

# Roadways, Input Sourcing, and Patterns of Specialization\*

Esteban Jaimovich<sup>†</sup>

September 2019

## Abstract

We propose a model where the internal transport network facilitates the sourcing of intermediate goods from geographically diffuse locations. A denser internal transport network promotes thus the growth of industries that rely on a large variety of inputs. The model shows that heterogeneities in internal transport infrastructures can become a key factor in shaping comparative advantage and specialization. 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 broader 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 specialization in complex goods. Furthermore, we show that a similar correlation arises as well when the density of the local transport network is measured by the density of their internal waterways, and also when road density is instrumented with measures of terrain roughness.

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

**JEL Classifications:** F11, O18, R12

---

\*I would like to thank the editor, the associate editor, two anonymous referees, Facundo Albornoz, Ainhoa Aparicio-Fenoll, Holger Breinlich, Giovanni Mastrobuoni, Vincenzo Merella, Ignacio Monzon, Ricardo Nunes, Joao Santos Silva, and Alexandra Scudo as well as seminar participants at the Collegio Carlo Alberto, University of Nottingham, University of Stirling, University of Surrey, the RES Meeting at University of Warwick and the DEGIT XXIV conference (Odense) for helpful comments and suggestions.

<sup>†</sup>University of Surrey. Mailing address: School of Economics, Guildford, GU2 7XH, United Kingdom. Email: e.jaimovich@surrey.ac.uk

# 1 Introduction

Transport costs represent a major factor influencing countries' trade flows and specialization. While this factor has been long and widely acknowledged by the trade literature, its attention has almost exclusively been centered on the impact of international shipping costs.<sup>1</sup> However, the evidence at hand suggests that internal transport costs are far from being a secondary component of total transportation costs.<sup>2</sup> In addition, the magnitude of internal transport costs tends to vary quite substantially across countries owing to large differences in their local transport infrastructures. As a result, significant disparities in internal transport costs between economies tend to be observed, which in turn translates into variations in the cost of production of goods in different countries.

While the presence of a sufficiently dense internal transport network is crucial to keep total costs low, certain industries are inherently more reliant on the transport network than others. This paper studies a specific channel whereby the density of the internal transport network shapes countries' comparative advantages and patterns of specialization. One key role of the internal transportation network is that it facilitates the sourcing of intermediate inputs from geographically diffuse locations. As a consequence, industries that require a broader variety of intermediate inputs tend to make more intense use of the network.<sup>3</sup>

To illustrate this idea, we introduce a simple model with two intermediate sectors and a continuum of final good industries. Activities are geographically dispersed as a result of the presence of regional comparative advantages. A denser road network allows cheaper transportation of the intermediate inputs to the location site of final good producers. A crucial feature of the model is that final goods differ in terms of the breadth of their intermediate input requirements. In particular, some final industries rely very intensively in only one intermediate sector, while others require a more balanced mix of the two intermediate sectors. Since transportation of inputs is costly, those industries that require a relatively balanced combination of the intermediate inputs turn out to benefit relatively more from a denser internal road network.

This simple mechanism yields a very clear prediction in terms of specialization within a

---

<sup>1</sup>For a few papers that have incorporated internal transport costs into international trade models, see Allen and Arkolakis (2014), Felbermayr and Tarasov (2015), Coşar and Fajgelbaum (2016), Ramondo, Rodriguez-Clare and Saborio-Rodriguez (2016), Redding (2016), Matsuyama (2017).

<sup>2</sup>See, e.g., Limao and Venables (2001), Anderson and Van Wincoop (2004), Hillberry and Hummels (2008), Mesquita Moreira *et al* (2013), Atkin and Donaldson (2015), Donaldson (2018), Agnosteva *et al* (2019).

<sup>3</sup>This idea was first suggested by Clague (1991a, 1991b) who argued that countries with poor infrastructure will specialize in 'self-contained' sectors (i.e., sectors that do not intensively rely on inputs from other sectors).

framework with open economies. Countries that enjoy a denser local transport network tend to display a comparative advantage in the goods whose production process requires a broader mix of the intermediate inputs. This is because these are the industries that make heavier use of the local transport network to source their inputs. Conversely, countries with underdeveloped transport networks tend to specialize in industries with narrow input bases, as this allows them to economize on input sourcing costs.

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 local transport networks of countries by the length of their roadways per square kilometer. Finally, we correlate countries' specialization 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 broader input bases.

The correlation between road density and specialization in industries with broader 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 contracts enforcement [Levchenko (2007) and Nunn (2007)]. We show that the correlation predicted by our model is still present once we also control for the effect of the rule of law. 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 broad input bases seem also to be strongly reliant on the internal transport network.

One additional concern is whether the found correlation can be interpreted at all as evidence of *causation* from road density to specialization in industries more strongly reliant on the transport network. Roadways are the result of investment choices. Hence, road infrastructure may positively respond to transport needs resulting from patterns of specialization, reversing 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 of lower-income countries, which are exactly the types of economies that tend to suffer from sparser road networks.

Arguably, while waterways cannot be molded and expanded as flexibly as road networks,

and hence they are less sensitive to issues of reverse causation, their evidence does not directly address this concern. In order to address more directly the possibility that reverse causality is behind our empirical results, drawing on Ramcharan (2009), we also instrument the density of a country’s road network with topographical measures of terrain roughness.<sup>4</sup> The instrumental variable approach confirms the previous findings, granting further support to the hypothesis that the density of the internal road network is an important determinant of comparative advantage in industries with broader input bases.

There is a growing literature studying the impact of the local transport infrastructure on international and intra-regional trade and specialization. For example, Volpe Martincus and Blyde (2013) study the access to foreign markets and international trade across regions in Chile, Coşar and Demir (2016) does so for Turkey, and Volpe Martincus, Carballo and Cusolito (2017) for Peru. Donaldson (2018) looked at reductions of price and output distortions across Indian regions after expansions of the local railroad network, and Donaldson and Hornbeck (2016) assess how the expansion of the railroad network in the US enhanced counties’ market access. 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, Duranton, Morrow and Turner (2014) and Coşar and Demir (2016) have tried to capture whether there is an effect of road infrastructure on certain types of industry specialization. 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. Our paper focuses on a different channel whereby the local transport infrastructure impacts comparative advantages: the notion that the spatial distribution of activities turns industries that need to source a large variety of intermediate inputs relatively more reliant on the internal transport network.

The internal transportation channel studied in this paper was first suggested by Clague (1991a, 1991b). There it is argued that poorer economies specialize in ‘self-contained’ sectors, as they lack a sufficiently developed infrastructure needed to sustain the production of industries requiring a large variety of inputs. These articles, however, did not articulate this hypothesis within an international trade model, nor did they empirically assess whether trade flows at the industry level are associated with actual measures of the internal transport network in a way consistent with it.<sup>5</sup> We formulate the hypothesis that the local transport infrastruc-

---

<sup>4</sup>Ramcharan (2009) shows that countries with rougher topography tend to exhibit less dense road networks. He argues that this is partly due to the impact of terrain roughness and grade variation on the cost of building and maintenance of transport networks.

<sup>5</sup>These articles provide evidence that the relative efficiency of underdeveloped economies is worse in industries

ture matters relatively more for industries with wider input bases within a trade model where transport costs, location choices, and comparative advantage are explicitly modeled. This leads to an endogenous determination of trade flows and specialization patterns, which respond to heterogeneities in transport infrastructures. In addition, we present evidence supporting the relevance of this mechanism exploiting cross-country variation in the density of road networks.<sup>6</sup>

Finally, our paper also relates to several strands of literature that have expanded upon the canonical Ricardian/Heckscher-Ohlin trade models based on heterogeneities in factor productivities/endowments. One set of papers have looked at enforcement institutions as a source of comparative advantage in industries producing complex goods requiring large variety of input-specific relationships [Antràs (2005), Acemoglu, Antràs and Helpman (2007), Levchenko (2007), Nunn (2007), Costinot (2009), Ferguson and Formai (2013)]. Another strand of literature has delved into the role of financial markets at fostering exports in industries that are heavy users of external finance [Beck (2002), Svaleryd and Vlachos (2005), Becker, Chen and Greenberg (2012), Manova (2013)]. Finally, another institutional source of comparative advantage is presented by Cuñat and Melitz (2012), who show that countries with more flexible labor market regulations tend to export more in industries subject to higher volatility.<sup>7</sup> Our paper highlights the impact of local infrastructures when industries differ in their degree of dependence on internal transportation of inputs.

The rest of the paper is organized as follows. Section 2 discusses the evidence on geographic clustering and coagglomeration of industries, which motivates some key assumptions in the model. Section 3 introduces the main features of the model in the case of a closed economy. Section 4 extends the model to a two-country setup, and derives the main predictions on trade flows. Section 5 contrasts the main predictions of the model with the data. Section 6 discusses some endogeneity issues and alternative interpretations of the results. Section 7 concludes. All relevant proofs can be found in the online appendix (Appendix A).

---

that rely on a large variety of intermediate inputs. While this could be the result of poorer economies having less developed transport networks, it could also be the result of other factors usually associated with underdeveloped economies, like weaker institutions, lower levels of human capital, etc.

<sup>6</sup>Yeaple and Golub (2007) show that the stock of roads affects total factor productivity and sectoral composition across 10 industries for a panel of 18 countries. While their analysis highlights that roads may be a source of comparative advantage in some industries, it does not link the effect of heterogeneities in transport infrastructures to specialization in industries with different degrees of input diversity.

<sup>7</sup>See Chor (2010) for a paper that aims at quantifying the importance of all these institutional sources of comparative advantage, alongside the more traditional ones stemming from heterogeneities in factor productivities and endowments as in Romalis (2004).

## 2 Geographic Clustering and Coagglomeration of Industries: Evidence in the Literature

The core argument in the paper rests on the idea that, in the presence of regional comparative advantages, industries that rely on a wider variety of intermediate inputs will turn out to be more sensitive to internal transportation costs. In particular, the need to source a broad variety of intermediate inputs exacerbates the impact of transportation costs because regional comparative advantages lead to a geographically dispersed distribution of input sources. Geographic dispersion means, in turn, that not all regions will be able to offer local producers the exact same variety of intermediate inputs. As a consequence, industries with broader input requirements will end up being more likely to need sourcing a larger share of their inputs from more distant locations.

The empirical evidence on industry localization shows that heterogeneous regional clustering is indeed a prevalent feature across different industries and countries. The most paradigmatic and often-cited examples are probably the high-tech industries Silicon Valley and the automobile industry in Detroit. Besides those specific examples, the first systematic attempt to measure the degree of concentration of industries in different regions was carried out by Ellison and Glaeser (1997) for U.S. manufacturing industries. They find that 446 out of 495 four-digit SIC sectors display excess geographic concentration, relative to the degree of geographic concentration that would be observed if firms in all industries would pick their locations following an *identical* random process, and after controlling for the degree of aggregate concentration of manufacturing industries in denser regions.<sup>8</sup> In other words, Ellison and Glaeser (1997) find that most of the manufacturing industries in the U.S. tend to regionally cluster their activities following idiosyncratic localization patterns.<sup>9</sup>

Alongside the uneven geographic clustering of industries, economies also tend to display clear coagglomeration patterns of certain groups of industries. This is a feature observed with particular clarity in the case of strongly-linked upstream-downstream sectors, as shown

---

<sup>8</sup>Evidence on geographically dispersed industry clustering is also consistently found for France by Maurel and Sedillot (1999), for the U.K. by Devereux, Griffith and Simpson (2004) and by Duranton and Overman (2005), for Japan by Mori, Nishikimi and Smith (2005), for Belgium by Bertinelli and Decrop (2005), for Germany by Koh and Riedel (2014), and for Canada by Behrens and Bougna (2015).

<sup>9</sup>Ellison and Glaeser (1997) ascribe the patterns of geographic localization by industries to the combined effect of regional (industry-specific) natural advantages and localized industry-specific spillovers. While the model presented in the next two sections relies on the effect of regional natural advantages to drive industry localization, it could also be extended to encompass localized industry-specific spillovers as well.

by Ellison and Glaeser (1997) and Ellison *et al* (2010) for manufacturing industries in the U.S., by Duranton and Overman (2008) for the U.K., and by Behrens (2016) for Canada. The presence of input-output links between upstream and downstream industries is in fact shown by Ellison *et al* (2010) to represent *quantitatively* the strongest factor leading to coagglomeration of industries. Furthermore, the main cause of such patterns of coagglomeration is argued to be that of economizing on transport cost of major inputs as a result of physical proximity.<sup>10</sup>

Two clear patterns of geographic clustering by industries therefore consistently emerge in the literature. The first is the heterogeneous concentration of different industries in different regions. The second is the coagglomeration of industries with strong input-output links. Such localization patterns are particularly effective at reducing transportation costs involved in sourcing the inputs required by industries with highly concentrated input requirements, and at shielding those industries from the effects of hikes in transportation costs.<sup>11</sup> In particular, geographic concentration of industries has been shown to be instrumental to lowering the distance to input sources [Hillberry and Hummels (2008) and Behrens and Sharunova (2015)], and that this is especially effective in the case of highly interlinked sectors [Cainelli and Iacobucci (2012)]. The flip side of the above patterns of localization is that industries relying on a broad variety of intermediate inputs will end up being either more likely to need sourcing many of them from more distant locations, or alternatively needing to source an important share of their intermediate inputs from less-efficient local producers.<sup>12</sup> In the next two sections, we build a model where such a mechanism will confer economies with better internal transport networks a comparative advantage in industries that rely on broader input bases.

---

<sup>10</sup>Transportation costs are not the only force leading to coagglomeration of industries. Marshallian externalities can also generate similar co-location patterns between strongly-linked industries.

<sup>11</sup>See, e.g., Behrens and Brown (2018) who show that a rise in internal transportation costs in Canada leads to further geographic dispersion of industry pairs with weak input-output links, while it has very little impact on the geographic localization of industry pairs with strong input-output links, given that strongly-linked industries tend to already be localized together, and thus well protected against hikes in input transportation costs.

<sup>12</sup>See Bernard, Moxnes and Saito (2019) for evidence that Japanese firms sourcing a larger number of input suppliers tend to do that from longer average distances, and that improvements in transport infrastructure is especially beneficial to the performance of those types of firms. See also Gibbons *et al* (2019) for evidence of how improvements in road infrastructure in the U.K. leads to higher productivity in industries that switch to sourcing intermediate inputs from more efficient providers as a result of lower internal transportation costs, and Datta (2012) and Gunasekera *et al* (2008) for similar evidence in the case of India and Sri Lanka, respectively. In contrast, Ramcharan (2009) shows that countries with less dense internal transport networks tend to concentrate more activities in fewer geographic locations, so as to compensate for their higher transportation costs resulting from poorer internal geographic connectivity.

## 3 A Closed Economy with Internal Transportation Cost

This section presents the environment and main features of the model in the specific case of a closed economy. This proves helpful for two reasons. First, it allows an easier description of the main building blocks of the model. Second, it facilitates the exposition of the main intuition for how the density of the transport network may heterogeneously affect the cost of production in sectors with different degrees of input breadth.

### 3.1 Setup and Environment

#### 3.1.1 Geography and Population

The economy is inhabited by a continuum of individuals with mass  $L$ . Each individual is endowed with one unit of labor time that is supplied inelastically to firms. The economy comprises two geographically separate regions denoted by  $R = A, B$ . Region  $R$  is inhabited by a mass of individuals  $L_R$ , where  $L_A + L_B = L$ . There is free labor mobility within the economy. For brevity, throughout the paper, we abstract from explicitly modelling the determination of  $L_A$  and  $L_B$ . However, those two variables could be thought of as resulting from the underlying equilibrium in the regional labor markets together with individuals' optimal location choices.<sup>13</sup>

Both regions  $A$  and  $B$  are, in principle, able to host the production of any existing intermediate and final goods. Production activities are carried out by firms, which must choose a specific location. All markets are assumed to be perfectly competitive.

#### 3.1.2 Intermediate Sectors

There exist two intermediate sectors (or industries), indexed by  $S = 0, 1$ . Each sector  $S$  comprises a unit continuum of intermediate goods (or varieties), which we index by  $i_0 \in [0, 1]$  and  $i_1 \in [0, 1]$ , for  $S = 0$  and  $S = 1$ , respectively. Whenever creating no confusion, we will skip the subscript  $S = 0, 1$  from  $i_{S=0,1}$ , and index intermediate goods in either sector by a generic  $i \in [0, 1]$ . Henceforth, we will refer to each  $i$  in sector  $S$  interchangeably as intermediate good  $i$  or variety  $i$  of sector  $S$ .

The technology in the intermediate sectors draws heavily from Eaton and Kortum (2002).<sup>14</sup> Producing one unit of variety  $i$  in sector  $S$  in region  $R$  requires  $1/Z_{R,S}(i)$  units of labor, where

<sup>13</sup>In Appendix D we explicitly show how  $L_A$  and  $L_B$  may arise as equilibrium objects in a context where individuals optimally choose the specific region where to live, consume, and work.

<sup>14</sup>Other models with an internal geography within an Eaton-Kortum framework can be found in Ramondo *et al* (2016), Redding (2017), Donaldson (2018), and Caliendo *et al* (2018).



each  $Z_{R,S}(i)$  is the result of an independent draw from a Fréchet distribution with location parameter  $T_{R,S} > 0$  and shape parameter  $\theta > 1$ . That is,

$$F_{R,S}(Z) = \exp(-T_{R,S} Z^{-\theta}). \quad (1)$$

For the rest of the paper, we assume:

**Assumption 1** *i)*  $T_{A,0} = T_{B,1} = 1 + T$ , where  $T > 0$ ; *ii)* and  $T_{A,1} = T_{B,0} = 1$ .

Assumption 1 entails that region  $A$  will display (on average) higher productivity in intermediate varieties belonging to sector 0, and region  $B$  will exhibit (on average) higher productivity in intermediate varieties of sector 1.<sup>15</sup> The main intention behind Assumption 1 is to create room for some degree of regional specialization across the two intermediate sectors.<sup>16</sup>

### 3.1.3 Final Goods Sector

In addition to the intermediate good sectors (and their respective varieties), there also exists a unit continuum of final goods, indexed by  $j \in [0, 1]$ . Final goods are purchased by individuals with preferences given by

$$U = \int_0^1 \ln(y_j) dj, \quad (2)$$

where  $y_j$  denotes the consumed amount of  $j$ .

Final goods are produced by combining intermediate good varieties within two-level structured production functions, where sectoral CES aggregators of intermediate varieties are nested into Cobb-Douglas production functions of the sectoral aggregators. In particular, total output of final good  $j \in [0, 1]$  is given by:

$$Y_j = \Psi_j \left[ \left( \int_0^1 x_{0,i}^{(\sigma-1)/\sigma} di \right)^{\frac{\sigma}{\sigma-1}} \right]^{1-\alpha_j} \left[ \left( \int_0^1 x_{1,i}^{(\sigma-1)/\sigma} di \right)^{\frac{\sigma}{\sigma-1}} \right]^{\alpha_j}, \quad (3)$$

$$\text{where: } \alpha_j \in [0, 1], \quad \Psi_j \equiv \alpha_j^{-\alpha_j} (1 - \alpha_j)^{-(1-\alpha_j)}, \quad \text{and } \sigma > 1.$$

---

<sup>15</sup>The implicit symmetry across regions' productivity in sectors 0 and 1 implied by Assumption 1 is posed to keep the model's structure as simple as possible. Our main qualitative results will carry over under more general assumptions, as long as each region retains *some* degree of aggregate advantage in one of the sectors. In particular, we could replace Assumption 1 by a more general setting, satisfying: *i)*  $T_{A,0} > T_{A,1}$ ; *ii)*  $T_{B,0} < T_{B,1}$ ; *iii)*  $T_{A,0} > T_{B,0}$ ; *iv)*  $T_{A,1} < T_{B,1}$ ; and the main results of the model will remain essentially unchanged.

<sup>16</sup>Notice, however, that since each  $Z_{R,S}(i)$  is independently drawn from a Fréchet distribution, both regions will end up producing in equilibrium a *positive* mass of intermediate varieties in *both* sectors (in other words, it is only the *degree* of specialization across intermediate varieties from sector 0 and 1 that will differ in equilibrium between regions  $A$  and  $B$ ).

In (3),  $x_{0,i}$  and  $x_{1,i}$  denote, respectively, the amount of intermediate variety  $i$  from sector 0 and sector 1 used in the production of final good  $j$ . The parameter  $\sigma$  denotes the elasticity of substitution between varieties of intermediate goods within each intermediate sector.

A crucial feature of the production functions in (3) is the fact that the Cobb-Douglas weights  $1 - \alpha_j$  and  $\alpha_j$  are assumed to be sector-specific. This implies that final good sectors may differ in terms of the intensity requirements of inputs from each of the two intermediate sectors. Final sectors with a small  $\alpha_j$  (resp. a large  $\alpha_j$ ) will tend to use intermediate varieties from sector 0 (resp. sector 1) more intensively. On the other hand, sectors whose  $\alpha_j$  lies close to 0.5 will tend to use a relatively balanced mix of varieties from both intermediate sectors.

For the remainder of the paper, we will assume that, when considering the whole unit continuum of final goods, the values of  $\alpha_j$  will turn out to be uniformly distributed within the unit interval. Abusing a bit the notation, we can thus index a generic final good  $j$  by the specific value of  $\alpha_j \in [0, 1]$ .

### 3.1.4 Internal Transportation Cost

We assume that transporting intermediate goods across the two regions entails an iceberg trade cost  $d_{A,B} = d_{B,A} = d > 1$ , where  $d_{R,-R}$  indicates the amount of intermediate good that must be shipped from  $R$  to  $-R$  for one unit of the good to arrive in  $-R$ . In the cases of intra-regional trade, transportation of intermediate goods is assumed to be costless; that is,  $d_{A,A} = d_{B,B} = 1$ .

Regions  $A$  and  $B$  are assumed to be linked to each other by an internal road network of density (or length)  $r > 0$ . A denser internal road network lowers the interregional trade cost of intermediate goods.<sup>17</sup> Namely,

**Assumption 2**  $d = d(r)$ , where  $d(r) > 1$  and  $d'(r) < 0$ , for all  $r \in \mathbb{R}_+$ .

With regards to final goods, to simplify the analysis, we will assume that they can be *costlessly* transported between region  $A$  and  $B$ . In our context, costless transportation of final goods will imply that, in equilibrium, the wage must be the same in  $A$  and  $B$ . Appendix D describes the implications of introducing also an iceberg trade costs for interregional transportation of final goods.<sup>18</sup>

---

<sup>17</sup>See Felbermayr and Tarasov (2015), Fajgelbaum and Schaal (2017), Santamaria (2018), and Allen and Arkolakis (2019), for papers that explicitly model how increases in the density of the internal transportation infrastructure leads to lower transportation costs of goods.

<sup>18</sup>The reason for the wage to be the same in  $A$  and  $B$  is that, since utility (2) depends only on consumption of final goods, in the absence of internal trade cost of final goods, all final goods will be priced identically in both

### 3.2 Location Choice and Cost of Production of Final Goods

Perfect competition implies that, in equilibrium, each final good  $j$  will be sold at a price equal to its marginal cost. Using (3), we can observe that the marginal cost of final good  $j$  produced in region  $R$ , denoted by  $c_{j,R}$ , is given by

$$c_{j,R} = p_{0,R}^{1-\alpha_j} p_{1,R}^{\alpha_j}, \quad (4)$$

where  $p_{0,R}$  and  $p_{1,R}$  are, respectively, the CES price indices of intermediate inputs from sector 0 and 1, faced by a firm located in region  $R$ .

Letting  $w$  denote the wage per unit of labor, and using (1), Assumption 1,  $d_{A,B} = d_{B,A} = d(r)$  and  $d_{A,A} = d_{B,B} = 1$ , the results in Eaton and Kortum (2002) applied to our context imply that for firms located in  $A$ :

$$p_{0,A} = \gamma w [(1+T) + d(r)^{-\theta}]^{-\frac{1}{\theta}} \quad \text{and} \quad p_{1,A} = \gamma w [1 + (1+T) d(r)^{-\theta}]^{-\frac{1}{\theta}}, \quad (5)$$

while for firms located in  $B$ :

$$p_{0,B} = \gamma w [(1+T) d(r)^{-\theta} + 1]^{-\frac{1}{\theta}} \quad \text{and} \quad p_{1,B} = \gamma w [d(r)^{-\theta} + (1+T)]^{-\frac{1}{\theta}}, \quad (6)$$

where  $\gamma \equiv [\Gamma(1 + \frac{1-\sigma}{\theta})]^{1/(1-\sigma)}$  and  $\Gamma(\cdot)$  is the Gamma function. Henceforth, we restrict the parameters to satisfy  $\sigma < 1 + \theta$ , so as to ensure having well-defined price indices.

From (4), (5) and (6), it follows that the marginal costs of  $j$ , depending on where it is being produced, are given by

$$c_{j,A} = \gamma w [(1+T) + d(r)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [1 + (1+T) d(r)^{-\theta}]^{-\frac{\alpha_j}{\theta}} \quad (7)$$

and

$$c_{j,B} = \gamma w [1 + (1+T) d(r)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [(1+T) + d(r)^{-\theta}]^{-\frac{\alpha_j}{\theta}}. \quad (8)$$

Since final goods are assumed to be transported costlessly between regions, each  $j$  will be produced in the region with the lower marginal cost.<sup>19</sup> Comparing  $c_{j,A}$  vis-a-vis  $c_{j,B}$  for different values of  $\alpha_j$  yields the following result:

---

regions, and thus free internal mobility will require earnings to be equalized in  $A$  and  $B$  for individuals to be indifferent. Appendix D shows that the presence of interregional trade cost for final goods does not *necessarily* lead to unequal wages in  $A$  and  $B$ . In particular, given the symmetric structure across regions implied by Assumption 1 and the log utility function (2), when all final goods are subject to the same iceberg cost, the internal equilibrium will still entail an identical wage in regions  $A$  and  $B$ . Under alternative assumptions the equilibrium wage may well differ across regions. Yet, the main qualitative results of the model will in general still hold true even in the presence of interregional income inequality.

<sup>19</sup>This result no longer holds true when final goods are also subject to internal transport cost. In those cases, some final goods will end up being produced in both regions and will be only sold locally, instead of being traded between regions – see Lemma ?? in Appendix D.

**Lemma 1** *In equilibrium, the geographic distribution of final good producers is as follows: i) all producers of final goods for which  $\alpha_j < 0.5$  locate in region A; ii) all producers of final goods for which  $\alpha_j > 0.5$  locate in region B; iii) the producers of the final good for which  $\alpha_j = 0.5$  are indifferent between the two regions, and choose randomly their location.*

The result in Lemma 1 states an intuitive geographic agglomeration result. Final good producers choose to locate in the region that exhibits a comparative advantage in the intermediate inputs they use more intensively. Intuitively, by setting up their firms in those regions, final good producers are able to economize relatively more on the transportation of the inputs they more strongly rely on.

Using the result in Lemma 1, together with (7) and (8), we can now write down the marginal cost of final good  $j$ :

$$c_j(r) = \begin{cases} \gamma w [(1+T) + d(r)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [1 + (1+T)d(r)^{-\theta}]^{-\frac{\alpha_j}{\theta}} & \text{if } \alpha_j \leq \frac{1}{2}, \\ \gamma w [1 + (1+T)d(r)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [(1+T) + d(r)^{-\theta}]^{-\frac{\alpha_j}{\theta}} & \text{if } \alpha_j \geq \frac{1}{2}. \end{cases} \quad (9)$$

The expressions in (9) show that the marginal cost of final good  $j$  is affected by the internal transport cost of intermediate inputs that are *not* sourced locally via the term  $d(r)$ . As Lemma 1 shows, final good producers locate in the region that is the *main* home of the intermediate sector that they use more intensively so as to minimize the number of inputs they must source from the *other* region. However, while this agglomeration pattern allows economizing on transportation costs, it does not do so equally across all sectors. In particular, locating in region A is especially beneficial for final sectors whose  $\alpha_j$  is very close to zero, whereas locating in B is so for those whose  $\alpha_j$  is very close to one. Instead, when  $\alpha_j$  lies near 0.5, locating production in either A or B will not drastically cut down on total transportation costs, since those sectors ultimately rely on a balanced combination of intermediate varieties from sector 0 and 1. As a result, internal transport costs will tend to affect more severely those sectors that use a relatively balanced combination of inputs. Yet, an interesting flip side of this argument is that improvements in  $r$  will end up benefiting relatively more exactly those sectors with intermediate values of  $\alpha_j$ . The following lemma states this result more formally.

**Lemma 2** *Consider two generic values of the internal road network density:  $r_1 < r_2$ . Then,*

1.  $c_j(r_1)/c_j(r_2) > 1$  for all  $\alpha_j \in [0, 1]$ .
2. The ratio  $c_j(r_1)/c_j(r_2)$  is strictly increasing in  $\alpha_j$  for all  $\alpha_j \in [0, \frac{1}{2})$  and strictly decreasing in  $\alpha_j$  for all  $\alpha_j \in (\frac{1}{2}, 1]$ . Moreover, the highest value of  $c_j(r_1)/c_j(r_2)$  is reached at  $\alpha_j = \frac{1}{2}$ .

Lemma 2 shows that larger values of  $r$  lead to lower costs of production for all final goods. However, the drop in  $c_j$  resulting from a higher  $r$  turns out to be proportionally greater in sectors with values of  $\alpha_j$  closer to  $\frac{1}{2}$ . This result will prove key in the next section where we allow for international trade, and it will imply that the density of the internal road network will become a source of comparative advantage. In particular, countries with denser road networks will tend to enjoy a comparative advantage in final sectors that rely on a relatively balanced mix of intermediate inputs (i.e., those characterized by values of  $\alpha_j$  around  $\frac{1}{2}$ ).

## 4 Two-Country Model

We now incorporate our previous framework into a world economy with two countries:  $H$  and  $F$ . We initially keep most of the main features of the environment in Section 3 essentially intact. Some of these specific features are subsequently relaxed in Section 4.2 and in the online appendix, to show how the main results extend to more general setups.

**Geography and Population:** Country  $H$  and country  $F$  are both characterized by an identical internal geography, comprising two separate regions:  $A_H$  and  $B_H$  in  $H$ ;  $A_F$  and  $B_F$  in  $F$ . Whenever it creates no confusion, we will skip the use of country subscripts, and simply refer to them as region  $A$  or  $B$  in country  $C = H, F$ . Both countries are populated by a mass  $L$  of individuals. Each individual is endowed with one unit of labor that is supplied inelastically to local firms. There is free labor mobility within countries. In contrast, labor is *immobile* between countries. We let  $w_H$  and  $w_F$  denote the wage in  $H$  and in  $F$ , respectively. Henceforth, we set  $w_F = 1$ , and use  $\omega \equiv w_H/w_F$  to denote the relative wage of  $H$  with respect to  $F$ . All individuals in the world economy share the same preferences, given by (2). All markets are subject to perfect competition everywhere.

**Final Goods Sectors:** Technologies to produce final goods differ between  $H$  and  $F$ . The production function of final good  $j \in [0, 1]$  in country  $C = H, F$  is given by:

$$Y_{j,C} = \zeta_{j,C} Y_j, \tag{10}$$

where  $Y_j$  is given by (3). We assume that the country specific parameters  $\zeta_{j,H}$  and  $\zeta_{j,F}$  are independently drawn from a Fréchet distribution function with location parameter equal to 1 and shape parameter  $\vartheta > 1$ :

$$F_C(\zeta) = \exp(-\zeta^{-\vartheta}). \tag{11}$$

All final goods are internationally tradeable, subject to an iceberg trade cost. Namely,  $\tau > 1$

units of final good  $j$  must be shipped from the exporter of  $j$  in order for the importer to receive one unit of  $j$ .<sup>20</sup>

**Intermediate Sectors:** Unlike for final goods, we assume that intermediate goods are all non-tradeable internationally.<sup>21</sup> The technology in the intermediate sectors is directly extrapolated from Section 3.1.2: producing one unit of variety  $i$  of sector  $S$  in region  $R_C = A_C, B_C$  (with  $C = H, F$ ) requires  $1/Z_{R_C,S}(i)$  units of labor, where each  $Z_{R_C,S}(i)$  is independently drawn from a Fréchet distribution like (1). Regarding each  $T_{R_C,S}$ , we assume that they are given in both  $H$  and  $F$  by Assumption 1.<sup>22</sup>

**Internal Transportation Cost:** Shipping intermediate goods between  $C$ 's regions entails an iceberg cost  $d(r_C)$ , where  $r_C$  is the density of  $C$ 's internal road network, with  $C = H, F$ . The properties of function  $d(\cdot)$  are still given by Assumption 2. Throughout the paper, we let  $H$ 's internal road network be denser than  $F$ 's. Namely,

**Assumption 3**  $r_H > r_F$ .

Assumption 3 will have two main implications in the context of our world economy model. First (and more importantly in our context), it will confer a source of comparative advantage to  $H$  in the types of final goods that rely more strongly on efficient internal transportation of inputs. Second, the fact that  $H$  will be able to ship inputs internally at a lower cost than  $F$  will, in turn, grant a source of aggregate absolute advantage by  $H$  over  $F$ .<sup>23</sup>

## 4.1 Equilibrium and Specialization with International Trade

Analogously to the closed-economy setup, in order to minimize marginal costs, final good producers in  $H$  and in  $F$  will locate in region  $A$  of their country when  $\alpha_j \leq 0.5$ , and in region

---

<sup>20</sup>Notice that this implicitly assumes that the international trade cost between *any* pair of foreign regions is equal to  $\tau$ . In other words, we disregard the possibility that one of the two regions is closer to international markets than the other one, like in models that study the division of industries between (export-oriented) coastal regions and (local-market-oriented) interior regions – e.g., Coşar and Fajgelbaum (2016). Allowing the internal geography of our model to feature a coastal and an interior region will certainly have implications in terms of regional specialization. However, this will not alter the main predictions of our model regarding patterns of *international* trade and specialization.

<sup>21</sup>This assumption is relaxed in Appendix E (online appendix), where we allow also intermediate goods to be traded internationally.

<sup>22</sup>More rigorously, (1) should be now written  $F_{R_C,S}(Z) = \exp(-T_{R_C,S}Z^{-\theta})$ , with  $C = H, F$ . Similarly, Assumption 1 should now state:  $T_{A_C,0} = T_{B_C,1} = 1 + T$ , with  $T > 0$ , and  $T_{A_C,1} = T_{B_C,0} = 1$ .

<sup>23</sup>Like in Section 3, we keep assuming costless internal transportation of final goods. Relaxing this assumption would not alter any of the main results of the model.

$B$  of their country when  $\alpha_j \geq 0.5$ . Should final good  $j$  be produced *somewhere* in country  $C = H, F$ , it would then be produced with a marginal cost:

$$c_j^C = \begin{cases} \frac{1}{\zeta_{j,C}} \gamma w_C [(1+T) + d(r_C)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [1 + (1+T) d(r_C)^{-\theta}]^{-\frac{\alpha_j}{\theta}} & \text{if } \alpha_j \leq \frac{1}{2} \\ \frac{1}{\zeta_{j,C}} \gamma w_C [1 + (1+T) d(r_C)^{-\theta}]^{-\frac{1-\alpha_j}{\theta}} [(1+T) + d(r_C)^{-\theta}]^{-\frac{\alpha_j}{\theta}} & \text{if } \alpha_j \geq \frac{1}{2}, \end{cases} \quad (12)$$

where the superscript  $C$  in  $c_j^C$  denotes the country of production of  $j$ .

To ease notation, we will let henceforth  $d_C \equiv d(r_C)$ , and denote by  $\lambda > 1$  the ratio of the internal trade cost in  $F$  relative to  $H$  (where  $\lambda > 1$  follows from Assumption 3):

$$\lambda \equiv \frac{d_F}{d_H} > 1. \quad (13)$$

Also, to ease notation, we will henceforth re-write the expressions in (12) simply as

$$c_j^C = \frac{1}{\zeta_{j,C}} \chi_{j,C}, \quad \text{with } C = H, F. \quad (14)$$

where  $\chi_{j,C}$  denotes the total cost of the intermediate inputs used to produce final good  $j$  in country  $C$ .

Costly international trade implies that consumer prices will, in general, differ between  $H$  and  $F$ , and that not all final goods will necessarily be traded between  $H$  and  $F$  in equilibrium. Whether or not a specific final good  $j$  will be traded internationally will depend on the gap between  $c_j^H$  and  $c_j^F$  relative to the international shipping cost  $\tau$ . When  $j$  is produced *only* by country  $C$ , it will be sold in country  $M$  at a price equal to  $P_{j,M}^C = [\mathbb{I}_{M \neq C} \cdot \tau + (1 - \mathbb{I}_{M \neq C})] \cdot c_j^C$ , where  $\mathbb{I}_{M \neq C}$  is an indicator function that is equal to 1 when  $M \neq C$  and 0 otherwise, and  $c_j^C$  denotes the marginal cost of producing  $j$  in  $C$ . On the other hand, if  $j$  remains untraded, the price in  $H$  (resp.  $F$ ) will be  $P_{j,H}^H = c_j^H$  (resp.  $P_{j,F}^F = c_j^F$ ). Country  $H$  will then be exporting good  $j$  to  $F$  when  $\tau c_j^H < c_j^F$ , while  $H$  will be importing  $j$  from  $F$  when  $c_j^H > \tau c_j^F$ .

To pin down whether a specific final good  $j$  will be exported by  $H$ , by  $F$ , or not exported by any of the two countries, we use (14), coupled with the fact that the productivity parameters  $\zeta_{j,C}$  are all independently drawn from (11), and compare the ensuing ratio  $c_j^H/c_j^F$  with the transport cost  $\tau$ . The probability that country  $C$  ends up exporting final good  $j$  to the other country (i.e.,  $-C$ ), denoted by  $\pi_C(j)$ , will be given by

$$\pi_C(j) = \frac{1}{1 + \left( \frac{\chi_{j,-C}}{\tau \chi_{j,C}} \right)^{-\vartheta}}. \quad (15)$$

By using (15), together with (??), the first result we can obtain concerns the equilibrium relative wage,  $\omega^*$ .

**Proposition 1** *In equilibrium, the wage in  $H$  is strictly greater than in  $F$  (i.e.,  $\omega^* > 1$ ). Furthermore,  $\omega^*$  is strictly increasing in  $\lambda$  (i.e.,  $\partial\omega^*/\partial\lambda > 0$ ), and it converges to 1 as  $r_F$  approaches  $r_H$  (i.e.,  $\lim_{\lambda \rightarrow 1} \omega^*(\lambda) = 1$ ).*

The result  $\omega^* > 1$  is a direct implication of the fact that Assumption 3 conveys an aggregate advantage by  $H$  over  $F$ . Hence, in equilibrium,  $\omega$  must rise above one, in order to allow  $F$  to export enough to  $H$  and satisfy the trade balance.

Bearing in mind (15) and Proposition 1, we can now summarize the equilibrium patterns of trade and specialization delivered by the model, by linking the probability that either  $H$  or  $F$  ends up exporting a final good  $j$  with a given value of  $\alpha_j$ .

**Proposition 2** *Let  $\pi_C(\alpha_j)$  denote the probability that country  $C = H, F$  exports in equilibrium the final good with Cobb-Douglas weight parameter  $\alpha_j \in [0, 1]$  to the other country. Then:*

1.  $\partial\pi_H/\partial\alpha_j > 0$  for all  $0 \leq \alpha_j < \frac{1}{2}$ , and  $\partial\pi_H/\partial\alpha_j < 0$  for all  $\frac{1}{2} < \alpha_j \leq 1$ .
2.  $\partial\pi_F/\partial\alpha_j < 0$  for all  $0 \leq \alpha_j < \frac{1}{2}$ , and  $\partial\pi_F/\partial\alpha_j > 0$  for all  $\frac{1}{2} < \alpha_j \leq 1$ .
3. The ratio  $\pi_H(\alpha_j)/\pi_F(\alpha_j)$  reaches its highest value at  $\alpha_j = \frac{1}{2}$ . Moreover, in equilibrium,  $\pi_H(\alpha_j = \frac{1}{2}) > \pi_F(\alpha_j = \frac{1}{2})$ .

The pattern of specialization described by Proposition 2 represents the main insight of the model. Owing to the geographically diffuse distribution of input sources, the density of the internal road network becomes a source of comparative advantage across industries. Final goods with values of  $\alpha_j$  close to  $\frac{1}{2}$  require a relatively balanced use of all intermediate varieties. This, in turn, means that a large share of their inputs will *necessarily* have to be transported through the road network. On the other hand, firms producing final goods with either *high* or *low* values of  $\alpha_j$  are able to source a large share of their intermediate inputs from the *same* location where their firms are placed, turning them much less reliant on the internal transport network. As a result, when  $H$  has a denser road network than  $F$ , exports originating from  $H$  will tend to be overrepresented by final goods with intermediate values of  $\alpha_j$ , whereas those originating from  $F$  will tend to be overrepresented by final goods with values of  $\alpha_j$  close to both extremes of the unit interval.

The result in Proposition 2 describes the patterns of specialization across final goods differing in terms of their value of  $\alpha_j$ , for *given* levels of the road density of  $H$  and  $F$ . The next result



complements Proposition 2 by showing that those patterns of specialization get magnified as the value of  $r_H$  increases, while holding constant the value of  $r_F$ .<sup>24</sup>

**Proposition 3** *Consider the case of an increase in the road density of country  $H$  while the road density of country  $F$  remains constant. Then, the patterns of specialization across final goods described by Proposition 2 will further intensify. More precisely,*

1.  $\frac{\partial(\partial\pi_H/\partial r_H)}{\partial\alpha_j} > 0$  for all  $0 \leq \alpha_j < \frac{1}{2}$ , and  $\frac{\partial(\partial\pi_H/\partial r_H)}{\partial\alpha_j} < 0$  for all  $\frac{1}{2} < \alpha_j \leq 1$ .
2.  $\frac{\partial(\partial\pi_F/\partial r_H)}{\partial\alpha_j} < 0$  for all  $0 \leq \alpha_j < \frac{1}{2}$ , and  $\frac{\partial(\partial\pi_F/\partial r_H)}{\partial\alpha_j} > 0$  for all  $\frac{1}{2} < \alpha_j \leq 1$ .

Proposition 3 shows how  $H$ 's comparative advantage in the final goods that require a more balanced use of the different intermediate varieties gets further magnified as  $r_H$  becomes even larger relative to  $r_F$ . This differential effect on the impact on exports stemming from variations in road density across countries is one of the main results of the model that we will contrast with cross-country data on exports by industries in Section 5.

## 4.2 Extension: Multiple Intermediate Industries

This section extends the model in Section 4.1 to incorporate  $N$  intermediate sectors, where we will interpret  $N$  as a large number. For analytical convenience, and without any loss of generality, we assume that  $N$  is an even number. We keep the assumption that each of the two economies consists of two separate regions:  $A_C$  and  $B_C$ . In order to keep this section concise, we will work with a framework that imposes symmetry in terms of final goods production functions, and also in terms of regional comparative advantage across intermediate industries.

There is a continuum of final goods produced by combining intermediate varieties from  $N$  different industries. The set of final goods comprises  $N$  different subsets, each of them with mass equal to  $N/(N-1)$ .<sup>25</sup> We index each of these subsets by the letter  $k = 1, 2, \dots, N$ . We also index each specific final good belonging to  $k$  by the letter  $j_k$ .

The production function corresponding to final good  $j_k$  in country  $C$  is given by:

$$Y_{j_k, C} = \zeta_{j_k, C} \times \Psi_{j_k} \times \left[ \left( \int_0^1 x_{k,i}^{(\sigma-1)/\sigma} di \right)^{\frac{\sigma}{\sigma-1}} \right]^{\frac{1}{N} + \gamma_{j_k}} \times \prod_{l \neq k} \left[ \left( \int_0^1 x_{l,i}^{(\sigma-1)/\sigma} di \right)^{\frac{\sigma}{\sigma-1}} \right]^{\frac{1}{N} - \frac{\gamma_{j_k}}{N-1}}, \quad (16)$$

<sup>24</sup>Note from the definition of  $\lambda$  in (13) that increasing  $r_H$  while holding  $r_F$  constant is one particular way to generate an increase in  $\lambda$ .

<sup>25</sup>The assumption of total mass equal to  $N/(N-1)$  for each subset of final goods is just posed for analytical convenience, and could be dispensed without any loss of generality.

where:  $\gamma_{j_k} \in [0, \frac{N-1}{N}]$ , and  $\Psi_{j_k} \equiv (\frac{1}{N} + \gamma_{j_k})^{-\frac{1}{N} + \gamma_{j_k}} \times (\frac{1}{N} - \frac{\gamma_{j_k}}{N-1})^{-\frac{N-1}{N} - \gamma_{j_k}}$ .

We keep assuming that each  $\zeta_{j_k, C}$  in (16) is independently drawn from (11). In addition, we assume that for each subset of final goods  $k \in \{1, 2, \dots, N\}$ , the values of the parameter  $\gamma_{j_k}$  are uniformly distributed along the interval  $[0, \frac{N-1}{N}]$ .<sup>26</sup>

To interpret the production functions in (16), note that when  $\gamma_{j_k} = (N-1)/N$  final good  $j_k$  will rely *exclusively* on intermediate varieties originating from intermediate sector  $k$ . Conversely, when  $\gamma_{j_k} = 0$ , final good  $j_k$  will rely *equally* on intermediate varieties from *all* the different intermediate sectors. Naturally, the closer the value of  $\gamma_{j_k}$  is to  $N/(N-1)$ , the stronger the importance of sector  $k$  as the *main* source of intermediate inputs for final good  $j_k$ .

Regarding the intermediate sectors, we keep assuming that one unit of intermediate variety  $i$  in sector  $k$  in region  $R_C$ , requires  $1/Z_{R_C, k}(i)$  units of labor, where each  $Z_{R_C, k}(i)$  is the result of an independent draw from a Fréchet distribution with location parameter  $T_{R_C, k} > 0$  and shape parameter  $\theta > 1$ . In order to maintain the implicit symmetry across regions previously determined by Assumption 1, we now postulate the following:

**Assumption 1 (bis)** *i*)  $T_{A_C, k} = 1 + T$  if  $k$  is an odd number (i.e.,  $k = 1, 3, \dots, N-1$ ), and  $T_{A_C, k} = 1$  if  $k$  is an even number (i.e.,  $k = 2, 4, \dots, N$ ); *ii*)  $T_{B_C, k} = 1$  if  $k$  is an odd number (i.e.,  $k = 1, 3, \dots, N-1$ ), and  $T_{B_C, k} = 1 + T$  if  $k$  is an even number (i.e.,  $k = 2, 4, \dots, N$ ).

Assumption 1 (bis) ensures that each of the two regions will exhibit, on average, higher labor productivity in half of the  $N$  intermediate sectors. All final good producers of country  $C$  belonging to the subset of final goods  $k = 1, 3, \dots, N-1$  (resp.  $k = 2, 4, \dots, N$ ) will locate in region  $A_C$  (resp. in region  $B_C$ ). As a result of this, the marginal cost of producing final good  $j_k$  in country  $C = H, F$ , will be given by:

$$c_{j_k}^C = \frac{1}{\zeta_{j_k, C}} \gamma w_C \left\{ [(1+T) + d_C^{-\theta}]^{-\frac{1}{\theta}} \right\}^{\frac{1+N}{2N} + \frac{\gamma_{j_k}}{2}} \left\{ [1 + (1+T) d_C^{-\theta}]^{-\frac{1}{\theta}} \right\}^{\frac{1}{2} - \frac{N}{N-1} \frac{\gamma_{j_k}}{2}}. \quad (17)$$

Denoting again with  $\chi_{j_k, C}$  the total cost of intermediate inputs used for final good  $j_k$  in country  $C$  (i.e.,  $\chi_{j_k, C} = c_{j_k}^C \cdot \zeta_{j_k, C}$ ), by using (17) we now obtain:

$$\frac{\chi_{j_k, H}}{\chi_{j_k, F}} = \omega \left[ \frac{(1+T) + (\lambda d_H)^{-\theta}}{(1+T) + d_H^{-\theta}} \right]^{\frac{1}{\theta} (\frac{1+N}{2N} + \frac{1}{2} \gamma_{j_k})} \left[ \frac{1 + (1+T) (\lambda d_H)^{-\theta}}{1 + (1+T) d_H^{-\theta}} \right]^{\frac{1}{\theta} (\frac{1}{2} - \frac{N}{2(N-1)} \gamma_{j_k})}. \quad (18)$$

From (18), two important properties immediately follow: *i*)  $\chi_{j_k, H} / \chi_{j_k, F} < 1$  if  $\omega = 1$ ; *ii*)  $\partial(\chi_{j_k, H} / \chi_{j_k, F}) / \partial \gamma_{j_k} > 0$ . These two properties will lead, respectively, to results analogous

<sup>26</sup>Notice that this implies that the total mass of final goods (summing up all subsets of final goods) equals  $N$ .

to those in Proposition 1 and Proposition 2. In particular, the fact that  $\chi_{j_k,H}/\chi_{j_k,F} < 1$  when  $\omega = 1$  will mean that in equilibrium we must have  $\omega^* > 1$ . On the other hand,  $\partial(\chi_{j_k,H}/\chi_{j_k,F})/\partial\gamma_{j_k} > 0$  will determine the patterns of specialization of  $H$  and  $F$ .

**Proposition 4** *Let  $\pi_C(\gamma_{j_k})$  denote the probability that country  $C = H, F$  exports in equilibrium the final good  $j_k$ . Then:  $\partial\pi_H/\partial\gamma_{j_k} < 0$ ,  $\partial\pi_F/\partial\gamma_{j_k} > 0$ , and  $\pi_H(\gamma_{j_k} = 0) > \pi_F(\gamma_{j_k} = 0)$ .*

The result in Proposition 4 is analogous to the one presented in Proposition 2 for the benchmark model with two intermediate sectors. According to Proposition 4,  $H$ 's exports will be overrepresented by final goods with relatively broad input bases (i.e., final sectors with values of  $\gamma_{j_k}$  close to zero), whereas  $F$ 's exports will be overrepresented by final goods with relatively concentrated input bases (i.e., final sectors whose  $\gamma_{j_k}$  is large).

## 5 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 trade flows to seek for evidence consistent with the main predictions of the model.

### 5.1 Main Variables of Interest

#### Input Narrowness

The first task is coming up with an empirical measure of the breadth of the set of intermediate inputs used by each industry. Although the model is quite stylized to allow a direct match between its technological environment and real-world data on inputs and outputs by sectors, we can still use it as a guide to construct measures of narrowness of the industries' input bases.

In the benchmark model with two intermediate sectors, industry  $j$  allocates a fraction  $1 - \alpha_j$  of their total expenditure in intermediate inputs on varieties from sector 0, and the remainder  $\alpha_j$  on varieties from sector 1. Industries with either very low or very high values of  $\alpha_j$  will therefore end up sourcing most of their inputs from only one intermediate sector, and thus exhibit a *narrow* intermediate input base. Conversely, industries with values of  $\alpha_j$  around one half will rely quite heavily on both intermediate sectors, displaying accordingly a *wide* intermediate input base. Similarly, in the extension with  $N$  intermediate sectors presented in Section 4.2, industry  $j_k$  allocates a share  $N^{-1} + \gamma_{j_k}$  of their total spending in intermediate inputs on varieties from the intermediate sector  $k$ , while each of the other  $N - 1$  intermediate

sectors receives a share  $N^{-1} - (N - 1)^{-1} \gamma_{j_k}$  of  $j_k$ 's total spending in intermediates. As a result, industries with relatively small values of  $\gamma_{j_k}$  –i.e., values of  $\gamma_{j_k}$  close to zero– will display *wide* intermediate input bases, whereas those with relatively large values of  $\gamma_{j_k}$  –i.e., values of  $\gamma_{j_k}$  close to  $(N - 1)/N$ – will exhibit *narrow* input bases.

When looking for an empirical counterpart of the notion of *narrowness* of input bases, we need an index that aims at capturing the degree of concentration of intermediate input expenditure shares by industries. In the remainder of this paper, we will rely on the Gini coefficient computed based on industries' expenditure shares across different intermediate inputs.<sup>27</sup> Notice that the link between such a Gini coefficient and the parameter  $\alpha_j$  (and its counterpart  $1 - \alpha_j$ ) in the production functions (3) of our benchmark model is quite direct. In particular, if we computed the Gini coefficient based on expenditure shares on inputs from sector 0 and 1 for final sector  $j$  (with parameter  $\alpha_j$ ), we would obtain that:

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

In other words, the Gini coefficient of  $j$  would be decreasing in the breadth of the input base of industry  $j$  (that is, as  $\alpha_j$  approaches  $\frac{1}{2}$ ). Analogously, if we computed the Gini coefficient based on expenditure shares in the extended model with  $N$  intermediate sectors in Section 4.2, we would obtain that:

$$Gini_{j_k} = \gamma_{j_k},$$

where recall that  $\gamma_{j_k} \in [0, (N - 1)/N]$ , and that the larger the value of  $\gamma_{j_k}$  the narrower the input base of sector  $j_k$  turns out to be.<sup>28</sup>

To construct a measure of input narrowness by industries 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). 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. Thus, we index by  $k = 1, 2, \dots, K$  each of the sectors present in the IO matrix and also in the trade data, and by  $n = 1, 2, \dots, N$  each of the sectors selling intermediate inputs.

---

<sup>27</sup>Imbs and Wacziarg (2003) have previously used the Gini coefficient to measure the degree of concentration of labor and value added across different economic sectors. We apply the same methodology as them, but in this case we use it to measure the degree of narrowness or concentration of the intermediate input bases of different industries.

<sup>28</sup>A formal proof that  $Gini_{j_k} = \gamma_{j_k}$  is provided at the end of Appendix A.

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 coefficient of sector  $k$  as follows:

$$Gini_k = \frac{2 \times \sum_{n=1}^N n \times S_{k,n}}{N \times \sum_{n=1}^N S_{k,n}} - \frac{N+1}{N}, \quad (19)$$

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

In Section 5.3, we will use  $Gini_k$  as given by (19) 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 most of their intermediate inputs from relatively few sectors. Conversely, small values of  $Gini_k$  are verified when the distribution of  $S_{k,n}$  is quite evenly spread across a large number of intermediates.

## Export Specialization

To measure the degree of export specialization 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 categorized following the Harmonized System (HS) 6-digit classification, with 5,192 products. We map the trade flows data based on the HS 6-digit classification to the BEA industry codes using the concordance table between the 2002 IO matrix commodity codes and the HS 10-digit classification from the BEA website (after grouping the HS 10-digit codes into HS 6-digit products).<sup>29</sup> In the cases in which an HS-6 product maps into more than one BEA code, we assign trade flows proportionally to each of the BEA sectors which it maps into.<sup>30</sup> Lastly, the IO industry codes of the 2002 classification are matched to those of the 2007 classification, which are the ones actually used in the computation of the Gini coefficients.<sup>31</sup>

<sup>29</sup>When mapping the HS 6-digit products into the BEA 2002 codes, several of the original industries in the Input-Output matrix are lost due to lack of export data on them (these are mainly non-tradeable services). In the end, after merging the COMTRADE export data with the IO matrix information, we are left with data on trade flows and input narrowness for 294 industries as coded by the BEA 2002 classification.

<sup>30</sup>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. (We excluded the 11 products that map into five or more IO codes.) None of the regression results in Section 5.3 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.

<sup>31</sup>Unfortunately, we are not aware of any correspondence table between BEA 2007 codes and the HS codes (only BEA 2002 codes are matched to the 10-digit HS commodities via the correspondence table in <https://www.bea.gov/industry/benchmark-input-output-data>). As a consequence of this, we link the industry input narrowness measures and total exports at the industry level via the BEA 2002 codes.

## Road Network

We take the road network length by countries from the data on roadways from the CIA World Factbook. Roadways are defined as ‘total length of the road network, including paved and unpaved portions’. The year of the data point for each country varies, ranging from year 2000 to 2016, with the median year of the sample being 2010. (See details in Table A.10 in Online Appendix C.) 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:  $r_c \equiv \text{roadways}_c / \text{area}_c$ . In some of the robustness checks, we use also two additional measures of transport density: waterways density (defined as  $\text{waterways}_c / \text{area}_c$ ) and railway density (defined as  $\text{railways}_c / \text{area}_c$ ). The data on length of waterways and railways are also taken from the CIA World Factbook.

## 5.2 Road Density and Patterns of Specialization: Testing the predictions of the model

The two-country model presented in Section 4 predicts that the country with the denser road network (i.e., country  $H$ ) will tend to export goods with intermediate values of  $\alpha_j$ , while the country with the sparser road network (i.e., country  $F$ ) will tend to specialize in goods with either high or low values of  $\alpha_j$ . Conceptually, this prediction can be interpreted as stating that countries with denser road networks will tend to exhibit a comparative advantage in the types of industries that rely on a broader (or more diverse) set of intermediate inputs.

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

$$\ln(\text{Exp}_{c,k}) = \beta \cdot (r_c \times Gini_k) + \chi \cdot \Delta_{k,c} + \varsigma_c + \kappa_k + \nu_{c,k}. \quad (20)$$

In the regression equation (20) the dependent variable is given by the logarithm of the total value of exports in industry  $k$  by country  $c$  to all other countries in the world in year 2014. The term  $(r_c \times Gini_k)$  interacts the measure of input narrowness defined in (19) with the measure of road density (i.e., length of roadways per square kilometer).  $\Delta_{k,c}$  denotes a vector of additional covariates that may possibly influence specialization across countries in industries differing in terms of the degree of input narrowness.  $\varsigma_c$  and  $\kappa_k$  denote country fixed effects and industry fixed effects, respectively, and  $\nu_{c,k}$  represents an error term.<sup>32</sup>

---

<sup>32</sup>The interaction-term approach followed by regression (20) is analogous to that in used by Rajan and Zingales

The main coefficient of interest in (20) is  $\beta$ . If countries with a denser road network (i.e., countries with a greater  $r_c$ ) tend to actually exhibit a comparative advantage in industries that require a wider set intermediate inputs (i.e., industries with a smaller  $Gini_k$ ), then the data should deliver a negative estimate of  $\beta$ .

### 5.3 Empirical Results

Table I displays the first set of estimation results corresponding to (20). Column (1) includes only our main variable of interest (i.e.,  $r_c \times Gini_k$ ), 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 broader set of intermediates (i.e., those exhibiting a lower  $Gini_k$ ).

In the remaining columns of Table I, we sequentially incorporate additional interaction terms that may also influence the patterns of specialization across industries with different levels of input narrowness. Column (2) adds an interaction term between  $Gini_k$  and an index of *Rule of Law*, taken from World Governance Indicators. The rationale behind including this term lies in the argument in Levchenko (2007) and Nunn (2007), who show that countries with better contract enforcement institutions display a comparative advantage in industries that are heavily dependent on relationship-specific investments. Within our context, industries that need to source a broader set of intermediates may benefit relatively more from a sound legal environment, as they need to establish relationships with a greater number of input providers. Given that countries with better institutions tend to be also richer and invest more in basic infrastructure, omitting this term could lead to an overestimation (in absolute value) of the correlation coefficient of interest in (20). The regression in column (2) yields indeed a negative and significant coefficient associated with the interaction term between rule of law and  $Gini_k$ , consistent with the previous literature on institutions and specialization. In addition, the magnitude of  $\hat{\beta}$  falls relative to column (1), but it remains negative and significant.

Another possible source of omitted variable bias is linked to financial markets. 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; Saint-Paul, 1992; Acemoglu and Zilibotti, 1997). We could then expect that countries

---

(1998) to assess whether countries with deeper financial sectors grow proportionally more in industries that rely more strongly on external finance; a similar approach is also used by Manova (2013) but with focus placed on bilateral exports. Similar regression equations with interaction terms between country-level and industry-level variables can also be found in Romalis (2004), Nunn (2007), Levchenko (2007), Chor (2010), among others.

**TABLE I**  
Export Specialization across Industries with Different Levels of Input Narrowness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road Density x $Gini_k$	-4.084*** (0.305)	-2.240*** (0.334)	-2.382*** (0.363)	-2.356*** (0.364)	-1.852*** (0.379)	-1.845*** (0.385)	-2.439*** (0.453)
Rule of Law x $Gini_k$		-5.414*** (0.441)	-2.639*** (0.648)	-2.435*** (0.702)	-3.132*** (0.762)	-3.045*** (0.766)	-2.999*** (0.765)
Financial Development x $Gini_k$			-3.743*** (0.760)	-3.547*** (0.822)	-1.704* (0.913)	-1.681* (0.915)	-1.500* (0.918)
log GDP per capita x $Gini_k$				-0.417 (0.602)	0.905 (0.656)	1.306** (0.654)	0.822 (0.658)
Capital Intensity x log ( $K/L$ )					0.010** (0.004)		0.010** (0.004)
Skill Intensity x Human Capital					0.008*** (0.001)		0.008*** (0.001)
(Pop) Density x $Gini_k$							-0.197** (0.101)
Road Dens x (Pop) Dens x $Gini_k$							0.064** (0.027)
Observations	42,578	41,947	40,692	40,692	31,892	31,892	31,892
R-squared	0.765	0.764	0.764	0.764	0.794	0.793	0.794
Number of Countries	166	163	157	157	134	134	134
Number of Industries	294	294	294	294	259	259	259

Robust standard errors reported in parentheses. All regressions include country and industry fixed effects. The dependent variable is the logarithm of exports in industry  $k$  by country  $c$  in 2014. Rule of law is taken from the World Governance Indicators (World Bank) for year 2014. Financial development is measured by (log) private credit over GDP, taken from World Bank Indicators averaged for years 2005-2014. GDP per capita, stock of physical capital, and the human capital index are taken from Penn T tables, all in year 2014. Physical capital intensity and skill intensity by industry are taken from the NBER-CES Manufacturing Industry Database, and correspond to year 2011. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

with more developed financial markets would also be better able to specialize in industries that require a wider input base. To deal with this concern, in column (3) we interact the Gini coefficients with an indicator of financial development: the log ratio of private credit to GDP. (This indicator is taken from the World Bank Indicators database, and averaged during years 2005-2014.) 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. Yet, the estimate of  $\beta$  still remains negative and significant.

Column (4) adds an interaction term between  $Gini_k$  and log GDP per capita. This term would control for the possibility that richer economies may be better able to produce goods with lower  $Gini_k$ , given that richer economies tend to exhibit a more diversified productive structure than poorer ones. As we can see, the estimate of  $\beta$  remains essentially unaltered.

In column (5), following the approach by Romalis (2004), we introduce two additional regressors to control for specialization driven by factor endowments: *i*) an interaction term between capital intensity of industry  $k$  and the stock of physical capital per worker in country  $c$ ; *ii*) an interaction term between the skill intensity of industry  $k$  and the stock of human capital in country  $c$ . The measures of capital and skill intensity at the industry level are



constructed from the NBER-CES Manufacturing Industry database, for year 2011.<sup>33</sup> Some industries are lost from the sample in column (4) since the NBER dataset contains information only for manufacturing industries. For comparability, in column (5) we display the results of the regression in column (4), but using the restricted sample. The coefficients associated with the factor intensities carry the expected sign, while the estimates of  $\beta$  remain negative and significant. Furthermore, the estimated coefficients are of similar magnitude in both columns.

Finally, the last column of Table I addresses the possibility of a differential effect of the road network on the pattern of specialization depending on the population density of the economy. One could 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. Column (7) assesses this possibility by introducing an interaction term between population density and  $Gini_k$  (i.e.,  $Gini_k \times Pop\_Density_c$ ), and a *triple* interaction term that also includes  $r_c$  (i.e.,  $r_c \times Gini_k \times Pop\_Density_c$ ). If road network length is especially important for specialization in economies that are *less* densely populated, then the triple interaction term should carry a positive estimate. As can be observed, this is indeed the case. Moreover, the estimate of  $\beta$  after introducing the triple interaction term is still negative and highly significant.

The results displayed in Table I not only deliver a consistently significant estimate of the coefficient  $\beta$  in regression (20), but also one that is economically sizeable. For example, focusing on column (5) as the most complete specification, the estimate of the interaction term  $r_c \times Gini_k$  (equal to  $-1.852$  in this case) implies that a one-standard-deviation expansion of  $r_c$  will be associated with an increase of total exports in the sector  $k$  at the 25th percentile of the  $Gini_k$  distribution that is 7.8% larger than that of the sector at the 75th percentile of the distribution. This effect is of a similar order of magnitude as those implied by column (5) for a one-standard-deviation increase in *Rule of Law* and in *Financial Development*: for the former, the differential effect on exports in the sectors at the 25th and 75th percentile of the  $Gini_k$  distribution is equal to 10.7%; for the latter, the differential effect is equal to 5.2%. The comparative statics of a one-standard-deviation increase in human capital stock, when compared at the 75th and 25th percentile of human capital intensity distribution, is also of a similar order of magnitude, yielding a differential effect of 9.9%. The analogous exercise carried out for a one-standard-

---

<sup>33</sup>Capital intensity is computed as the total stock of physical capital per worker by industry. Skill intensity is measured by the average wage by industry. (See Becker, Gray and Marvakov (2013) for details on the NBER-CES Manufacturing Industry database.) Both the measure of physical capital per worker and the index of human capital are drawn from the Penn Tables database. (The human capital index is based on the average years of schooling from Barro & Lee (2013) and an assumed rate of return of education based on Mincer estimates.)

**TABLE II**  
High-Income and Low-Income Subsamples

	(1)	(2)	(3)	(4)	(5)	(6)
Road Density x $Gini_k$	-2.790*** (0.309)	-6.680*** (1.361)	-2.156*** (0.355)	-4.680*** (1.475)	-1.653*** (0.373)	-4.418*** (1.710)
Rule of Law x $Gini_k$			-1.496 (1.019)	-3.150* (1.709)	-2.198** (1.109)	-2.161 (1.974)
Financial Development x $Gini_k$			-3.706*** (1.139)	-3.360*** (1.250)	-3.446*** (1.257)	-0.718 (1.351)
log GDP per capita x $Gini_k$			-0.733 (1.607)	1.403 (1.038)	1.859 (1.721)	1.907* (1.075)
Capital Intensity x $\log(K/L)$					0.006 (0.012)	0.011 (0.009)
Skill Intensity x Human Capital					0.013*** (0.002)	-0.014*** (0.002)
Observations	22,633	19,945	21,220	19,472	16,370	15,522
R-squared	0.803	0.665	0.790	0.669	0.807	0.692
Countries Sample (high/low income)	High	Low	High	Low	High	Low
Number of Countries	83	83	76	81	65	69
Number of Industries	294	294	259	259	259	259

Robust standard errors reported in parentheses. All regressions include country and industry fixed effects. The dependent variable is  $\log(\text{Expo}_{c,k})$  in the year 2014. The high-income sample comprises countries whose GDP per capita is above the sample median, and the low-income sample countries with GDP per capita below it. The median sample income lies between that of Dominican Rep. (\$12,511 PPP) and China (\$12,472 PPP) in 2014. Rule of Law is taken from World Governance Indicators for year 2014. Financial development is measured by (log) private credit over GDP, averaged for years 2005-2014. Physical capital and skill intensity by industry are taken from the NBER-CES Manufacturing Industry Database, and correspond to year 2011. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

deviation in the physical capital per head is associated with a differential effect of 2.1%.

Next, Table II displays some of the regressions previously presented in Table I, but now splitting the sample of countries in two subsamples, 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’. The results show that the effect of road density on the patterns of specialization 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.<sup>34</sup>

### Additional Robustness Checks

Further robustness checks are provided in the online appendix (Appendix B). Tables A.2 and A.3 change the measure of transport density. In Table A.2, we show the results of a set of regressions substituting  $r_c$  in (20) by railway density, computed as the total railway network length of country  $c$  per square kilometer. In Table A.3, we expand our measure transport

<sup>34</sup>This difference in magnitude could suggest the presence of some sort of decreasing marginal effect of road density, since richer economies tend to exhibit denser road networks than poorer ones (see Figure 1 later on).

network length to include (in addition to roadways) also the total length of internal railways and waterways. All the results in Table A.2 and A.3 follow a similar pattern as those shown in Table I.

Table A.4 shows that all the previous results are also robust to: *i*) excluding very small countries (both in terms of area and population), *ii*) excluding very large countries (in terms of area), *iii*) controlling for the effect of area and population (in both cases interacted with the measure of input narrowness), *iv*) including the interaction between total GDP and input narrowness, and *v*) excluding from the sample those countries whose road networks were measured before year 2010 (which is the median year in the sample).

The rationale for those additional robustness checks is the following. In the case of very small countries, on the one hand, they may find it easier to link together geographic locations while, on the other hand, they may be less able to provide sufficient opportunities for input diversity. Next, regarding very large countries in terms of area, the concern could be that some of those countries may contain very large swathes of uninhabitable land, which may end up turning our measure of road density somewhat imprecise in those cases. Controlling for the size of the country (both in terms of area and population) takes into account the possibility that larger countries may face more opportunities for input diversity, regardless of the density of their internal transport network. Similarly, including total GDP can control for the possibility that there exist minimum size requirements to open up some sectors in the economy. Finally, restricting the sample to countries whose road networks were measured after 2010 helps in harmonizing the data year on trade flows and road density, and shows that the found effects are not contaminated by countries whose data on road networks is relatively older.<sup>35</sup>

Lastly, Table A.5 shows the results of some of the regressions of Table I, but computing the variable ‘road density’ using only the total length of *paved* roads. All the main results hold through. Moreover, the estimate for the main interaction term is in absolute terms greater when using only paved roads than in the regression that relies on all types of roads. This is, in fact, the type of result one should expect to see if paved roads provide better transportation than unpaved roads do.

---

<sup>35</sup>For this same harmonization purpose, column (7) of Table A.4 shows also the results when using only data on road density for years 2010 and 2011. The estimates remain essentially unaffected as well in this case.

## 6 Endogeneity and Alternative Interpretations

The previous section presented a robust association between road density in country  $c$  and its degree of specialization in industries that rely on a wide set of inputs. While those results are certainly consistent with the main predictions of the model, they cannot be taken as hard evidence of its underlying mechanism. Two separate issues deserve further discussion and analysis. First, the correlation found in the previous regressions could as well be the result of road infrastructure responding to transport needs stemming from industry specialization (i.e., 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 a different feature: the degree of product complexity of industry  $k$ . In the next two subsections we aim to address these two points more explicitly.

### 6.1 Endogeneity and Reverse Causation

Our model 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. The exogeneity of  $r_c$  represents thus a critical assumption that warrants further discussion in case the previous empirical results are intended to be interpreted as evidence of a *causal* effect from road density to specialization.

Endogenizing countries' road networks can easily lead to a model where  $\beta$  in (20) can be confounding an effect from road density to specialization, together with reverse causality from the latter to the former. For example, suppose that for some reason the final good production functions differ between  $H$  and  $F$ , and that  $H$  turns out to be relatively more productive in the final sectors that rely on a broader set of intermediate inputs. If countries were able to 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  $\beta$  could end up capturing (at least partially) an effect going from patterns of specialization to road density.

The rest of this subsection provides some further support for the notion that the density of the internal transport network is instrumental to specialization in industries with wider input bases. First, we show that a similar correlation to that one found in Section 5.3 arises when the density of the transport network is measured by the density of internal waterways. Next, we show that the results in Section 5.3 remain true when we instrument the density of a country's road network with topographical measures of terrain roughness.

### 6.1.1 Patterns of Specialization and Waterways Density

This first part intends to provide some further evidence consistent with the main mechanism of the model, by relying on a measure of countries' transport network that is *less* sensitive to reverse causality concerns than  $r_c$  is. 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.<sup>36</sup> Arguably, while countries can still expand their waterways 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 roads.

One additional aspect we investigate here is the possibility that waterways impact specialization *heterogeneously* at different stages of development. For various 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 of waterways may have influenced patterns of development before railroads and roads became more widespread worldwide, waterways are no longer a mode of transportation that seems to be associated with economies' current level of development. In fact, a quick look at simple cross-country correlations in Figure 1 shows that income per head and road density display a clear positive correlation, while the association between income per head and waterways density is rather weak.<sup>37</sup>

Table III displays the results of a regression equation analogous to (20), but where  $r_c$  is replaced by a measure of *waterways density*. The table shows the results for three different countries sample: entire sample, high-income countries, and low-income countries.

The regressions based on the whole set of countries tend to yield an estimate that is negative. However, this aggregate result masks important heterogeneities in the impact of waterways density on specialization for richer versus poorer economies. Column (2) shows that waterways density carries no impact at all in the subsample of above-median income economies. By contrast, column (3) exhibits a negative and highly significant coefficient. This suggests that, in the case of poorer economies, those that enjoy a denser network of waterways tend to export relatively more in industries that require a wider intermediate input base.

---

<sup>36</sup>The CIA World Factbook measures waterways as the total length of navigable rivers, canals and other inland bodies of water.

<sup>37</sup>One possible interpretation of Figure 1 is that, as economies grow richer, roadways tend to gradually overshadow waterways for internal transportation. From this perspective, we could then expect waterways to represent an important determinant of specialization in poorer economies, but losing preeminence in richer economies where roadways can more easily make up for an insufficiently dense internal waterway network.

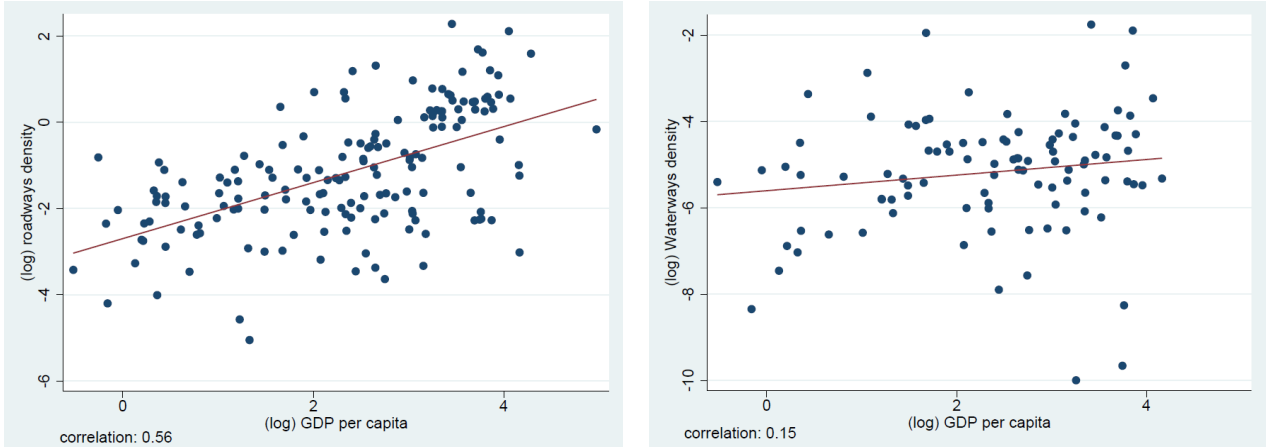


Figure 1: Roadways and waterways density against GDP per head

Columns (4)-(6) re-run the regressions in columns (1)-(3) but also including the original interaction term  $r_c \times Gini_k$ . The results for the impact of waterways density on specialization in (5) and (6) follow a very similar pattern as those in (2) and (3). Moreover, the regressions also show that the coefficient for road density remains negative and significant.<sup>38</sup> Finally, Table A.7 in Appendix B shows the results of a set of regressions that include a triple interaction term between  $r_c$ ,  $Gini_k$  and log income, rather than splitting the sample of countries according to income. All the results remain qualitatively consistent with those in Table III.

### 6.1.2 Instrumental Variables: Terrain Roughness and Roadway Density

This second part intends to address more directly the concern of reverse causation from industry specialization to road density. To do so we instrument  $r_c$  with measures of terrain roughness in country  $c$ . The idea is drawn from Ramcharan (2009), who shows that countries with rougher terrain surface tend to exhibit less dense road networks.<sup>39</sup> In the context of our paper, if the roughness of the terrain affects the density of the internal road network, but it does not exert a systematic impact on specialization across industries with varying degrees of input narrowness

<sup>38</sup>Note that the results in columns (5) and (6) of Table III are not directly comparable to those in columns (3) and (4) of Table II due to the loss of some countries in the samples of Table III.

<sup>39</sup>Ramcharan (2009) studies the spatial concentration of economic activities within countries, and how this is affected by their surface topography. The author argues that countries with rougher topography tend to display stronger spatial concentration of economic activity, and that this is partly explained by the poorer land transportation associated with rougher terrain.

**TABLE III**  
Waterways Density as Measure of Transport Network

	(1)	(2)	(3)	(4)	(5)	(6)
Waterways Density x $Gini_k$	-0.262*	0.016	-1.120***	-0.137	0.194	-1.100***
	(0.148)	(0.140)	(0.347)	(0.170)	(0.172)	(0.347)
Rule of Law x $Gini_k$	-4.205***	-4.727***	-7.389***	-3.576***	-4.257***	-4.533*
	(0.838)	(1.546)	(2.511)	(0.902)	(1.561)	(2.811)
Financial Development x $Gini_k$	-1.724*	-2.403*	0.484	-1.857*	-2.505*	0.624
	(1.077)	(1.377)	(1.664)	(1.085)	(1.382)	(1.663)
log GDP per capita x $Gini_k$	0.457	5.220*	0.283	0.620	5.752**	0.830
	(0.867)	(2.944)	(1.256)	(0.874)	(2.962)	(1.268)
Capital Intensity x log ( $K/L$ )	0.007	-0.014	0.021*	0.008	-0.013	0.021*
	(0.006)	(0.014)	(0.012)	(0.006)	(0.014)	(0.012)
Skill Intensity x Human Capital	0.007***	0.020***	-0.018***	0.007***	0.020***	-0.018***
	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)
Road Density x $Gini_k$				-1.381**	-1.426**	-7.249***
				(0.711)	(0.703)	(2.386)
Observations	22,357	11,750	10,607	22,357	11,750	10,607
R-squared	0.811	0.798	0.712	0.811	0.798	0.712
Countries Sample	All	High	Low	All	High	Low
Number of Countries	93	46	47	93	46	47
Number of Industries	259	259	259	259	259	259

Robust standard errors in parentheses. All regressions include country and industry fixed effects. The dependent variable is  $\log(\text{Exp}_{c,t})$  in year 2014. Waterways is taken from the CIA World Factbook. It comprises total length of navigable rivers, canals and other inland water bodies. Waterway density equals internal waterways per sq km. Columns (2) and (5) include only countries whose income per capita is above the sample median. Columns (3) and (6) include only countries with income per capita below the sample median. \*\*\*p<0.01, \*\*p<0.05, \*p<0.1

via other alternative channels, then it can serve as a valid instrument for  $r_c$ .<sup>40</sup>

In Table IV we show the results of the two-stage least square (2SLS) regressions using three alternative measures of terrain roughness: *i*) the difference between the maximum and minimum land elevation in country  $c$ , taken from the CIA Factbook; *ii*) the standard deviation of elevation of country  $c$  measured at 30" degree grids (approx. 1km cells), taken from Ramcharan (2009); *iii*) the percentage of mountainous terrain, taken from Fearon and Laitin (2003). Table A.8 in

<sup>40</sup>Notice that a violation of the exclusion restriction in this context requires more than simply terrain roughness having an impact on industry specialization. For the exclusion restriction to be violated, it must be the case that terrain roughness affects specialization across industries in a way that is also correlated with their degree of input narrowness (besides the effect mediated by the impact of terrain roughness on road density). For example, the exclusion restriction may be threatened if economies with rougher terrain tend to also display more heterogeneous climatic conditions and land configurations, and this allows them to enjoy a more diverse productive structure. Conversely, it may be that rougher terrain reduces the share of inhabitable land, curtailing productive heterogeneity, and thereby possibly leading to a violation of the exclusion restriction via a negative impact of roughness on specialization in industries with wide input breadth. While we cannot test the validity of the exclusion restriction, the results in Table A.9 in Appendix B (see also the discussion therein) are in principle encouraging about how concerned we should remain about the possibility of a direct impact of terrain roughness, besides that one mediated by its effect on road density.

Appendix B shows that all three measures of terrain roughness are negatively correlated with road density, even after controlling for several other country-level variables.<sup>41</sup>

Columns (1) and (2) display the results of the 2SLS regressions based on the difference between the maximum and minimum land elevation in  $c$  as an instrument for  $r_c$ . Arguably, this variable may be seen as a relatively crude measure for terrain roughness. However, it has the upside of being available for the exact same samples as those in Section 5.3. As a result, columns (1) and (2) of Panel A can be directly compared to their respective OLS counterparts in columns (4) and (5) in Table I. Next, columns (4) and (6) display the estimation results of the 2SLS regressions in which the instrument for road density is based on the standard deviation of land elevation. Given that for some of the countries in the original sample this information is missing, columns (3) and (5) show their respective OLS estimates for the corresponding sample. Finally, columns (8) and (10) report the 2SLS results when using the percentage of mountainous terrain as instrumental variable. Again, for comparability, columns (7) and (9) report the OLS estimates for the corresponding country samples.

The main message to draw from Table IV is that the 2SLS regressions consistently yield a negative and significant estimate for our coefficient of interest. These results reinforce the support for the hypothesis that the density of the internal transport network is an important determinant of specialization in industries with wide input bases, by exploiting the variation in the internal road network across countries predicted by their degrees of terrain roughness.

One additional point to note is that the 2SLS estimates for  $\beta$  tend to be consistently greater in absolute magnitude than their OLS counterparts. This would in principle run against the direction of the bias that would stem from the reverse causality concern discussed in the second paragraph of Section 6.1 (i.e., the notion that economies specializing in industries that rely on a wide variety of inputs may tend to invest more in transport infrastructure). One possible reason behind these results is that the instrument may also be alleviating some degree of measurement error in our indicator of road density. In that respect, recall that  $r_c$  is computed using the total length of roads by country. This disregards the fact that different roads may differ substantially in terms of quality and width, and it is also summing up together paved and unpaved roads. Furthermore, the total length or the road network is not taking into account the possibility of a

---

<sup>41</sup>All the measures of terrain roughness used here aim at capturing large-scale terrain irregularities. In a sense, these seem to be the types of irregularities that can most severely hinder internal transport networks. Other papers, e.g. Nunn and Puga (2012), have resorted to the methodology developed by Riley et al. (1999) so as to measure small-scale terrain irregularities. While small-scale terrain roughness measures seem more appropriate for capturing the presence of small geographic formations that may provide natural sources of protection to certain groups of people, they may not represent the main source of obstruction to dense transport networks.



**TABLE IV**  
Two-Stage Least Squares Regressions using Terrain Roughness as Instrument for Roadway Density

	PANEL A: SECOND-STAGE RESULTS (AND OLS COMPARISONS)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	2SLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Road Density x $Gini_k$	-5.006*** (1.513)	-3.188* (1.670)	-3.364*** (0.509)	-6.199** (2.800)	-2.204*** (0.544)	-6.166* (3.573)	-3.347*** (0.506)	-5.638* (3.258)	-2.000*** (0.515)	-5.645* (3.355)
Rule of Law x $Gini_k$	-1.260 (0.943)	-2.469** (1.100)	-2.285*** (0.720)	-1.306 (1.187)	-2.886*** (0.786)	-1.264 (1.635)	-2.082*** (0.737)	-1.108 (1.564)	-3.147*** (0.790)	-1.543 (1.661)
Fin Dev x $Gini_k$	-3.340*** (0.838)	-1.678* (0.917)	-4.063*** (0.878)	-3.978*** (0.890)	-2.605*** (0.956)	-2.849*** (0.963)	-3.975*** (0.870)	-4.060*** (0.874)	-1.507 (0.938)	-1.719* (0.948)
log GDP pc x $Gini_k$	-0.088 (0.623)	1.077* (0.677)	0.161 (0.667)	0.620 (0.781)	1.170* (0.714)	1.857** (0.908)	0.056 (0.641)	0.437 (0.820)	0.926 (0.676)	1.562* (0.868)
K Intens x log ( $K/L$ )		0.010** (0.004)			0.009** (0.0045)	0.009** (0.0045)			0.008* (0.004)	0.008* (0.004)
Skill Intens x Human K		0.008*** (0.001)			0.007*** (0.001)	0.007*** (0.001)			0.008*** (0.001)	0.008*** (0.001)
Observations	40,692	31,892	35,988	35,988	29,229	29,229	36,544	36,544	30,067	30,067
R-squared	0.763	0.794	0.764	0.764	0.794	0.793	0.757	0.757	0.795	0.795
Number of Countries	157	134	135	135	121	121	138	138	126	126
Number of Industries	294	259	294	294	259	259	294	294	259	259

PANEL B: FIRST-STAGE RESULTS (Dependent Variable: Road Density x Gini)										
Terrain Roughness x $Gini_k$	-0.013*** (0.0004)	-0.013*** (0.001)		-0.452*** (0.017)		-0.392*** (0.019)		-0.007*** (0.0002)		-0.007*** (0.0003)
Rule of Law x $Gini_k$	0.344*** (0.011)	0.371*** (0.012)		0.290*** (0.011)		0.350*** (0.012)		0.398*** (0.011)		0.413*** (0.012)
Fin Dev x $Gini_k$	0.109*** (0.011)	0.059*** (0.013)		0.066*** (0.011)		-0.029*** (0.011)		-0.004 (0.009)		-0.032*** (0.010)
log GDP pc x $Gini_k$	0.144*** (0.007)	0.164*** (0.010)		0.165*** (0.009)		0.184*** (0.010)		0.152*** (0.008)		0.163*** (0.009)
K Intens x log ( $K/L$ )		0.005 (0.008)				0.010 (0.007)				0.008 (0.007)
Skill Intens x Human K		-0.010 (0.015)				-0.017 (0.013)				-0.010 (0.014)

Robust standard errors in parentheses. All regressions include country and industry fixed effects. The dependent variable in Panel A is the  $\log(\text{Exp}_{c,t})$  in year 2014. 'Terrain Roughness' is measured as follows: *i*) in columns (1) and (2) by the difference between the maximum and the minimum elevation in country  $c$  (source: CIA World Factbook); *ii*) in columns (4) and (6) by the std. deviation of elevation in country  $c$  computed at the 30" degree resolution [source: Ramcharan (2009)]; *iii*) in columns (8) and (10) by the percentage of mountainous in terrain in country  $c$  (source: Fearon and Laitin, 2003). \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

very inefficient layout of the network. All these issues could end up reducing the precision with which  $r_c$  captures the notion that a denser road network allows cheaper internal transportation of inputs. Therefore, when instrumenting  $r_c$ , we may not only be dealing with problems of endogeneity, but also with the fact that in some cases our measure of road density may be quite imprecisely gauging the efficiency of the internal road network.

## 6.2 Alternative Interpretations of the Input Breadth Measures

The analysis in Section 5 was based on the notion that the degree of input breadth of industry  $k$  can serve as an indicator for how reliant this industry is on the internal transport network. The

need to source a large variety of inputs can certainly make a particular sector heavily dependent on efficient transportation; however, it can also mean that the sector is highly sensitive to sound contracts enforcement. Indeed, Blanchard and Kremer (1997) and Levchenko (2007) have previously used input-output data to compute diversification indices for intermediate input purchases across industries, and use them to proxy the degree of complexity of sectors: sectors with more diversified (i.e., less concentrated) input bases are considered to be more complex. In their analysis, more complex sectors require better contract enforcement to work efficiently. From this viewpoint, countries with better functioning institutions should exhibit a comparative advantage in industries with broad intermediate input bases. Levchenko (2007) shows that this is indeed the case for US imports: the US imports a higher share of goods with greater diversity of intermediate inputs from countries with better rule of law.

The previous regressions have conditioned on the interaction between rule of law in country  $c$  and the Gini coefficient for intermediate inputs in industry  $k$ . The estimate of  $\beta$  remained consistently negative and significant, regardless of the introduction of this additional control. In that regard, our results seem to suggest that *both* institutions and local transport networks are instrumental and complementary to the growth of industries with wide input bases. This section will attempt to further strengthen this argument

Countries with better institutions are in general richer, and also exhibit a denser transportation infrastructure network. If  $Gini_k \times r_c$  in (20) were *not* capturing any type of impact related to how transport-intensive industry  $k$  is, but *only* the effect of rule of law in country  $c$  through its correlation with  $r_c$ , then the correlation found in Table I should arise more prominently for industries that are relatively more dependent on judicial quality. The regressions reported in Table V show this is actually not the case in the data.

Columns (1) and (2) in Table V show the results of the simple correlation reported initially in column (1) of Table I, after splitting the set of industries in two subsamples: low contract intensity vs. high contract intensity. To do so, we take the measure of contract intensity by industries from Nunn (2007), and split the sample of industries according to whether they rank below or above the median value of contract intensity.<sup>42</sup> If the  $Gini_k$  were simply proxying for how sensitive to efficient contract enforcement industry  $k$  is, then the estimate in column (1) should turn out to be significantly milder than that one in column (2). The regressions show, however, that the negative correlation is significant in both subsamples, and moreover the magnitude of the estimates is not significantly different from one another.

---

<sup>42</sup>Nunn (2007) reports contract intensity measures for 222 industries coded according to NAICS 1997. We lose some of the original industries in Table I when matching the NAICS 1997 codes to those of BEA 2002.

**TABLE V**  
Regressions on Industry Subsamples: Effects at Different Leves of Contract Intensity

	(1)	(2)	(3)
Road Density x $Gini_k$	-3.443*** (0.652)	-1.929*** (0.483)	-2.070*** (0.440)
Rule of Law x Contract Intensity			0.441*** (0.057)
Rule of Law x $Gini_k$			-1.974** (0.926)
Financial Development x $Gini_k$			-0.533 (1.082)
log GDP per capita x $Gini_k$			0.897 (0.772)
Capital Intensity x log ( $K/L$ )			0.007*** (0.002)
Skill Intensity x Human Capital			0.008* (0.005)
Observations	13,179	14,231	20,823
R-squared	0.704	0.820	0.790
Contract Intensity	low	high	all
Number of Countries	166	166	134
Number of Industries	91	91	163

Robust standard errors in parentheses. All regressions include country and industry fixed effects. The dependent variable is  $\log(\text{Expo}_{c,k})$  in year 2014. Contract intensity by industry measures are taken from Nunn (2007). The original measures are coded according to the NAICS 1997 classification and matched to the BEA codes. Nunn (2007) measures contract intensity in  $k$  as the proportion of inputs of  $k$  classified as differentiated by Rauch (1999). Rule of law is from the World Governance Indicators (year 2014). Financial development is (log) private credit over GDP, averaged for years 2005-14. GDP per capita, total GDP, the stock of physical capital (per capita), and the human capital index ( $H_c$ ) are all taken from the Penn Tables (year 2014). Capital and skill intensity by industry are from the NBER-CES Manuf. Industry Database (for year 2011). \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Finally, to complement our previous results, column (3) displays the outcomes of a regression analogous to that one in column (5) of Table I, but including the interaction term between rule of law in country  $c$  and contract intensity of industry  $k$  (namely, ‘Rule of Law  $\times$  Contract Intensity’). Consistently with the previous results in the literature, the regression shows that countries with better institutions exhibit a comparative advantage in the industries with higher contract intensity. Moreover, the regression results are in line with those in column (5) of Table I, and support the prediction that countries with denser road networks export relatively more in sectors that require a broader variety of intermediate inputs.

## 7 Concluding Remarks

We proposed a simple trade model where the density of the internal transport network represents a key factor in shaping comparative advantage and specialization. The underlying mechanism rests on the idea that shipping intermediate inputs across spatially diffuse locations is costly. As a consequence, industries that rely on a wider set of intermediate inputs become heavier users of the internal transport network. The model shows that countries with denser internal transport infrastructures exhibit a comparative advantage in industries that rely on a broader variety of intermediate goods.

Drawing on intermediate goods transactions from the US input-output matrix to measure industries' input breadth, we have also shown that the patterns of specialization predicted by the model are broadly consistent with the trade flows observed in the data. Several caveats apply nonetheless to the empirical evidence. First, patterns of specialization could be influencing investment in road infrastructure, and thus be behind the correlation found in the data. In that respect, the fact that the same correlation appears when substituting roadway density by waterway density seems reassuring, as waterways are harder to expand in response to increased transport needs. Furthermore, our empirical results also remain in line with the predictions of the model when we instrument the density of countries' road networks with indicators of roughness of the terrain. Second, the measure of input breadth by industries could alternatively be capturing a stronger need for contract enforcement when the input base is wider. We showed however that the correlation predicted by the model is still present when the confounding effect of institutional quality (by country) and judiciary intensity (by industry) is accounted for. Lastly, road density is a relatively imprecise way to capture the efficiency in connectedness of dispersed locations. Road networks not only differ in length, but also in width, quality, etc. In addition, road density fails to account for inefficiencies in the layout of internal roads. It would be certainly desirable to use a more detailed and refined measure of internal road networks. Yet, it is hard to envision a clear reason why such sources of measurement error could end up systematically biasing the empirical results in the direction predicted by the model.

As a final remark, the paper has abstracted from analyzing the possibility of a heterogeneous impact of transport networks on the internal geographic configuration of industries with different degrees of input breadth. Access to international markets could indeed create a force leading to an agglomeration of industries with wider input bases along "coastal" regions (i.e., regions closer to international markets). Furthermore, this force pushing certain industries towards coastal regions could be especially important in countries lacking a sufficiently dense internal transport network. We leave this question open for future research.

## 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] Agnosteva, D., Anderson, J. and Yotov, Y. (2019). ‘Intra-national Trade Costs: Assaying Regional Frictions’, *European Economic Review*, vol. 112, pp. 35-50.
- [4] Allen, T. and Arkolakis, C. (2014). ‘Trade and the Topography of the Spatial Economy’, *Quarterly Journal of Economics*, vol. 129, pp. 1085–1140.
- [5] Allen, T. and Arkolakis, C. (2019). ‘The Welfare Effects of Transportation Infrastructure Improvements’, NBER WP No. 25487.
- [6] Anderson, J. E., and Van Wincoop, E. (2004). ‘Trade costs’, *Journal of Economic Literature*, vol. 42, pp. 691–751.
- [7] Antràs, P. (2005). ‘Incomplete Contracts and the Product Cycle’, *American Economic Review*, vol. 95, pp. 1054-1073.
- [8] Atkin, D. and Donaldson, D. (2015). ‘Who’s Getting Globalized? The Size and Implications of Intra-national Trade Costs’, Mimeo, MIT.
- [9] 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.
- [10] Beck, T. (2002). ‘Financial Development and International Trade. Is There a Link?’, *Journal of International Economics*, vol. 57, pp. 107-131.
- [11] 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.
- [12] Becker, R., Gray, W. and Marvakov, J. (2013). ‘NBER-CES manufacturing industry database: Technical notes’, NBER.
- [13] Behrens, K. (2016). ‘Agglomeration and clusters: Tools and insights from coagglomeration patterns’, *Journal of Canadian Economics*, vol. 49, pp. 1293-1339.

- [14] Behrens, K. and Bougna, T. (2015). ‘An anatomy of the geographical concentration of Canadian manufacturing industries’, *Regional Science and Urban Economics*, vol. 51, pp 47-69.
- [15] Behrens, K. and Brown, M. (2018). ‘Transport costs, trade, and geographic concentration: Evidence from Canada’, in *Handbook of International Trade and Transportation*, B. Blonigen and W. Wilson (Eds), EE Publishing.
- [16] Behrens, K. and Sharunova, V. (2015). ‘Inter- and Intra-Firm Linkages: Evidence from Micro-geographic Location Patterns’, CEPR Discussion Paper No. 10921.
- [17] Bernard, A., Moxnes, A. and Saito, Y. (2019). ‘Production Networks, Geography and Firm Performance’, *Journal of Political Economy*, Vol. 127, pp. 639–688
- [18] Bertinelli, L. and Decrop, J. (2005). ‘Geographical agglomeration: Ellison and Glaeser’s index applied to the case of Belgian manufacturing industry’, *Regional Studies*, vol. 39, pp. 567-583
- [19] Blanchard, O. and Kremer, M. (1997). ‘Disorganization’, *Quarterly Journal of Economics*, Vol. 112, pp. 1091-1126.
- [20] Cainelli, G. and Iacobucci, D. (2012). ‘Agglomeration, Related Variety, and Vertical Integration’, *Economic Geography*, Vol. 88, pp. 255-277.
- [21] Caliendo, L., Parro, F., Rossi-Hansberg, E. and Sarte, P. D. (2018). ‘The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy’, *Review of Economic Studies*, vol. 85, pp 2042-2096.
- [22] Chor, D. (2010). ‘Unpacking sources of comparative advantage: A quantitative approach’, *Journal of International Economics*, vol. 82, pp. 152-167.
- [23] Clague, C. (1991a). ‘Relative Efficiency, Self-Containment, and Comparative Costs of Less Developed Countries’, *Economic Development and Cultural Change*, vol. 39, pp. 507-530.
- [24] Clague, C. (1991b). ‘Factor Proportions, Relative Efficiency and Developing Countries’ Trade’, *Journal of Development Economics*, Vol. 35, pp. 357-380.
- [25] Coşar, K. and Fajgelbaum, P. (2016). ‘Internal Geography, International Trade, and Regional Specialization’, *American Economic Journal: Microeconomics*, vol. 8, pp. 24-56.
- [26] Costinot, A. (2009). ‘On the Origins of Comparative Advantage’, *Journal of International Economics*, vol. 77, pp 255-264.

- [27] Cuñat, A. and Melitz, M. (2012). ‘Volatility, labor market flexibility, and the pattern of comparative advantage’, *Journal of the European Economic Association*, vol. 10, pp 225-254.
- [28] Datta, S. (2012). ‘The impact of improved highways on Indian firms’, *Journal of Development Economics*, Vol. 99, pp. 46-57.
- [29] Devereux, M., Griffith, R. and Simpson, H. (2004). ‘The geographic distribution of production activity in the UK’, *Regional Science and Urban Economics*, vol. 34, pp. 533-564.
- [30] Donaldson, D. (2018). ‘Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure’, *American Economic Review*, vol. 108, pp. 899-934.
- [31] Donaldson, D. and Hornbeck, R. (2016). ‘Railroads and American Economic Growth: A "Market Access" Approach’, *Quarterly Journal of Economics*, vol. 131, pp 799-858.
- [32] Duranton, G., Morrow, P. and Turner, M. (2014). ‘Roads and Trade: Evidence from the US’, *Review of Economic Studies*, vol. 81, pp. 681-724.
- [33] Duranton, G. and Overman, H. (2005). ‘Testing for Localization Using Micro-Geographic Data’, *Review of Economic Studies*, Vol. 72, pp. 1077–1106.
- [34] Duranton, G. and Overman, H. (2008). ‘Exploring the detailed location patterns of UK manufacturing industries using microgeographic data’, *Journal of Regional Science*, vol. 48, pp. 213–243.
- [35] Eaton, J. and Kortum, S. (2002). ‘Technology, geography, and trade’, *Econometrica*, vol. 70, pp. 1741-1779.
- [36] Ellison, G., Glaeser, E. and Kerr, W. (2010). ‘What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns’, *American Economic Review*, vol. 100, pp. 1195-1213.
- [37] Ellison, G. and Glaeser, E. (1997). ‘Geographic concentration in US manufacturing industries: a dartboard approach’, *Journal of Political Economy*, vol. 105, pp. 889-927.
- [38] Fajgelbaum, P. and Redding, S. (2014). ‘External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914’, NBER WP No. 20217.
- [39] Fajgelbaum, P. and Schaal, E. (2017). ‘Optimal Transport Networks in Spatial Equilibrium’, NBER WP No. 23200.
- [40] Fearon, J. and Laitin, D. (2003). ‘Ethnicity, Insurgency, and Civil War’, *American Political Science Review*, vol. 97, pp. 75–90.

- [41] Felbermayr, G. and Tarasov, A. (2015). ‘Trade and the Spatial Distribution of Transport Infrastructure’, CESifo Working Paper 5634.
- [42] Ferguson, S. and Formai, S. (2013) ‘Institution-Driven Comparative Advantage and Organizational Choice’, *Journal of International Economics*, vol. 90, pp. 193-200.
- [43] Gaulier, G. and Zignago, S. (2010). ‘BACI: International Trade Database at the Product-level. The 1994-2007 Version’, CEPII Working Paper 2010-23.
- [44] Gibbons, S., Lyytikäinen, T., Overman, H. and Sanchis-Guarner, R. (2019). ‘New road infrastructure: The effects on firms’, *Journal of Urban Economics*, vol. 100, pp. 35-50.
- [45] Greenwood, J. and Jovanovic, B. (1990). ‘Financial Development, Growth, and the Distribution of Income’, *Journal of Political Economy*, vol. 98, pp. 1076-1107.
- [46] Gunasekera, K. Anderson, W. and Lakshmanan T. (2008). ‘Highway-induced development: evidence from Sri Lanka’, *World Development*, vol. 36, pp. 2371-2389.
- [47] Hillberry, R. and Hummels, D. (2008). ‘Trade responses to geographic frictions: A decomposition using micro-data’, *European Economic Review*, vol. 52, pp. 527-550.
- [48] Imbs, J. and Wacziarg, R. (2003). ‘Stages of Diversification’, *American Economic Review*, vol. 93, pp. 63-86.
- [49] Koh. H. J. and Riedel, N. (2014). ‘Assessing the Localization Pattern of German Manufacturing and Service Industries: A Distance-based Approach’, *Regional Studies*, vol. 48, pp. 823-843.
- [50] Levchenko, A. (2007). ‘Institutional Quality and International Trade’, *Review of Economic Studies*, vol. 74, pp. 681-724.
- [51] Limao, N. and Venables, A. (2001). ‘Infrastructure, geographical disadvantage, transport costs, and trade’, *World Bank Economic Review*, vol. 15, pp. 451-479.
- [52] Manova, K. (2013). ‘Credit Constraints, Heterogeneous Firms, and International Trade’, *Review of Economic Studies*, vol. 80, pp. 711-744.
- [53] Maurel, F. and Sedillot, B. (1999). ‘A measure of the geographic concentration in french manufacturing industries’, *Regional Science and Urban Economics*, vol. 29, pp. 575-604.
- [54] Matsuyama, K. (2017). ‘Geographical Advantage: Home Market Effect in a Multi-Region World’, *Research in Economics*, vol. 71, pp. 740-758.



- [55] Mesquita Moreira M., Blyde, J., Volpe, C. and Molina, D. (2013). ‘Too Far to Export: Domestic transport costs and regional export disparities in latin america and the caribbean’, IADB.
- [56] Mori, T., Nishikimi, K. and Smith, T. (2005). ‘A Divergence Statistic for Industrial Localization’, *Review of Economics and Statistics*, vol. 87, pp. 635-651.
- [57] Nunn, N. (2007). ‘Relationship-specificity, incomplete contracts, and the pattern of trade’, *Quarterly Journal of Economics*, vol. 122, pp.569-600.
- [58] Nunn, N. and Puga, D. (2012). ‘Ruggedness: The blessing of bad geography in Africa’, *Review of Economics and Statistics*, vol. 94, pp.20-36.
- [59] Rajan, R. and Zingales, L. (1998). ‘Financial Dependence and Growth’, *American Economic Review*, vol.88, pp. 559-586.
- [60] Ramcharan, R. (2009). ‘Why an economic core: domestic transport costs’, *Journal of Economic Geography*, vol. 9, pp. 559-581.
- [61] Ramondo, N., Rodriguez-Clare, A. and Saborio-Rodriguez, M. (2016). ‘Trade, Domestic Frictions, and Scale Effects’, *American Economic Review*, vol. 106, pp. 3159-84.
- [62] Rauch, J. (1999). ‘Networks versus markets in international trade’, *Journal of International Economics*, vol. 48, pp. 7-35.
- [63] Redding, S. (2016). ‘Goods Trade, Factor Mobility and Welfare’, *Journal of International Economics*, vol. 101, pp. 148-167.
- [64] Riley, S., DeGloria, S. and Elliot, R. (1999) ‘A Terrain Ruggedness Index That Quantifies Topographic Heterogeneity’, *Intermountain Journal of Sciences*, vol. (5), pp. 23-27.
- [65] Romalis, J. (2004). ‘Factor Proportions and the Structure of Commodity Trade’, *American Economic Review*, vol. 94, pp. 67-97.
- [66] Saint-Paul, G. (1992), ‘Technological Choice, Financial Markets and Economic Development’, *European Economic Review*, vol. 36, pp. 763-781.
- [67] Santamaria, M. (2018), ‘The Gains from Reshaping Infrastructure: Evidence from the Division of Germany’, mimeo, Universitat Pompeu Fabra.

- [68] 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.
- [69] 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.
- [70] 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.
- [71] Yeaple, S. and Golub, S. (2007). ‘International productivity differences, infrastructure, and comparative advantage’, *Review of International Economics*, vol. 15, pp. 223-242.