourcing

[18] Imbs and Wacziarg (2003) have previously used the Gini coefficient to measure the degree of concentration of labour and value added across different sectors in the economy. In this paper, we apply a similar methodology, but we use it to measure the degree of narrowness/concentration of the intermediate input base of different sectors in the economy. There are other measures that could alternatively be used to capture the same concept; e.g., coefficient of variation, log-variance, Herfindahl index. We use those alternative measures in our empirical analysis in Section 4.3 as robustness check of the results when using the Gini.

[19] This data is publicly available from https://www.bea.gov/industry/io_annual.htm.

most of their intermediate inputs from relatively few sectors (i.e., when the distribution of shares $S_{k,n}$ is very concentrated in few intermediates). Conversely, small values of $Gini_k$ tend to occur when sector k purchases their intermediate inputs from a relatively wide input base (i.e., when the distribution of $S_{k,n}$ is quite evenly spread across a large number of intermediates).[20] Notice finally the link between $Gini_k$ in (23) and $Gini_j$ in (22): the former boils down to the latter when $N = 2$, and $S_{k,n} = \alpha_{k,n}$ with $\alpha_{k,1} + \alpha_{k,2} = 1$.

Export Specialisation

In order to measure the degree of export specialisation by sectors we use the data on trade flows from COMTRADE compiled by Gaulier and Zignago (2010). We use only trade flows in year 2014. The data are categorised following the Harmonized System (HS) 6-digit classification. There are in total 5,192 6-digit products.

We are interested in countries' specialisation in sectors/industries ranked according to their degree of input narrowness. We need thus to map the trade flows data based on the HS 6-digit classification to the industry codes used by the BEA for the United States IO matrix. We do so by using the concordance table between the 2002 IO matrix commodity codes and the HS 10-digit classification from the BEA, and grouping the HS 10-digit codes into HS 6-digit products. In the cases in which an HS-6 product maps into more than one BEA's code, we assign their trade flows proportionally to each of the BEA sectors in which it maps.[21] Lastly, the Input-Output industry codes of the 2002 classification are matched to those of the 2007 classification, which are the ones actually used to compute the measures of input narrowness based on (23). In the end, after mapping the HS 6-digit products into the BEA IO industry codes, we are left with data on bilateral trade flows for 273 industries.

Road Network

The last key variable in our model is the length of the road network of country c (r_c). We take the road network length by countries from the data on *roadways* from the CIA World

[20] In the extreme (unequal) case in which $S_{k,n'} = 1$ for some n' and $S_{k,n} = 0$ for all $n \neq n'$, (23) yields $Gini_k = 2 - [(N+1)/N]$, which approaches 1 as $N \to \infty$. On the other hand, in the case when $S_{k,n} = S_k > 0$ for all $n = 1, ..., N$, we would have $Gini_k = 0$.

[21] There are 526 HS-6 products that map into two BEA Input-Output industry codes, 96 products that map into three IO codes, 33 products that map into four IO codes, and 11 products that map into five or more IO codes. None of the regression results in Section 4.3 are significantly altered when all the HS-6 products that map into more than one BEA Input-Output industry code are dropped from the sample.

Factbook. Roadways are defined as 'total length of the road network, including paved and unpaved portions'. The World Factbook contains data on roadways for 223 countries. The year of the data point for each country varies, ranging from year 1999 to 2016, with the median year of the sample being 2010. The sample we use in Section 4.3 contains 136 countries, whose measured length of roadways years range from 2000 to 2016. When defining our empirical counterpart of the variable r_c, we divide the length of the road network by the total area of the country; that is, $r_c \equiv roadways_c/area_c$.

4.2 Road Density and Patterns of Specialisation: Testing the predictions of the model

The two-country model presented in Section 3 predicts that when heterogeneities in road networks across the economies are sufficiently large, the patterns of specialisation and trade flows follow those depicted by the lower panel of Figure 3. More formally, when the condition $\gamma^{-1} > (1+\varepsilon)^2$ holds true, the country with the longer road network (i.e., country H) will export goods with intermediate values of α_j, while the country with the shorter road network (i.e., country F) will export goods with values of α_j located on the extremes of the unit continuum. Conceptually, this prediction can be interpreted as stating that countries with longer road networks will tend to exhibit a comparative advantage in the types of goods that require a wider (or more diverse) set of intermediate inputs.

From an empirical viewpoint, if road network length differences across countries shaped somehow their patterns of specialisation as our model predicts, we should then observe the following: economies with a greater r_c will tend to export relatively more of the goods produced in industries with a smaller value of $Gini_k$ vis-a-vis economies with smaller r_c. We test this prediction by means of the following regression:

$$\ln(1 + Expo_{k,c}) = \beta \cdot (r_c \times Gini_k) + \boldsymbol{\delta} \cdot \boldsymbol{\Delta}_{c,k} + \varsigma_c + \kappa_k + \upsilon_{c,k}, \tag{24}$$

In the regression equation (24) the dependent variable is given by the natural logarithm of (1 plus) the total value of exports in industry k by country c to all other countries in the world in year 2014. The logarithm is applied to $1 + Expo_{k,c}$, rather than simply using $\ln(Expo_{k,c})$, in order to possibly include in the regressions observations where $Expo_{k,c} = 0$.[22] The term $(r_c \times Gini_k)$ interacts the measure of input narrowness defined in (23) with the measure of

[22] None of the results in Section 4.3 are substantially altered when we drop observations with $Expo_{k,c} = 0$, and we use instead $\ln(Expo_{k,c})$ as dependent variable in (24). The results excluding observations with zero exports are available from the author upon request.

road density (i.e., length of roadways per square kilometer). $\mathbf{\Delta}_{c,k}$ denotes a vector of additional covariates that may possibly influence specialisation across countries in industries differing in terms of the degree of input narrowness. ς_c and κ_k denote country fixed effects and industry fixed effects, respectively. Finally, $\upsilon_{c,k}$ represents an error term.

The main coefficient of interest in (24) is β. If, as the model predicts, countries with a denser road network (i.e., countries with a greater r_c) indeed tend to exhibit a comparative advantage in goods from industries that require a wider set intermediate inputs (i.e., industries with a smaller $Gini_k$), the data should then deliver a negative estimate of β.

4.3 Empirical Results

Table 1 displays the first set of estimation results corresponding to (24). In column (1) we include only our main variable of interest (i.e., the interaction term between road density in country c and the degree of input narrowness of industry k), together with the exporter and industry dummies. The correlation is negative and highly significant, suggesting that countries with denser road networks tend to export relatively more of the final goods whose production process requires a wider intermediate input base (i.e., those exhibiting a lower $Gini_k$).

In column (2) we incorporate some additional interaction terms that may also influence the patterns of specialisation of countries across industries with different levels of input narrowness. This column includes interaction terms between $Gini_k$ and GDP per capita, total GDP, and population. Omitting these terms may lead to a biased estimate of β if some of their components are correlated with r_c, while at the same time they impact the patterns of specialisation across industries with different values of $Gini_k$. For example, larger economies in terms of total GDP may be better able to produce goods with lower $Gini_k$ if there is a fixed cost to open some intermediate sectors. Similarly, more populated countries may enjoy a more diverse labour force (or more diverse human capital), which could also allow producing a wider set of goods. More importantly, it may be the case that richer economies can export relatively more in industries with wider input bases simply because they tend to be more diversified.[23] In turn, since richer economies also tend to enjoy denser road networks, omitting this interaction term could lead to an overestimation (in absolute value) of the correlation coefficient of interest in (24). The results in column (2) show that this is indeed the case. The estimate for the interaction term

[23] It may also be the case that richer economies may be relatively more specialised in sectors with low $Gini_k$ because they have better financial markets or institutions that help establishing links with intermediate suppliers. Column (2) will also partly control for these channels. In addition, in Table 2 we directly incorporate measures of financial development and rule of law in the regression equation.

TABLE I

Export Specialisation across Industries with Different Levels of Input Narrowness

	(1)	(2)	(3)	(4)	(5)	(6)
Road Density x Input Narrowness	-5.012***	-3.501***	-3.501**	-4.056***	-0.122***	-1.084***
	(0.571)	(0.658)	(1.429)	(0.771)	(0.025)	(0.201)
GDP per capita x Input Narrowness		-0.110***	-0.110	0.050	0.000	0.023**
		(0.035)	(0.100)	(0.044)	(0.001)	(0.011)
GDP x Input Narrowness		-0.414	-0.414	-0.817***	-0.022**	-0.236***
		(0.292)	(0.655)	(0.320)	(0.010)	(0.083)
Population x Input Narrowness		-0.004	-0.004	-0.002	-0.0001	-0.0001
		(0.004)	(0.007)	(0.005)	(0.0001)	(0.001)
Road Density			4.948***			
			(1.343)			
GDP per capita			0.165*			
			(0.094)			
GDP			0.833			
			(0.616)			
Population			0.005			
			(0.007)			
Observations	37,128	37,128	37,128	37,128	37,128	37,128
R-squared	0.768	0.768	0.493	0.767	0.767	0.768
Country FE	Yes	Yes	No	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Number Countries	136	136	136	136	136	136
Narrowness Measure	Gini	Gini	Gini	Coef Var	Log Var	Herf

Robust standard errors in parentheses. The dependent variable is the logarithm of total exports in industry k by country c plus 1; i.e. $\log(1+Expo_{c,k})$, and includes observations with $Expo_{c,k} = 0$. The total number of industries is 237 in all regressions. GDP per capita, GDP, and population are taken from the Penn Tables. All data corresponds to year 2014. *** p<0.01, ** p<0.05, * p<0.1

between GDP per capita and $Gini_k$ is negative and significant, and the absolute value of $\widehat{\beta}$ falls relative to column (1). However, this estimate still remains negative and highly significant.

Next, in column (3) we introduce directly all the above-mentioned regressors, in addition to their interaction terms. (We have to drop in this case the country fixed effects ς_c, as they are perfectly collinear with the country-level variables.) As we can observe, our main correlation of interest remains unaffected.

Finally, the last three columns show the results of the regression in column (2) when using three alternative measures of input narrowness: the coefficient of variation, the log-variance, and the Herfindahl index. (Like for $Gini_k$ in (23), these indices are computed using the industry k's expenditure shares across all intermediates, $S_{k,n}$.) Once again, the estimate of β remains negative and highly significant in all cases.

Robustness Checks

In Table 2 we keep incorporating some additional interaction terms to gauge the robustness of previous results, and to rule out some further possible sources of bias due to omitted variables.

There is a large body of literature that sustains that financial markets are instrumental to opening new sectors and increasing the variety of industries in the economy (e.g., Greenwood and Jovanovic, 1990; Acemoglu and Zilibotti, 1997 and 1999). From this perspective, we could expect that countries with more developed financial markets would also be better able to specialise in industries that require the use of a wider input base. As a result, by not including interaction terms with measures of financial development into (24), we may end up with a biased estimate of β. To deal with this concern, in columns (1) and (2) we include two separate indicators of financial development taken from the World Bank Indicators database: the ratio of private credit to GDP and the ratio of stock market capitalisation to GDP. In both cases the indicators are averaged during years 2005-2014. When measured by private credit, the effect of financial development interacted with $Gini_k$ is significant and it carries a sign consistent with the past literature on growth and diversification. Moreover, including it into the regression leads to a smaller estimate (in absolute terms) of β, consistent with the above argument. Instead, when using stock market capitalisation the effect becomes insignificant, while the previous estimates remain (quantitatively) almost intact. Note, however, that the number of countries in the sample of column (2) shrinks substantially relative to column (1).

Columns (3) and (4) add an interaction term between $Gini_k$ and an index of human capital in country c. The human capital index is taken from the Penn Tables, and corresponds to year 2014.[24] This additional term would control for the possibility that industries which are human capital intensive may also be producing goods that require a wide set of intermediate inputs. Although in column (4) this additional interaction term is negative and significant, our main estimate of interest remains essentially intact.

Next, in columns (5) and (6) we introduce an interaction term between $Gini_k$ and an index of Rule of Law, taken from World Governance Indicators for year 2014. The rationale behind this is related to the argument in Levchenko (2007) and Nunn (2007), who shows that countries with better contract enforcement institutions display a comparative advantage in industries that rely more strongly on relationship-specific investments. Industries that need to source a larger

[24] The index of human capital in the Penn Tables is based on the average years of schooling from Barro & Lee (2013) and an assumed rate of return of education based on Mincer estimates. The results in column (3) and (4) do not change much when using human capital measures based on alternative sources.

set of intermediate inputs may benefit relatively more from a sound legal environment, as they need to establish relationships with a greater number of intermediate good providers. While the results with this additional interaction term seem to cast some support to this argument, they do not undo or invalidate our previous results regarding the estimates of β.

Lastly, the last two columns of Table II address the possibility of a differential effect of the road network on the pattern of specialisation depending on the population density of the economy. While in our two-country model we take the geographic and population structure of the economies as identical, in the real world this assumption is clearly untenable. One could thus expect that more densely populated countries may display also a greater concentration of activities in fewer locations. Hence, all else equal, more densely populated countries may need to resort less strongly on a vast road network than sparsely populated countries do. Columns (7) and (8) assess this possibility by introducing an interaction term between population density and $Gini_k$, and a *triple* interaction term which also includes r_c. If road network length is especially important for specialisation as predicted by our model in economies that are *less* densely populated, then the triple interaction term should carry a positive estimate. As we can readily observe, this is indeed the case. Moreover, the estimate of β after introducing the triple interaction term is still negative and highly significant.

Table III provides another set of robustness checks to our previous results. It runs a set of regressions as some of those previously shown in Table I and II, but splitting that sample of countries in two according to whether their income is above or below the median. The odd numbered columns show the results for the subsample of 'high-income countries', while the even numbered columns do that for the 'low-income countries'. (Notice that in some of the regressions the subsamples are not balanced due to data availability.) The results show that the effect of road density on pattern of specialisation holds true both for richer and poorer countries. In addition to that, the effect seems to be consistently greater in magnitude for the subsample of economies whose income is below the median.

Finally, Table II (bis) in Appendix displays the results of the same regressions as in Table II, but when the dependent variable is given by export shares rather than the logarithm of exports. As we can observe, all the results remain qualitatively intact.

TABLE II

Export Specialisation across Industries with Different Levels of Input Narrowness: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road Density x Narrowness	-2.880***	-3.765***	-2.865***	-2.810***	-2.563***	-2.220***	-6.790***	-6.910***
	(0.645)	(0.622)	(0.666)	(0.636)	(0.708)	(0.666)	(1.451)	(1.436)
GDP per capita x Narrowness	-0.070*	-0.133***	-0.088**	-0.112***	-0.073	-0.082*	-0.077*	-0.079*
	(0.041)	(0.039)	(0.044)	(0.043)	(0.049)	(0.050)	(0.049)	(0.050)
GDP x Narrowness	-0.355	-0.404	-0.352	-0.040	-0.321	-0.043	-0.547*	-0.298
	(0.287)	(0.270)	(0.291)	(0.272)	(0.292)	(0.272)	(0.294)	(0.276)
Population x Narrowness	-0.004	-0.005	-0.004	-0.011***	-0.005	-0.012***	-0.001	-0.007*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Fin Dev (priv cred) x Narrowness	-0.041***		-0.040**		-0.027		-0.018	
	(0.015)		(0.016)		(0.019)		(0.019)	
Fin Dev (stk mkt cap) x Narrowness		0.009		0.012		0.019		0.012
		(0.012)		(0.012)		(0.013)		(0.013)
Human Capital x Narrowness			0.281	-4.288***	0.479	-3.251**	0.398	-4.331***
			(1.092)	(1.399)	(1.121)	(1.520)	(1.143)	(1.528)
Rule of Law x Narrowness					-1.137	-1.850*	-0.730	-1.015
					(1.126)	(1.034)	(1.140)	(1.067)
(Pop) Density x Narrowness							-2.262***	-3.193***
							(0.528)	(0.615)
Road Density x (Pop) Density x Nwness							1.873***	2.219***
							(0.414)	(0.406)
Observations	36,036	24,843	33,306	23,478	33,306	23,478	33,306	23,478
R-squared	0.769	0.744	0.769	0.747	0.769	0.747	0.770	0.748
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number Countries	132	91	122	86	122	86	122	86

Robust standard errors reported in parentheses. The dependent variable is log(1+$Expo_{c,k}$) and includes observations with $Expo_{c,k} = 0$. The total number of industries is 273. Data on GDP per capita, GDP, population and the human capital index are taken from the Penn Tables and corresponds to year 2014. Private credit over GDP and stock market capitalisation over GDP are taken from the World Bank Indicators, averaged for years 2005-2014. Rule of Law is taken from the World Governance Indicators and corresponds to year 2014. *** p<0.01, ** p<0.05, * p<0.1

TABLE III

Robustness Check: High-Income and Low-Income subsamples

	(1)	(2)	(3)	(4)	(5)	(6)
Road Density x Input Narrowness	-3.115***	-8.755***	-1.892***	-9.215***	-1.743***	-7.991***
	(0.601)	(2.235)	(0.621)	(2.334)	(0.611)	(2.329)
GDP per capita x Input Narrowness	-0.127***	0.393	-0.133***	0.504	-0.144***	0.231
	(0.043)	(0.284)	(0.053)	(0.441)	(0.053)	(0.493)
GDP x Input Narrowness	0.218	-6.240	0.732**	-8.543*	0.869***	-4.829
	(0.279)	(4.792)	(0.295)	(5.060)	(0.280)	(5.095)
Population x Input Narrowness	-0.018***	0.038	-0.027***	0.055*	-0.032***	0.030
	(0.006)	(0.027)	(0.006)	(0.029)	(0.006)	(0.029)
Fin Dev (priv cred) x Input Narrowness			-0.025	-0.111*		
			(0.020)	(0.062)		
Fin Dev (stock mkt cap) x Input Narrowness					0.013	0.021
					(0.013)	(0.029)
Human Capital x Input Narrowness			-8.924***	4.024**	-10.049***	0.313
			(1.910)	(1.967)	(1.887)	(3.036)
Rule of Law x Input Narrowness			1.741	-3.215	0.343	-10.147***
			(1.432)	(2.197)	(1.152)	(2.915)
Observations	18,564	18,564	16,653	16,653	15,561	7,917
R-squared	0.772	0.661	0.773	0.661	0.750	0.678
Country FE	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Countries Sample	High Income	Low Income	High Income	Low Income	High Income	Low Income
Number Countries	68	68	61	61	57	29

Robust standard errors in parentheses. The dependent variable is log(1+$Expo_{c,k}$) and includes observations with $Expo_{c,k}$ = 0. The total number of industries is 273. The High Income subsample comprises countries whose GDP per capita was above the median in 2014, and the Low Income subsamble those per capita was below it. The median income of the sample lies between that of Ecuador ($10,968 PPP) and Peru ($10,993 PPP) *** p<0.01, ** p<0.05, * p<0.1

5 Beyond the Model: Endogeneity and Alternative Interpretations

The results in the previous section yield a robust correlation between road density in country c and its degree of specialisation in industries that rely on a wide input base. Those results are certainly consistent with the main predictions of the model. Yet, they cannot be taken as hard evidence of the core mechanism underlying the model's predictions. In that regard, two separate issues deserve some further discussion and analysis. First, the correlation found in the previous regressions could be the result of road infrastructure responding to transport needs resulting from industry specialisation (in other words, the negative estimate of β may be the result of reverse causation). Second, our interpretation of a lower value of $Gini_k$ as reflecting greater need of industry k for the local transport infrastructure is debatable, as previous authors have looked at that variable as capturing the degree of product complexity of industry k. In the next two subsections we aim to address more explicitly these two points.

5.1 Endogeneity, Reverse Causation and Waterways Density

Our model has resorted to two critical assumptions that warrant some further discussion in case the previous empirical results are intended to be taken as evidence of a *causal* effect from road density to specialisation. Firstly, it has taken r_c as exogenously given. The length of a country's road network is however the result of investment choices in infrastructure, and hence it will respond to a host of economic variables and incentives. Secondly, the model has assumed away any sort of intrinsic differences in productivities *directly* linked to the production functions of final goods laid out in equation (3). In fact, all differences in countries' productivities across final sectors stem *indirectly* from the heterogeneities in the intensity of inputs implied by the parameter α_j.

Relaxing the two above-mentioned assumptions can easily lead to a model where the partial correlation parameter β in (24) can be confounding an effect from road density to specialisation, together with reverse causality from the latter to the former. For example, suppose that for some reason the final good production functions differ across countries, in a way such that H is relatively more productive than F in the final sectors whose α_j lies near one half.[25] In a context like this one, if countries can invest in expanding their road networks, we could well expect r_H to be larger than r_F simply because the incentives to do so are greater in H than in F. From an empirical viewpoint, this reasoning means that the partial correlation coefficient β in (24) could end up capturing (at least partially) an effect going from patterns of specialisation to road density.

One solution to the above problem would be to find an instrumental variable to generate variation in r_c that is plausibly exogenous. This section does not go that far. However, it intends to provide some further evidence consistent with the main mechanism of the model, relying on a measure of countries' transport network that is *less* sensitive to reverse causality concerns than r_c. We measure now the internal transport network of an economy by the density of their *waterways* network. We draw the data on waterways from the CIA World Factbook, and define *waterways density* as waterways length per square km.[26] Arguably, while countries can still

[25] For example, we may have that final good productions functions are given by (3) for $\alpha_j \in [0, 0.5 - \epsilon]$ and for $\alpha_j \in [0.5 + \epsilon, 1]$, where $0 < \epsilon < 0.5$, and by

$$Y_j = (1 + \phi_c) \frac{1}{\alpha_j^{\alpha_j} (1 - \alpha_j)^{1-\alpha_j}} X_{0,j}^{1-\alpha_j} X_{1,j}^{\alpha_j}, \quad \text{for } \alpha_j \in (0.5 - \epsilon, 0.5 + \epsilon), \quad \text{with } \phi_H > \phi_F.$$

[26] The CIA World Factbook measures waterways as the total length of navigable rivers, canals and other inland bodies of water.

affect the density of their waterways network by investing in creating canals or improving the navigability of some rivers and bodies of water, the scope for this is far more limited than in the case of building and expanding roadways.

One additional aspect we exploit here is the possibility that waterways density impacts specialisation *heterogeneously* at different stages of development. For a number of reasons, richer economies tend to have much denser road networks than poorer ones. In particular, poorer economies may find it harder to undertake the necessary investment to build a sufficiently developed road infrastructure. On the other hand, while the presence and density of waterways may have influenced patterns of development worldwide before alternative means of transport became more widespread, it can hardly be expected to impact the current level of development of economies in a systematic way. In fact, a quick look at simple cross-country correlations in Figure 4 shows that, while income per head and road density display a clear positive correlation, the association between income per head and waterways density is rather weak.

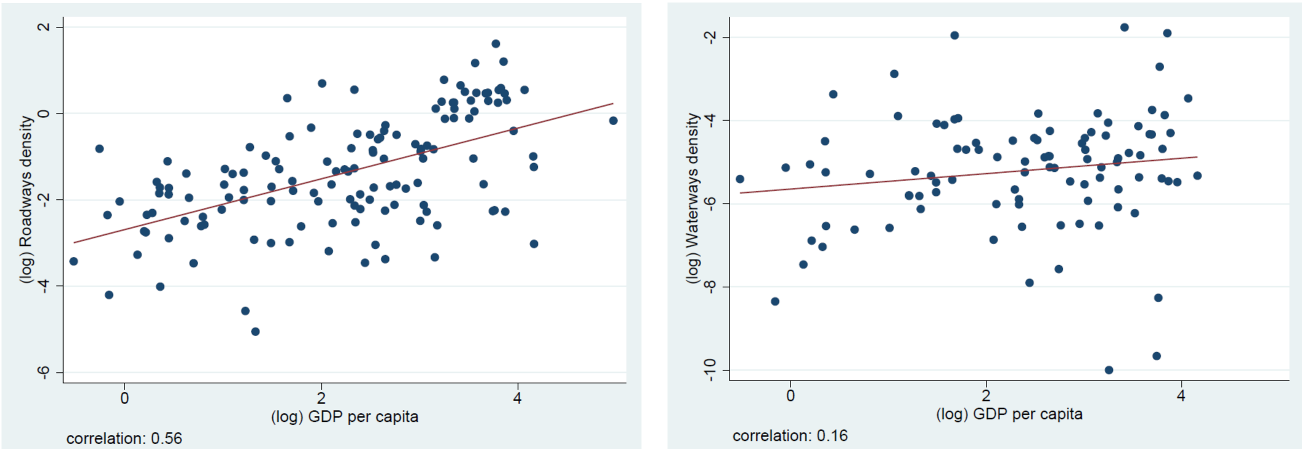


Figure 4: Roadways and waterways density against GDP per head

An interpretation of the simple correlations in Figure 4 is that, as economies grow richer, roadways tend to gradually supersede waterways as the main type of internal transport network. From this perspective, we could then expect waterways to represent an important determinant of patterns of specialisation in poorer economies, but that they should lose preeminence in richer economies where roadways can more easily make up for an insufficiently deep internal waterway network.

Table 4 displays the results of a regression equation analogous to (24), but where r_c is replaced by a measure of *waterways density*. The table shows the results of three sets of regressions for three different countries sample: entire sample, high-income countries, and low-income countries. The regressions based on the whole set of countries yield an estimate that is

TABLE IV
Waterways Density as Measure of Transport Network

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Waterways Density x Input Narrowness	-0.291*	0.044	-0.941***	-0.178	0.095	-1.118**	-0.253*	0.123	-1.097***
	(0.170)	(0.149)	(0.359)	(0.168)	(0.148)	(0.511)	(0.160)	(0.145)	(0.368)
GDP per capita x Input Narrowness	-0.278***	-0.283***	0.224	0.108	0.124	0.077	0.215**	0.083	0.697
	(0.038)	(0.061)	(0.253)	(0.103)	(0.109)	(0.356)	(0.107)	(0.112)	(0.553)
GDP x Input Narrowness	-0.042	0.253	-2.051**	-0.316	0.799	-2.762***	-0.073	1.178*	-2.566***
	(0.287)	(0.588)	(0.824)	(0.299)	(0.628)	(1.030)	(0.278)	(0.630)	(0.902)
Population x Input Narrowness	-0.010***	-0.006	0.006	-0.006	-0.034	0.012	-0.013***	-0.048*	0.016*
	(0.004)	(0.026)	(0.008)	(0.004)	(0.028)	(0.008)	(0.004)	(0.029)	(0.010)
Fin Dev (priv cred) x Input Narrowness				-0.025	-0.029	0.090			
				(0.019)	(0.020)	(0.098)			
Human Capital x Input Narrowness				-1.482	-10.383***	1.146	-7.042***	-10.309***	-6.171*
				(1.416)	(2.169)	(2.329)	(2.004)	(2.064)	(3.970)
Rule of Law x Input Narrowness				-4.578***	-2.465	-7.872***	-5.697***	-3.353**	-17.666***
				(1.406)	(1.743)	(2.676)	(1.453)	(1.615)	(4.538)
Fin Dev (stock mkt cap) x Input Nwness							-0.001	0.002	-0.123
							(0.019)	(0.019)	(0.094)
Observations	26,481	13,377	13,104	24,843	12,558	12,285	18,018	12,285	5,733
R-squared	0.779	0.771	0.695	0.780	0.762	0.701	0.756	0.745	0.738
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries Sample (income level)	All	High	Low	All	High	Low	All	High	Low
Number Countries	97	49	48	91	46	45	66	45	21

Robust standard errors reported in parentheses. The dependent variable is log(1+$Expo_{c,k}$), and includes observations with $Expo_{c,k}$= 0. The total number of industries is 237 in all regressions. Waterways is taken from the CIA World Factbook, and comprises total length of navigable rivers, canals and other inland water bodies. Waterway density equals waterways per sq km. All data corresponds to year 2014. *** p<0.01, ** p<0.05, * p<0.1

negative, and it is also mildly significant in two out of three cases. However, this aggregate result masks important heterogeneities in the effect of waterways density on export specialisation in the case of richer versus poorer economies. Columns (2), (5), and (8) essentially show that waterways density carries no impact whatsoever in the subsample of above-median income economies. In contrast, columns (3), (6), and (9) consistently exhibit a negative and highly significant coefficient. This result suggests that, in the case of poorer economies, those that enjoy a denser network of waterways tend to export relatively more of the goods that require a wider intermediate input base.[27]

5.2 Alternative Intepretations of the Input Breadth Measures

[TBD.

[27] Table IV (bis) in Appendix # shows that this result keeps holding true when the regressions in Table IV also include the interaction term ($r_c \times Gini_k$). The main difference between the results in Table IV and Table IV (bis) is that, in the case of richer economies, the coefficient for the interaction term between *waterways density* and $Gini_k$ turns quantitatively larger and significant.

6 Concluding Remarks

[TBD]

Appendix A: Proofs

Proof of Proposition 1. We first prove by contradiction that $\omega = 1$ cannot hold in equilibrium. Given that the expression in (17) entails that total imports from F by H increase with ω, while total exports by H to F decrease with ω, it will then follow that in equilibrium we must necessarily have $\omega^* > 1$, and that this equilibrium will be unique. We carry out the proof of $\omega^* > 1$ by splitting the possible parametric configurations of the model in three subsets.

i) Case 1: $\gamma(1+\tau)^2 \geq 1$. In this case, when $\omega = 1$, using the LHS of (17), it follows that total exports by H are equal to:

$$Expo_H = \frac{\ln(1+\varepsilon) - \ln(1+\tau)}{2\ln(1+\varepsilon) + \ln(\gamma)}. \tag{25}$$

Notice that $\gamma(1+\tau)^2 \geq 1$ implies the RHS of (25) is never greater than one half, while Assumption 2 implies it is strictly above zero. Using now the RHS of (17), we can obtain that total imports by H are:

$$Impo_H = 1 - \frac{\ln(1+\varepsilon) + \ln(1+\tau) - \ln(\gamma)}{2\ln(1+\varepsilon) - \ln(\gamma)}. \tag{26}$$

Comparing (25) versus (26), while bearing in mind $\gamma < 1$, yields $Expo_H > Impo_H$. Hence, when $\gamma(1+\tau)^2 \geq 1$, the equilibrium must necessarily encompass $\omega > 1$.

ii) Case 2: $\gamma(1+\tau)^2 < 1 < \gamma(1+\varepsilon)^2$. Using setting again (17), we obtain:

$$Expo_H = \frac{\ln(1+\varepsilon) - \ln(1+\tau) - \ln(\gamma)}{2\ln(1+\varepsilon) - \ln(\gamma)}, \tag{27}$$

while total imports by H are still given by (26). When $\gamma(1+\tau)^2 < 1$, the RHS of (27) yields a value strictly larger than one half, while the RHS of (26) is always strictly smaller than one half. As a consequence, $Expo_H > Impo_H$ also when $\gamma(1+\tau)^2 < 1 < \gamma(1+\varepsilon)^2$, and the equilibrium must necessarily encompass $\omega > 1$ in that range too.

iii) Case 3: $\gamma(1+\varepsilon)^2 < 1$. Using once again (17), notice that total exports by H are still given in this case by (27), which yields a value strictly above 0.5 and strictly below 1. In addition, total imports by H are still given by (26), which yields a value strictly above zero, but strictly below one half. Hence, when $\gamma(1+\tau)^2 \geq 1$, the equilibrium must necessarily encompass $\omega > 1$ as well.

Next, to prove that ω^* is strictly decreasing in γ, it suffices to note that the boundaries $A(\gamma, \varepsilon, \tau, \alpha_j)$ for ω in the expressions $\mu_{\alpha_j \leq 0.5}\left(\alpha_j \middle| \, \omega \lesseqgtr A(\gamma, \varepsilon, \tau, \alpha_j)\right)$ and $\mu_{\alpha_j > 0.5}\left(\alpha_j \middle| \, \omega \lesseqgtr A(\gamma, \varepsilon, \tau, \alpha_j)\right)$ in (17) are all strictly decreasing in γ.

Lastly, to prove the different bounds on ω^* we proceed by contradiction for each of them. First, suppose that $\gamma(1+\varepsilon)^2 > 1$. Notice that if $\omega^* \geq (1+\tau)\gamma^{-0.5}$, then using (17) we can observe that the mass of final goods exported by F would be at least one half. However, this is incompatible with the fact that in equilibrium $\omega^* > 1$. Hence, it must be that $\omega^* < (1+\tau)\gamma^{-0.5}$. Next, notice that when $\omega^* \geq (1+\varepsilon)/(1+\tau)$, the exports by H fall to zero, while H's imports are strictly positive; hence, this cannot hold in equilibrium either, and it must be that $\omega^* < (1+\varepsilon)/(1+\tau)$. Second, suppose now that $\gamma(1+\varepsilon)^2 < 1$... ■

References

[1] Acemoglu, D., Antràs, P. and Helpman, E. (2007). 'Contracts and Technology Adoption', *American Economic Review*, vol. 97, pp. 916-943.

[2] Acemoglu, D. and Zilibotti, F. (1997). 'Was Prometheus Unbound by Chance? Risk, Diversification and Growth', *Journal of Political Economy*, vol. 105, pp. 709-751.

[3] Acemoglu, D. and Zilibotti, F. (1999). ' Information Accumulation in Development', *Journal of Economic Growth*, vol. 4, pp. 5-38.

[4] Agnosteva, D., Anderson, J. and Yotov, Y. (2014). 'Intra-national Trade Costs: Assaying Regional Frictions', NBER WP No. 19872.

[5] Anderson, J. E., and Van Wincoop, E. (2004). 'Trade costs', *Journal of Economic Literature*, vol. 42, pp. 691–751.

[6] Antràs, P. (2005). 'Incomplete Contracts and the Product Cycle', *American Economic Review,* vol. 95, pp. 1054-1073.

[7] Atkin, D. and Donaldson, D. (2015). 'Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs', Mimeo, MIT.

[8] Barro, R. and Lee, J. (2013). 'A New Data Set of Educational Attainment in the World, 1950-2010', *Journal of Development Economics,* vol 104, pp.184-198.

[9] Beck, T. (2002). 'Financial Development and InternationalTrade. Is There a Link?', *Journal of International Economics*, vol. 57, 107-131.

[10] Becker, B., Chen, J. and Greenberg, D. (2012). 'Financial development, fixed costs, and international trade', *Review of Corporate Finance Studies*, vol. 2, pp. 1-28.

[11] Blanchard, O. and Kremer, M. (1997). 'Disorganization', *Quarterly Journal of Economics*, Vol. 112, pp. 1091-1126.

[12] Coşar, K. and Fajgelbaum, P. (2016). 'Internal Geography, International Trade, and Regional Specialization', *American Economic Journal: Microeconomics*, vol. 8, pp. 24-56.

[13] Costinot, A. (2009), 'On the Origins of Comparative Advantage', *Journal of International Economics*, vol. 77, pp 255-264.

[14] Donaldson, D. (forthcoming). 'Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure', *American Economic Review.*

[15] Dornbusch, R., Fischer, S. and Samuelson, P. (1977). 'Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods', *American Economic Review*, vol. 67, pp. 823-839.

[16] Duranton, G., Morrow, P. and Turner, M. (2014). 'Roads and Trade: Evidence from the US', *Review of Economic Studies*, vol. 81, pp. 681-724.

[17] Fajgelbaum, P. and Redding, S. (2014). 'External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914', NBER WP No. 20217.

[18] Gaulier, G. and Zignago, S. (2010). 'BACI: International Trade Database at the Product-level. The 1994-2007 Version', CEPII Working Paper 2010-23.

[19] Greenwood, J. and Jovanovic, B. (1990). 'Financial Development, Growth, and the Distribution of Income', *Journal of Political Economy*, vol. 98, pp. 1076-1107.

[20] Hillberry, R. and Hummels, D. (2008). 'Trade responses to geographic frictions: A decomposition using micro-data', *European Economic Review*, vol. 52, pp. 527-550.

[21] Imbs, J. and Wacziarg, R. (2003). 'Stages of Diversification', *American Economic Review*, vol. 93, pp. 63-86.

[22] Levchenko, A. (2007). 'Institutional Quality and International Trade', *Review of Economic Studies*, vol. 74, pp. 681-724.

[23] Limao, N. and Venables, A. (2001). 'Infrastructure, geographical disadvantage, transport costs, and trade', *World Bank Economic Review*, vol. 15, pp. 451–479.

[24] Manova, K. (2013). 'Credit Constraints, Heterogeneous Firms, and International Trade', *Review of Economic Studies*, vol. 80, pp. 711-744.

[25] Matsuyama, K. (2017). 'Geographical Advantage: Home Market Effect in a Multi-Region World', *Research in Economics*, vol. 71, pp. 740-758.

[26] Nunn, N. (2007). 'Relationship-specificity, incomplete contracts, and the pattern of trade', *Quarterly Journal of Economics*, vol. 122, pp.569-600.

[27] Ramondo, N., Rodriguez-Clare, A. and Saborio-Rodriguez, M. (2016). 'Trade, Domestic Frictions, and Scale Effects', *American Economic Review*, vol. 106, pp. 3159-84.

[28] Redding, S. (2016). 'Goods Trade, Factor Mobility and Welfare', *Journal of International Economics*, vol. 101, pp. 148-167.

[29] Svaleryd, H. and Vlachos, J. (2005) 'Financial markets, the pattern of industrial specialization and comparative advantage: Evidence from OECD countries', *European Economic Review*, vol. 49, pp. 113-144.

[30] Volpe Martincus, C. and Blyde, J. (2013). 'Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment', *Journal of International Economics*, vol. 90, pp. 148-161.

[31] Volpe Martincus, C., Carballo, J. and Cusolito, A. (2017). 'Roads, exports and employment: Evidence from a developing country', *Journal of Development Economics*, vol. 125, pp. 21-39.