ISLANDS AS 'BAD GEOGRAPHY.'

INSULARITY, CONNECTEDNESS, HISTORY AND TRADE COSTS

Luca De Benedictis* Anna Maria Pinna†

February 15, 2018

Abstract

In this paper we explore how geography can be changed by history, and how the past can condition the present. We focus the analysis on islands and on how insularity is related to high levels of trade costs. After controlling for geographical characteristics, for unobservable heterogeneity and for spatial connectedness, we show that islands that were touched by navigation routes between 1662 and 1855, especially those giving shelter to vessels that anchored their ports, show - three centuries later - lower trade costs with respect to similar islands that were not involved in historical navigation routes.

Keywords: Islands, Geography, Connectedness, Trade, Gravity model, Historical trade routes.

JEL Classification: F10, F14.

*Rossi-Doria Center & DED - University of Macerata. luca.debenedictis@unimc.it

†DSEA - University of Cagliari and CRENoS. ampinna@unica.it

We are grateful to the government of Sardinia (Italy) for financial support, under CRP-27162. A preliminary version of the paper has been presented at DEGIT XVIII in Lima, and subsequently at ITSG Cagliari, ETSG in Munich, at the Conference on "Economics of Global Interactions" in Bari, Ridge in Montevideo, EIEF in Rome, NHH in Bergen, Roma Tre and the Polytechnic University of Marche, Universitat de les Illes Balears and the World Bank, and the PSE. We thank Jean-François Arvis, Costas Azariadis, Antoine Berthou, Kjetil Bjorvatn, Massimiliano Bratti, Fabio Cerina, Robert Chirinko, Rosario Crinò, Giuseppe de Arcangelis, Dave Donaldson, Michael Ferrantino, Lionel Fontagné, Ewout Frankema, Giulia Felice, Stefano Gagliarducci, Russell Hillberry, Omar Licandro, Carlo Mastinu, Emanuela Marrocu, Gianmarco I.P. Ottaviano, Fabiano Schivardi, Joana Silva, Massimo Tamberi, Stig Tenold, Anthony Venables, Christian Volpe Martincus, Clive Wilkinson, Jeffrey Williamson and Alan Winters for the comments received. We also thank Vania Licio for excellent research assistantship. Luca De Benedictis wishes to thank EIEF and NHH for the hospitality. The usual disclaimer applies.

"*No man is an iland,*
intire of it selfe;
every man is a peece of the Continent,
a part of the maine; ...."
John Donne

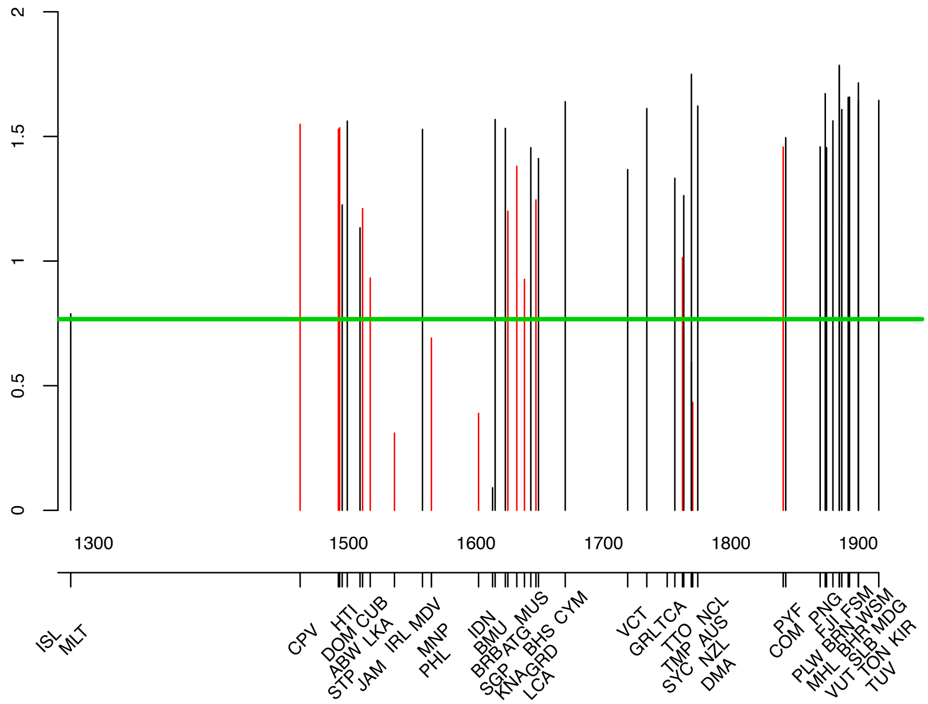
1 Introduction

In this paper we have the purpose of evaluating how spatial and historical connectedness can influence, modify and overturn natural conditions. Focusing on countries, we study how being an island affects - literally - their isolation, measured in terms of trade costs today. Subsequently, we quantify the level of spatial and historical connectedness of each island and demonstrate that islands connected to the mainland through navigation routes show - after three centuries - lower trade costs, and that this difference is statistically and economically meaningful.

The analysis comprises four steps. First, we show to what extent insular countries differ with respect to others in terms of average trade costs, in the period 1995-2010; second, we measure spatial connectedness through geographical network topology and historical connectedness through detailed information on maritime trade routes contained in vessels' logbooks between 1662 and 1855; third, we show for all countries in the dataset, that connectedness is important in reducing trade costs, and quantify the role of spatial and historical connectedness; fourth, we demonstrate that islands lying along historical navigation routes that gave shelter to vessels that stopped for relatively long periods show lower trade costs now than similar islands not on those routes. The main goal of the whole analysis is to give an account of how geography can be changed by history, and how the past can condition the present.

Evidence of the first, third and fourth steps is summarized in Figure 1, where an average comprehensive general measure of bilateral trade costs, $\bar{\tau}_{ij}$, between island i and all its trade partners j, is calculated for the period

Figure 1: Trade costs of islands



Note: The figure represents the logarithmic transformation of average bilateral trade costs (Chen and Novy, 2011) of islands between 1995 and 2010. Islands are identified by their ISO3 character code and are ordered on the horizontal axis according to the first year as a European colony. Colors distinguish islands lying along navigation routes between 1662 and 1855 (in red). The horizontal green line indicates the average level (in logs) of bilateral trade costs for continental countries ($\bar{\tau}_{ij} = 0.767$). The level of trade costs for Ireland, colonized in 1536, is $\bar{\tau}_{IRLj} = 0.31$, for the Philippines [1565] $\bar{\tau}_{PHLj} = 0.69$, for Indonesia [1602] $\bar{\tau}_{INDj} = 0.38$, for Singapore [1613] $\bar{\tau}_{SGPj} = 0.09$, for Australia [1770] $\bar{\tau}_{AUSj} = 0.43$, and for New Zealand [1769] $\bar{\tau}_{NZLj} = 0.59$. A precise definition of the insularity taxonomy will be discussed further on.

1995-2010, according to the procedure proposed by Chen and Novy (2011),[1] and is plotted against the year of the island's modern discovery by European

[1] $\bar{\tau}_{ij}$ is the (logarithmically transformed) tariff equivalent, $\frac{\bar{p}_{ij}}{\bar{p}_{ii}} - 1$, of the wedge between the consumer price, $\bar{p}_{ii}$, of the domestic good in country i and the price, $\bar{p}_{ij}$, when the same good is sold in country j, with $i \neq j$. All details on the data and procedure used will be discussed later.

colonizers.[2]

The horizontal green line refers to (the level of) $\bar{\tau}_{ij}$ for the average continental country, and is around 0.7. The $\bar{\tau}_{ij}$ for island countries is in general way above 1, with only Ireland (IRL), the Philippines (PHL), Indonesia (IDN), Singapore (SGP), Australia (AUS), and New Zealand (NZL) showing a trade cost lower that the average trade cost for continental countries. What is also evident from Figure 1 is the great variability of trade costs across islands, ranging from 0.1 to 1.75. Even islands that are geographically contiguous, such as the Dominican Republic (DOM) and Haiti (HTI), display different average trade costs; i.e. $\bar{\tau}_{DOMj} = 1.23$ and $\bar{\tau}_{HTIj} = 1.53$. Lastly, islands which were connected through historical navigation routes - depicted in red in Figure 1 – were discovered at an early stage and show, in general, lower trade costs today. This is the claim we want to support through the following analysis, in spite of the fact that current trade routes totally differ from those navigated in the XVII Century, and that the large majority of islands we discuss are no longer on those routes.

The case of islands' trade costs is both interesting in itself and methodologically challenging. 'Bad geography' is generally associated with specific latitudes (La Porta *et al.*, 1997, Spolaore and Wacziarg, 2013), terrain conformation (Nunn and Puga, 2012) or lack of direct access to the sea (Faye *et al.*, 2004). In particular, landlocked countries are often considered as unfortunate territories because their geo-political condition makes them dependent on their neighbors' territorial characteristics or policy choices. In this paper we want to evaluate if too much direct access to the sea and the absence of spatially contiguous continental neighbors - not being "... *a peece of the Continent, a part of the main* ..." - can also make trade costs so high as to hamper the participation of an island country in international economic exchanges, reducing the feasible ways of increasing per capita income and promoting economic growth (Helpman, 2009).[3]

[2] E.g. the island of Malta (MLT) - an archipelago in the Mediterranean Sea – was first settled around 5200 BC. The island was first ruled by a 'European dynasty', the Aragonese House of Barcelona, from 1282 on, which is the date considered in Figure 1. Before 1282, the island was under control of the Greeks, the Phoenicians, the Carthaginians and the Romans, and remained under the Byzantine Empire until 870, when it fell to the Arabs. After that the Normans conquered the island until Spanish colonization. The level of $\bar{\tau}_{MLTj}$ - the average trade cost with respect to its trade partners for the period 1995-2010 – is 0.78.

[3] In principle, the effect of isolation could also be a blessing instead of a curse, in an era of low transportation (e.g. sea motorways) and communication costs. Isolated countries

The analysis is however hampered by several obstacles. The first is a clear definition of insularity - that we overcome through a novel index - distinguishing it from other geographical conditions, such as e.g. small country size in terms of territorial extension or remoteness, that are as common to islands, as to other countries. The second is to tackle the difficulties associated with the quantification of trade costs (Anderson and van Wincoop, 2004), that we address by calculating a comprehensive measure of bilateral trade costs, based on a theoretically-grounded gravity model of international and domestic trade, as in Chen and Novy (2011) and Novy (2013). Focusing on islands requires building a substantially new dataset, including 191 countries and 18145 country pairs; while the emphasis on geography and spatial connectedness limits the use of traditional fixed-effects estimators to control for unobserved heterogeneity, given the time-invariance of the main covariates of interest. Finally, the quantification of historical connectedness requires the use of further new data. Along the lines of Nunn (2009), we operationalized the information contained in *all existing* logbook records of vessels traveling between European ports and the rest of the world, between 1662 and 1855 (García-Herrera *et al.*, 2005, Konnen and Koek, 2005, Russell and Cohn, 2012, Pascali, 2017). The documentary sources, mostly kept in European archives in Spain, UK, the Netherlands and France, allow to trace major navigation routes, the length of the different journeys, day-by-day, and, most of all, the anchorage and the in-harbor stops of vessels in a specific number of islands.

The complexity of the fact that past connected countries show lower trade costs now hinges on endogeneity. This is why we carefully selected the object of our investigation and elaborate an empirical strategy that imposes fewer constraints than traditional fixed or random effects estimators. Islands, aside from being surrounded by the sea, share the common characteristic of entering modern history through a fluke (Briguglio and Kaminarides, 1993, Feyrer and Sacerdote, 2009, Diamond and Robinson, 2010). Their random discovery rules out by definition the potential selection bias in the analysis (e.g. they were not selected in the past because of a convenient position in present navigation) and allows the meaningful application of the strong ignorability assumption (Imbens and Rubin, 2015), making the case of islands a clear

can develop a tourism sector based on exclusivity or become a tax haven adopting a strategy of institutional free-riding. As a matter of fact, according to the World Bank, the GDP per capita of islands is higher than that of average countries (see also Spolaore and Wacziarg (2013) on the sign of the island-dummy in growth regressions).

experiment on the effect of geography on contemporary economic outcomes. On the other hand, contemporary maritime trade routes (i.e. current transportation costs) are fundamentally different from late XIX and XX century shipping routes, as sailing technology is different (the correlation between past and present navigation routes is almost inexistent). This excludes the endogeneity of the islands' treatment effect due to persistence of historical trade routes over today's sea lanes. Moreover, the covariates used in the analysis are time-invariant (e.g. geographical characteristics) or largely predetermined (e.g historical characteristics), further excluding the possibility of reverse causality. Finally, to support the causal interpretation of the results, we control for omitted variable bias and the possible endogeneity in the classification of islands (due to man-made splitting of geographical areas, a.k.a national borders) adapting the Hausman and Taylor (1981) instrumental variable estimator to our needs.

The rest of the paper is structured as follows. A short review of the three streams of literature that are instrumental in the analysis anticipates the bulk of the data description (definition of the insularity index, measurement of bilateral trade costs and characterization of covariates, including the measures of spatial and historical connectedness) and the empirical setting. We provide novel evidence on the role of geography (as discussed in macro and international trade theory and empirics) focusing on the specificity of islands, adding results to the recent literature that explores the role of historical events in shaping conditions in modern economies. Next we discuss the empirical results. A section outlining possible issues that warrant further discussion follows. Concluding remarks are presented in the last section. An online Appendix discusses the data at length and integrates the analysis included in the present text.[4]

[4] All information on the sources of the variables, the final datasets, the Stata do-files and the R codes necessary to replicate the analysis presented in the paper and in the online Appendix are available to guarantee replicability of the analysis.

2 Building blocks: Geography, islands and historical events

2.1 Geography and economic outcomes

The role of geography in economic development has filled the research agenda of development economists examining cross-country correlates of GDP per capita (Gallup *et al.*, 1999, Spolaore and Wacziarg, 2013). While there is little doubt that geographic factors are highly correlated with economic development, there is little consensus on how this correlation should be interpreted. Geography as a key determinant of climate and temperature, natural resources, disease, ease of transport, and diffusion of technology, can directly affect productivity, human capital accumulation and the use of other factor resources.[5] Hibbs and Olsson (2004), in search of an empirical validation of Diamond (1999),[6] control for biogeographic endowments (i.e. initial biological conditions: the number of animals and plants suitable for domestication and cultivation at each location 12,000 years ago) in a cross-country regression of contemporary levels of development on geographic variables. They find supporting evidence of Diamond's hypotheses, with geography being empirically more relevant than biology.

On the other hand, several authors claim that the influence of geography on economic development is merely indirect, through institutions and trade. The very influential evidence put forward by Acemoglu *et al.* (2001, 2002), showing that after controlling for the effects of institutions, geography did not matter for economic performance in their cross-sectional sample of countries, convincingly stress the primacy of institutions over geography in causally determining the real wealth of nations. According to this view, geography plays an important secondary role, which in the specific case of Acemoglu

[5] Spolaore and Wacziarg (2013), in their magnificent survey on the 'Deep determinants' of economic growth, show how " . . . a small set of geographic variables (absolute latitude, the percentage of a country's land area located in tropical climates, a landlocked country dummy, an island country dummy) can jointly account for 44% of contemporary variation in log per capita income, with quantitatively the largest effect coming from absolute latitude (excluding latitude causes the R^2 to fall to 0.29). This result [documents] the strong correlation between geography and income per capita."

[6] Diamond (1999) traces the contemporary level of economic development of countries to biological and geographical characteristics of territories that their inhabitants were able to exploit during the Neolithic transition. See also Ashraf and Galor (2013) for a recent discussion of the issue.

et al. (2001) determines the burden of diseases on settlers, which in turn shaped the type of institutional experience of colonies, and, through this channel, influenced the type of modern institutions and the present fortunes of economies. The indirect role of geography has further been clarified by Rodrik *et al.* (2004), focusing on trade. Geography in fact is an important determinant of the extent to which a country can become integrated in world markets, regardless of its own trade policy. A distant, remote, landlocked, isolated country faces greater trade, and therefore integration, costs.

This paper contributes to the literature on the interplay between geography, institutions and trade, exploring how extreme geographical conditions, such as insularity, affect trade costs and the level of integration in world markets, and how connectedness, especially historical connectedness, modifies the natural condition of being, or not being, "*. . . a part of the main*." The issue is of relevance since, despite the importance of trade costs as drivers of the geographical pattern of economic activity around the globe, most contributions to their understanding remain piecemeal (Arvis *et al.*, 2013).

The recent literature on the gravity equation (Eaton and Kortum, 2002, Anderson and van Wincoop, 2003) has theoretically shown that the position of a country with respect to its partners, i.e. bilateral resistance, including physical distance between the countries, has to be considered relatively to its position with respect to all its feasible alternatives,[7] i.e. its multilateral resistance. In structural gravity models (Anderson and Yotov, 2010) unknown bilateral trade costs are approximated by a set of observable variables that includes geography in its monadic dimension - introducing controls for landlocked countries and islands - and in its dyadic dimension - introducing controls for border sharing, common language and common colonial past. The common practice of controlling for multilateral resistance through a set of country dummy variables (Feenstra, 2003) rules out the possibility of separating geography from all other factors contributing to multilateral resistance. Aiming to isolate the effect of insularity on trade costs, taking into account at the same time the role of multilateral resistance, this paper takes a different direction from the one of structural gravity models.

[7] See also Chaney (2008) and Helpman *et al.* (2008) on the issue of selection on foreign markets, and De Benedictis and Taglioni (2011), Head and Mayer (2014), Baltagi *et al.* (2014) on the related issue of the use of non-linear estimators in presence of many zeros in the trade matrix.

2.2 Islands as 'Bad Geography'

The role of specific geographical conditions as determinants of income and economic growth and as constraints to the effect of economic integration has received increasing attention in the literature. Milner and Zgovu (2003) and Hoekman and Nicita (2008) maintain that they are the primary reason for which developing countries are unable to benefit from trade preferences. Geographical impediments hamper international trade raising transportation costs. As Hummels (2007) points out, " ... as tariffs become a less important barrier to trade, the contribution of transportation to total trade costs ... is rising."[8] Moreover, the interest on extreme geographical conditions shown by policy frameworks such as the Almaty Program of Action (2003), the Conference on Landlocked Developing Countries (2014) or the EU Posei Program (2006), suggests that more evidence on how geography imposes costs on countries' economies is needed.

Insularity is not in generally considered the worst condition in terms of 'bad geography.' According to both empirical and theoretical literature, the most straightforward case of extreme geographical condition is the lack of direct access to the sea. Landlockness is considered as a fundamental cause of heterogeneity among countries. One out of four countries in the world is landlocked; in Africa, it is one out of three. On the contrary, having direct access to the sea is the geographical condition that has been found to be the most advantageous for a country's economy: coastal countries are wealthier and experience 30% more trade than landlocked countries (see the references in Limao and Venables (2001)). But direct access to the sea can generate extreme geographical conditions. The discontinuity between islands and the mainland increases costs by eliminating alternatives in the connection system of an island and by raising the opportunity cost of possible alternatives. The small and remote nature of the majority of island countries (Briguglio and Kaminarides, 1993, Briguglio, 1995, Mimura *et al.*, 2007, Becker, 2012) should be considered in view of these characteristics, revealing the crucial physical difference between islands and coastal countries. At the same time, not all islands are made the same.

In a recent work, Licio and Pinna (2013) constructing a new dataset,

[8] The same evidence is confirmed by Bertho *et al.* (2016), who state: "maritime transport costs today matter more than tariffs. Ad valorem MTC [maritime transport costs] of exports to the United States are on average more than three times higher than the average US tariff, and in New Zealand are more than twice as high."

discuss which dimensions are better aimed at capturing the heterogeneity of the insular state. Following Licio and Pinna (2013), we define a taxonomy of world countries, ordered according to their degree of insularity. On the lower extreme are the Landlocked countries ($\mathcal{L}$), having no access to the sea, followed by Continental countries ($\mathcal{C}$), having access to the sea and with less than 2% of their territory composed of islands,[9] and Partial islands ($\mathcal{P}$), having access to the sea and with more than 2% of their territory composed of islands, and, finally, Island countries ($\mathcal{I}$), consisting entirely of islands. This taxonomy defines a brand new ordered Index of Insularity in which all countries are islands, with a different degree of insularity.

This taxonomy is used in the following analysis, describing how islands are different in terms of trade costs and explaining why.

2.3 The long-lasting effects of historical events

Before going into detail, we need to consider a new stream of literature that is strongly related to our own analysis. As summarized by Nunn (2009), the primary goal of this literature is to examine "whether historic events are important determinants of economic development today." Acemoglu *et al.* (2001, 2002) and La Porta *et al.* (1997) paved the way to the analysis of the potential importance of an historic event - colonial rule, in both cases – for long-term economic development. From the earliest subsequent studies, dealing essentially with the correlation of historically related variables with present-day economic outcome, the literature has developed in two directions. The first one explores new identification strategies of causal effects of history, the second deals with the quantification of historical episodes, the digitalization of historical archives, the collection and compilation of new datasets based on historical data. The information content of such data has rapidly moved from sparse cross-sections to very detailed longitudinal structures.

Our contribution moves along this track, quantifying the information contained in all available logbook records for the period 1662 to 1855, drawing on British, Dutch, French and Spanish archives.[10] The data, extracted from

[9] In Licio and Pinna (2013), the authors distinguish continental countries without islands from those having a negligible part (less than 2%) of territory in islands. We collapse these two categories into a single one.

[10] Portuguese archives no longer exist due to the Big Fire that destroyed Lisbon at the end of the XIX century; US naval archives cover a period subsequent to the one we explore; Chinese historical archives are not available to the general public, but China was entering

the original CLIWOC climatology database (a more detailed description of the database is included in Section 3), allows to describe the main navigation routes in the XVIII and XIX centuries. Having this information (the day-by-day length of the different journeys, the islands touched by the routes, the anchorage of vessels) at an international level is unique. We fully exploit the quality of the data in quantifying the emergence of a *culture of openness* in an international context, due to historical connectedness.

This is however not the first contribution on the role played by historical roads or communication routes in shaping the geographical distribution of contemporary economic outcomes (O'Rourke and Williamson, 2001). Dell (2010) in her seminal work on the persistent effect of Peru's mining mita shows that the geographical propagation of the negative effect of the forced labour system, instituted by the Spanish government in Peru and Bolivia in 1573, is related to the road system, and today mita districts still remain less integrated into road networks. Volpe Martincus *et al.* (2012) use the (distance to the) Inca road network for instrumenting the present road network, addressing the potential endogeneity of transportation infrastructure to domestic and international trade. Similar analyses on transport infrastructure have been conducted by Fajgelbaum and Redding (2014) for Argentina, by Banerjee *et al.* (2012) and Faber (2014) for China, Donaldson (2016) for India, Jedwab *et al.* (2014) for Kenya and Donaldson and Hornbeck (2016) for the US. Two papers share with our own a common interest in islands and historical navigation routes. Feyrer and Sacerdote (2009) examine whether colonial origins affect modern GDP per capita of islands. They focus the analysis on Atlantic, Pacific and Indian Oceans islands arguing, as we do, that the random nature of the discovery and colonization makes islands a natural experiment and allows to use an instrumental variable strategy for the length of colonization, using wind direction and wind speed. Differently, Pascali (2017), searching for an asymmetric exogenous shock to explore how an increase in trade affected economic development, uses the change in maritime trade routes generated by the emergence of steamships in the 1870s. To the best of our knowledge, Pascali (2017) is the only paper that shares with us a multi-country approach to the long-lasting effects of historical maritime transportation events.

into an autarchic period in the XVI century, a period that lasted 300 years (Diamond, 1999). We will discuss the issue of geographical selection of historical data later on, and we account for that in the empirical part of the analysis.

3 Navigation routes between 1662 and 1855: The vessels logbooks database

In the empirical analysis that follows, we make extensive use of the data included in the CLIWOC database (García-Herrera *et al.*, 2005, Konnen and Koek, 2005). The Climatological Database for the World's Oceans 1662-1855 was constructed between 2000 and 2003 by several institutions, European and non-European universities and research institutes under the EU funded EVK2-CT-2000-00090 project. The goal of the project was to collect and digitalize meteorological information reported in all available ships' logbook records contained in national archives for the period under scrutiny.

The version of the database we re-elaborate in order to provide a new information source on world territories interested by historical trade routes is the 2.1 version, released in 2007. The first trip in the database is from a Dutch ship, the Maarseveen, which left Rotterdam bound for Batavia, on October 15, 1662, while the last one is again for a Dutch ship, the Koerier, which left Curacao returning to the Netherlands on June 21, 1855. The total number of logbooks included in the dataset is 1,758 comprising 287,114 daily observations. The database provides daily information on 5227 voyages (during some of them data were recorded at different times each day). The period spans from 1662 to 1855, but the database concentrates mainly on navigations after 1750.[11] Several European countries were simultaneously involved in voyages to the opposite part of the world. English and Dutch routes show a higher density in the data but the richness of CLIWOC information stems from the fact that Spanish and French navigations are also present and spread over the years after 1750.

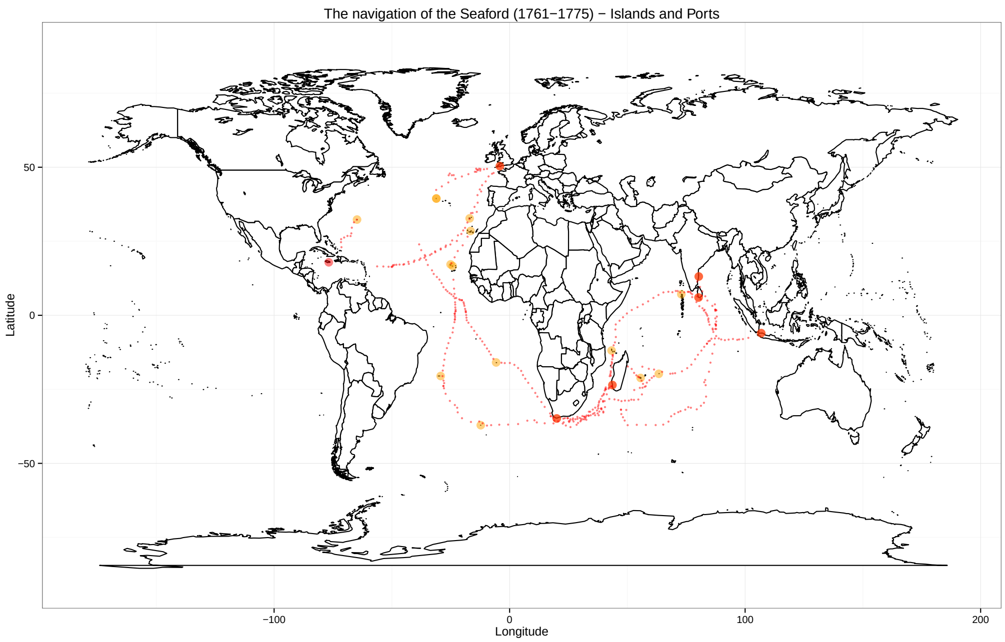
Logbooks included general information on the state of the vessel, the name of the captain, the port of origin and the destination; wind direction and strength and vessel's speed: logbooks also registered other aspects of the weather and precipitation, the state of the sea and sky, thunder, lightning, and eventually the proximity of mainland. Every record in the logbooks

[11] The number of trips from 1662 to 1749 are only 13. We listed them in Table 2 in the online Appendix showing how information on locations is reported originally in the database. The identifier for each trip and calculations of the number of days of the journey is from our elaborations. The recorded navigations are based on 1922 historical ships. The actual number of ships is 2010, some of them have the same navigation data recorded in more than one archive with different logbook numbers giving rise to duplicates in the trips' records and number of ships.

includes the location of the vessel, in terms of longitude and latitude.

Figure 2:

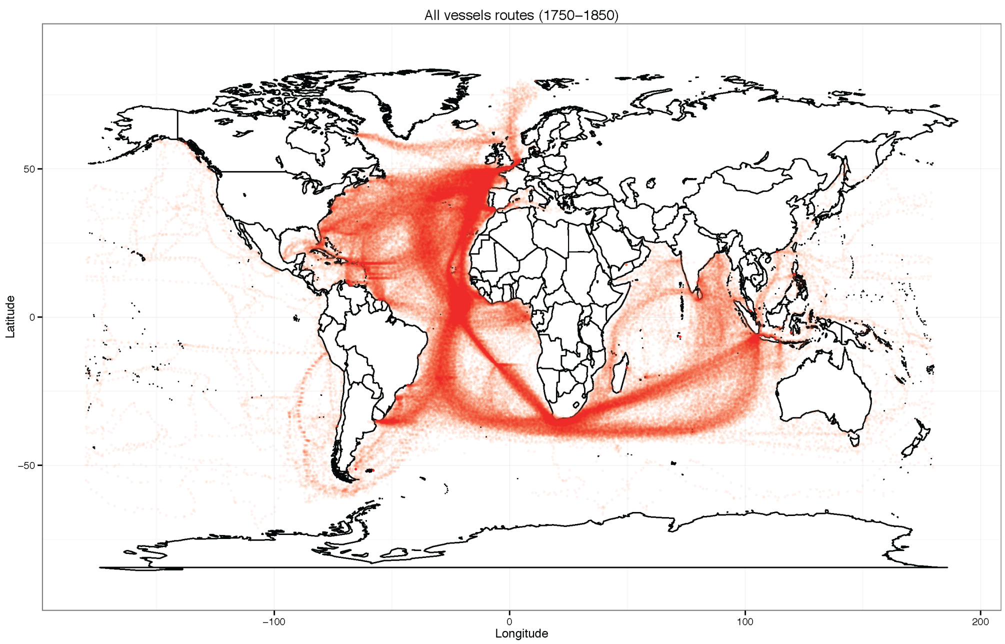
The navigation of the Seaford



Note: Figure 2 depicts the navigation of the vessel Seaford between 1761 and 1775. Yellow spots indicate the islands touched by the Seaford; red spots indicate ports. The latitude and longitude of sailing days is depicted by the red dotted line. Data drawn from the CLIWOC database. Elaborations are our own.

In Figure 2 we describe, as an example, the navigation of one single vessel included in the records of the CLIWOC database: the Seaford, leaving Plymouth on the first day of February 1761 with destination Madras, in India. It anchored in Madras on the 5th of July, 1761, after six month sailing. It then continued its journey until March 1775. The last record we have of the vessel corresponds to a logbook note written when leaving the Bermudas Islands. During fourteen years of traveling the Seaford touched the ports of Cape Town, St. Marys Road in Madagascar, Point Galle in Ceylon (Sri Lanka), Jakarta (Indonesia) and Jamaica. It also stopped for a few hours or several days in Tenerife in the Canaries Islands, at Capo Verde, in the Island of Trindade (Brazil), in the Island of Tristian De Cunha (UK), in the Mauritius, and in the Comoros Islands. In Figure 2 we marked the islands touched by the Seaford with a yellow spot, and ports with a red spot, while the latitude and longitude of sailing days are depicted by the red dotted line.

Figure 3:
Trade routes: 1750-1850



Note: Figure 3 depicts the navigation of all vessels between 1750 and 1850. The latitude and longitude of sailing days are depicted by the red dotted line. Data are drawn from the CLIWOC database release 2.1. Elaborations are our own and will be described in a later section.

In the same way we are able to trace all routes sailed by all vessels in the database. As shown by the journeys of the Seaford, it is during a sailing trip that islands could have had a special role. Islands are, in fact, the first most likely territories to be encountered by ships navigating the oceans. In Figure 3 we plot all available observations on the spatial position of vessels between 1750 and 1855. Major routes are immediately visible and it is also relatively simple to keep records of the islands touched by the different routes.

We will take advantage of this information later on, as well as of that on the anchorage and ports of call of vessels in a specific number of islands.

4 Measuring trade costs

It is now time to focus on our dependent variable: trade costs.

To produce a comprehensive aggregate measure of bilateral trade costs, that takes into account all possible costs associated with international trade, we built upon some insights from the structural gravity equation literature (Anderson and van Wincoop, 2003, Anderson and Yotov, 2010, 2012, Fally, 2015).[12]

From Anderson and van Wincoop (2003) we know that the bilateral trade flow between country i and country j can be expressed[13] as:

$$X_{ij} = \frac{Y_i}{\Pi_i^{1-\sigma}} \cdot d_{ij}^{1-\sigma} \cdot \frac{E_j}{P_j^{1-\sigma}} \tag{1}$$

where Y_i is total output in country i, E_j is total expenditure in country j, d_{ij} is a measure of bilateral distance, that must be interpreted as a trade cost factor: the gross bilateral cost of importing a good (one plus the tariff equivalent), $p_{ij} = d_{ij}p_i$; and $\sigma > 1$ is the elasticity of substitution between varieties in a Dixit-Stiglitz utility function.[14] The terms $\Pi_i^{1-\sigma}$ and $P_j^{1-\sigma}$ are

[12] Also the World Bank (Arvis *et al.*, 2013) has recently produced a sectoral measure of the same class of indices used in this analysis, using the Inverse Gravity Framework methodology (Novy, 2013). The Trade Costs Dataset (ESCAP and World Bank, 2013), which is the result of this computational effort, provides estimates of bilateral trade costs in agriculture, manufactured goods and total trade for the period 1995-2010. It includes symmetric bilateral trade costs for 178 countries, computed for each country-pair using bilateral trade and gross national output. The Trade Costs and our own datasets do not fully overlap either in the time series or in the cross-sectional dimension. Furthermore, our measure builds on the revision of the Comtrade database by the Cepii Research Center, whose work is aimed at producing comparable numbers between export values declared at origin and import values declared at destination, using mirror statistics.

[13] Even if the obtained measure is time-variant, for the sake of simplicity, we disregard the time subscript, t, from the notation. Moreover, the index can be applied to sectoral data, X_{ij}^k, without any substantial change in the way the equation is expressed for trade flows in sector k.

[14] As emphasized by Novy (2013) and Head and Mayer (2014), equation 1 can be derived from different trade models. In spite of being consistent with Armington (1969) preferences, the parameter σ would indicate the constant elasticity of substitution between varieties in a monopolistic competition trade model á la Krugman. In Melitz (2003) and Chaney (2008) the same parameter refers to the exponent of the Pareto distribution of firms' productivity (the higher σ the lower would be the productivity dispersion among firms). Finally, in Eaton and Kortum (2002) the parameter σ indicates the exponent of the Fréchet distribution defining the countries' productivity across product varieties. See

the"inward" and "outward" multilateral resistance terms, capturing the interconnectedness among countries that is revealed through the price index in the importing market, $P_j^{1-\sigma}$, and through the price index $\Pi_i^{1-\sigma}$ capturing the degree of competition faced by the exporting country. After Anderson and van Wincoop (2003), the multilateral resistance terms highlight the fundamental relevance of considering distance in relative terms, and not only in absolute terms, as expressed by d_{ij}.

Denoting by σ the elasticity of substitution among product varieties, the varieties considered in the expenditure function of consumers must necessarily include both domestic varieties and foreign varieties. Accordingly, the gravity Equation 1 should consider not only foreign trade but also domestic trade, X_{ii} and X_{jj}.[15] On that we follow Jacks *et al.* (2008), Chen and Novy (2012) and Novy (2013).

Denoting by N the total number of countries, for consistency we must have that:

$$Y_i \equiv \sum_{i \neq j}^{N-1} X_{ij} + X_{ii}; \tag{2}$$

$$E_j \equiv \sum_{i \neq j}^{N-1} X_{ij} + X_{jj}. \tag{3}$$

X_{ii} and X_{jj} are in general not observed and must therefore be estimated or - as in our case - can be calculated using Equation 2 and Equation 3.

Imposing $Y_j \equiv E_j$, we obtain for country i domestic trade,

$$X_{ii} = Y_i Y_i \left(\frac{d_{ii}}{\Pi_i P_i} \right)^{1-\sigma}, \tag{4}$$

from which we can isolate the multilateral resistance variables:

$$\Pi_i P_i = \left(\frac{X_{ii}}{Y_i Y_i} \right)^{\frac{1}{\sigma-1}} d_{ii}, \tag{5}$$

also De Benedictis and Taglioni (2011) on this point.

[15] The domestic trade component is usually disregarded in gravity models, making the model inconsistent with the data. Wei (1996) derives domestic trade in order to derive the notion of home bias from a micro-founded gravity equation. According to his definition: " ...a country's home bias ...[is the] imports from itself in excess of what it would have imported from an otherwise identical foreign country (with same size, distance and remoteness measure)." See also Wolf (2000).

and $\Pi_i P_j$ can be measured using domestic variables only (given σ). We exploit this to solve the gravity model for bilateral trade costs.

Let us now multiply the gravity equation 1 by the corresponding gravity equation for trade flows in the opposite direction: X_{ji}. We obtain,

$$X_{ij}X_{ji} = (Y_i Y_j)^2 \left(\frac{d_{ij}d_{ji}}{\Pi_i P_j \Pi_j P_i}\right)^{1-\sigma}. \qquad (6)$$

substituting for $\Pi_i P_i$ and $\Pi_j P_j$, we end up with the following expression:

$$\frac{d_{ij}d_{ji}}{d_{ii}d_{jj}} = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{\sigma-1}}. \qquad (7)$$

Jacks *et al.* (2008) show that, taking the square root of equation 7 and deducting one it is possible to get an expression for the tariff equivalent:

$$\mathcal{T}_{ij} \equiv \left(\frac{d_{ij}d_{ji}}{d_{ii}d_{jj}}\right)^{\frac{1}{2}} - 1 = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{2(\sigma-1)}} - 1, \qquad (8)$$

where $\mathcal{T}_{ij}$ is the geometric average of international trade costs between country i and country j, relative to domestic trade costs within each country.

The support of the variable is $\mathcal{T}_{ij} \in [-1, \infty)$, that empirically is restricted to $\mathcal{T}_{ij} \geq 0$.[16] When $d_{ij}d_{ji} = d_{ii}d_{jj}$ this leads to $\mathcal{T}_{ij} = 0$, which means that for a country it is equivalent to trade domestically and internationally. In other words, there is no particular advantage in trading at the domestic level instead of at the international one. In this case international borders play no role.

In fact, $\mathcal{T}_{ij}$ is a tariff equivalent of comprehensive trade costs; an indirect measure that includes everything (transport, trade policy, culture, institutions, etc): e.g., if the value of $\mathcal{T}_{ij} = 0.5$, this implies that the price in country

[16] Intuitively, when countries trade more internationally than they do domestically that is reflected in low trade costs, that will be high in the opposite case. The benchmark case, that is usually taken as a lower bound for $\mathcal{T}_{ij}$, is when in both countries total output is equally traded inside and outside the country. In this event $\mathcal{T}_{ij} = 0$ for all levels of σ. When in the hypothetical case the domestic trade of one of the two countries is null, the ratio $\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}} = 0$, and for any given σ, $\mathcal{T}_{ij} = -1$. No such events happen in the dataset used here. The online Appendix contains descriptive statistics on the distribution of $\mathcal{T}_{ij}$ and on some limited cases of unusual behavior of the index. As the ratio in equation 8 rises above one, with countries trading more domestically than internationally, international trade costs rise relative to domestic trade costs, and $\mathcal{T}_{ij}$ takes positive values that reach the upper bound of the index, when countries do not trade internationally and $\mathcal{T}_{ij} = +\infty$.

j of a good imported from i is increased with respect to the mill-price by the trade cost factor: $p_{ij} = p_{ii}(1 + 0.5)$.

Since the index is a product of the two countries' trade flows, the level of trade of one country influences the trade cost of the other country at the bilateral level. In this respect, $\mathcal{T}_{ij}$ is a symmetric measure of bilateral trade costs. Its construction and interpretation is model-dependent, and can be time-variant and sector specific, if this is required by the analysis. In our case, we use it at the aggregate level, and we calculate $\mathcal{T}_{ij}$ for the period 1995-2010, obtaining the mean value for every pair of countries in the database. To reduce the effect of the skewed distribution of $\mathcal{T}_{ij}$, in the empirical analysis we log-transformed the trade cost measure: $\tau_{ij} \equiv ln\mathcal{T}_{ij}$.

5 Descriptives

The comprehensive measure of bilateral trade cost, $\mathcal{T}_{ij}$ (and its log transformation, τ_{ij}), is calculated as in Equation 8 for 191 countries and 18145 country pairs. We used bilateral trade data from the Cepii revision of the Comtrade UN database (Gaulier and Zignago, 2010) to derive the aggregate measure of bilateral trade x_{ij}, and we calculate internal trade x_{ii} using data on GDP reported in the World Bank WDI dataset. As for σ, we use estimates from the literature on trade elasticity (Eaton and Kortum, 2002, Anderson and Yotov, 2012), mainly working with $\sigma = 11$.[17] The domain of existence of $\mathcal{T}_{ij}$ imposes some considerations on the treatment of zero-trade values. When countries do not trade internationally our measure is indefinite. In this case we would lose all observations with really high international trade costs with respect to internal ones. We include all the zero trade observations by transforming zero trade values (in dollars) into ones. Following this procedure we calculate our measure for every year between 1995 and 2010 (details of the time series are included in the online Appendix), and we construct averages at the pair level.[18]

[17] We also used different levels of σ from 9 to 5 to check for the robustness of the results. In general, higher elasticity implies lower trade costs.

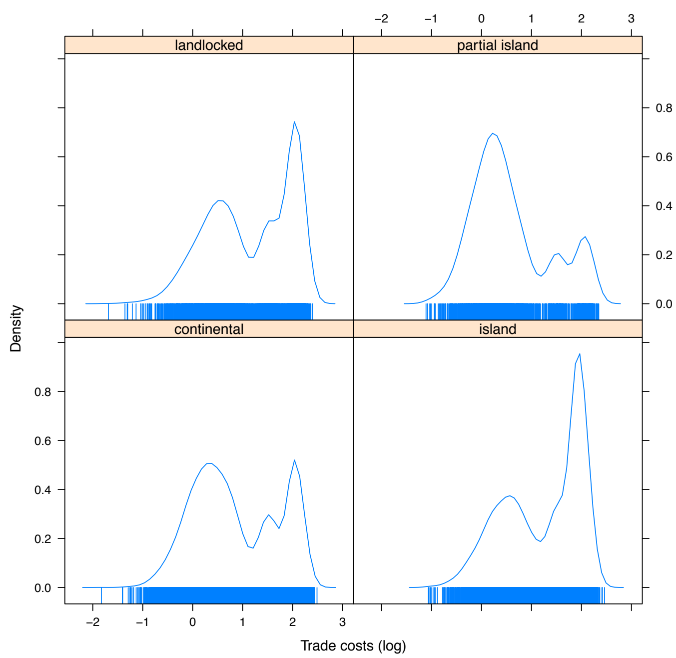
[18] A possible extension of the analysis would be to split the data into two different samples, before and after the Great Trade Collapse in 2008, to account for different trade costs in periods of growing world trade or downward sloping trade trend.

5.1 The distribution of trade costs

Figure 4 illustrates the empirical distribution of τ_{ij} for four categories of the Insularity taxonomy between 1995 and 2010. Among the 191 countries included in the dataset, 32 are $\mathcal{L}$ (17%), 88 are $\mathcal{C}$ (46%), 17 are $\mathcal{P}$ (9%) and 54 are $\mathcal{I}$ (28%). The kernel densities clearly show a trimodal empirical distribution with a different balance between the low, medium and high values of trade costs.

Figure 4:

Kernel density of trade costs: Country pairs averages, 1995-2010



Note: Our elaborations on BACI-CEPII data and the World Bank WDI data. τ_{ij} is computed according to equation 8. The smoothing parameter of the kernel function is set optimally. The spikes aligned below the kernel densities depict the position of each observation.

Striking differences appear across the classes of the taxonomy. $\mathcal{L}$-countries and $\mathcal{I}$-countries show higher levels of trade costs, and a prevalent right mode; in $\mathcal{C}$-countries and, especially, in $\mathcal{P}$-countries the right mode is less accentuated. This last group of countries is remarkably characterized by very low τ_{ij}. Country pairs with minimal trade costs are however present in all groups of countries, as shown by the left outliers in the levels of τ_{ij} visualized by the spikes below the kernel densities.[19]

5.2 Trade Costs and Insularity

To further explore the relationship between trade costs and insularity, we show in the top panel of Figure 5 a spatial scatter plot having longitude on the horizontal axis and latitude on the vertical axis, as in a cartogram, and where countries, identified by Iso3 codes, are located according to the geographical coordinates of their capital cities. The dots are colored according to the four levels of the Insularity index (black=$\mathcal{L}$; green=$\mathcal{C}$; blue=$\mathcal{P}$; red=$\mathcal{I}$). The size of the dots is proportional to the level of the average country τ_{ij} between 1995 and 2010.[20]

In the $\mathcal{I}$ group, as for the others, we have countries with very high average bilateral trade costs, such as Tonga (TON), and countries with low average bilateral trade costs, such as the United Kingdom (GBR) or Singapore (SGP). In general, as previously observed in Figure 4, high or low trade costs do not seem to be a peculiar feature of any specific group of countries in the Insularity index.

The bottom panel of Figure 5 uses the same data, measuring the average bilateral distance for country i on the horizontal axis, and average bilateral trade costs of the same country on the vertical axis, both in logs. Every country i is therefore identified by a pair of values, the first is the average bilateral distance with every trading partner j, the second is the average bilateral trade cost (in logs) 1995-2010, between i and its trade partners. Countries are identified by Iso3 codes and dots are colored according to levels of the Insularity index, as in Figure 5. In this case, to highlight the variability of bilateral trade cost for every country i, the size of the dots is proportional to the standard deviation of the country τ_{ij}, 1995-2010.

Let us take Italy (ITA) as an example; the country has low average bilateral trade costs and also a low standard deviation of τ_{ij}, but it trades with

[19] The dynamics of the changes in τ_{ij} occurring between 1995 and 2000 is evidence of the strong persistence in the distribution of trade costs, even in periods of international turmoil. During this time span, the changes appear to be more relevant for $\mathcal{C}$ and $\mathcal{L}$-countries, much less so in $\mathcal{I}$-countries, especially in terms of high levels of trade costs. Details of the temporal evolution of τ_{ij} are included in the online Appendix.

[20] See also the 3D version of this scatterplot (including a nonparametric surface visualizing estimated trade costs) and the related heat map included in the online Appendix.

Figure 5:

Spatial scatter plot: trade costs (1995-2010), insularity and distance

Note: Our elaborations on BACI-CEPII data and the World Bank WDI data. τ_{ij} is computed according to equation 8. Dots are colored according to the different levels of the Insularity index. Countries are identified by Iso3 codes. In the top panel, the size of the dots is proportional to the average trade cost of the country between 1995 and 2010, in logs; in the bottom panel, the size of the dots is proportional to the standard deviation of the country τ_{ij} 1995-2010 (average).

countries located at a low average distance. Taking the United States (USA) as a comparison, the two moments of the distribution of bilateral trade costs are quite similar but the US trades on average with partners which are at a greater distance with respect to Italy.

The general tendency is a positive correlation between distance and trade costs. This tendency is accentuated for $\mathcal{L}$-countries and $\mathcal{I}$s, less so for $\mathcal{C}$-countries. Variability of $\bar{\tau}_{ij}$, measured by its standard deviation, is substantially unrelated to distance. Finally, $\mathcal{I}$-countries can be roughly divided into two groups for each variable of interest: in terms of trade costs, the bottom panel of Figure 5 shows a prevalence of islands with high $\bar{\tau}_{ij}$, but also a substantial number of islands with low $\bar{\tau}_{ij}$ (i.e. Singapore (SGP)); in terms of average bilateral distance, the islands in the Pacific Ocean are all characterized by trade with countries located at a very long distance, while islands in the Atlantic Ocean are not. European islands form a third separate group. As a side evidence, large islands (such as Australia (AUS), Indonesia (IDN), Japan (JAP), and the United Kingdom (GBR)) are associated with low trade costs. We will return to this issue later on.

Table 1: Trade costs: Insularity index.

$\mathcal{I}_{ij}$	$\mathcal{L}_{ij}$	$\mathcal{P}_{ij}$	N	mean (log)	geometric mean (level)	sd	se
0	0	0	7132	0.767	2.326	0.756	0.009
0	0	1	2890	0.456	1.637	0.727	0.014
0	0	2	272	0.212	1.259	0.692	0.042
0	1	0	5440	1.094	3.198	0.757	0.010
0	1	1	1088	0.785	2.280	0.718	0.022
0	2	0	992	1.277	3.799	0.778	0.025
1	0	0	8824	1.217	3.549	0.747	0.008
1	0	1	1768	0.873	2.498	0.804	0.019
1	1	0	3320	1.493	4.616	0.607	0.010
2	0	0	2640	1.256	3.687	0.746	0.015

A short summary of the (arithmetic and geometric) average level of τ_{ij} and $\mathcal{T}_{ij}$ is included in Table 1. For every combination of the Insularity index categories we report the number of cases (e.g. among the possible 88×87=7656 pairs of continental countries, only 7132 are included in the database), the level of the arithmetic mean of τ_{ij}, including that of continental countries plotted as a reference line in Figure 1, its standard deviation and standard errors and the geometric mean of $\mathcal{T}_{ij}$

The highest level of τ_{ij} is for pairs of trade partners in which one of the two is an $\mathcal{I}$ and the other is $\mathcal{L}$. All things equal, this is the worst possible combination (τ_{ij}=1.493). In second place are two $\mathcal{L}$-countries (τ_{ij}=1.277); third comes $\mathcal{I}$ (one island: τ_{ij}=1.217; two islands: τ_{ij}=1.256). Islands are also characterized by high variability in trade costs, having a standard deviation of 0.804.

5.3 Geographical covariates

The first hypothesis is that islands are characterized by geographical specificities - besides being an island - that are different from other countries. The denomination of "island" could, therefore, hide some relevant geographical dimension that could explain why "islands" look so different in their trade costs.

Table 2: Linear correlation among geographical covariates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Ruggedness$_i$	1.000							
(2) Distance to ice-free coast$_i$	-0.013	1.000						
(3) Tropical$_i$	-0.214*	-0.291*	1.000					
(4) Average temperature$_i$	-0.282*	-0.352*	0.667*	1.000				
(5) Average precipitation$_i$	-0.020*	-0.394*	0.729*	0.404*	1.000			
(6) Distance from the equator$_i$	0.111*	0.186*	-0.768*	-0.885*	-0.485*	1.000		
(7) Land area$_i$	-0.101*	0.468*	-0.165*	-0.234*	-0.177*	0.151*	1.000	
(8) Distance$_{ij}$	-0.017*	-0.106*	0.234*	0.145*	0.259*	-0.172*	0.012	1.000

Note: * indicates significance level of $p < 0.001$. In the first row (labels) (1) stands for Ruggedness$_i$; (2) for Distance to ice-free coast$_i$; (3) for Tropical$_i$; (4) Average temperature$_i$; (5) for Average precipitation$_i$; (6) for Distance from the equator$_i$; (7) Land area$_i$; (8) Distance$_i$. Further description of the data and of the data sources is included in the online Appendix.

The geographical dimensions that we consider, and that we use as covariates in the empirical model described in Section 6, are ruggedness (Terrain Ruggedness Index),[21], distance from nearest ice-free coast, percentage of

[21] The Terrain Ruggedness Index (TRI) was originally devised by Riley *et al.* (1999) to quantify topographic heterogeneity in wildlife habitats. It is calculated as the sum change in elevation between a grid cell and its eight neighbor grid cells. Source: Nunn and Puga (2012), who constructed country averages.

tropical territory, annual average temperature, annual average precipitation, land area (of individual countries and country pairs), continental dummies,[22] distance from the equator,[23] and percentage of the land surface area of each country in a tropical climatic zone.[24] The last geographical variable, bilateral distance (between countries' centroids), has been calculated according to the great circle distance formula. More information on each variable and on data sources is included in the online Appendix.

The geographical variables are correlated as summarized in Table 2. The highest correlation is, unsurprisingly, between being a tropical country and average temperature, average precipitation, and distance from the equator. All other variables capture different geographical characteristics of countries.

5.4 Connectedness

It is now time to go back to John Donne's *Meditation*. Islands are not always severely isolated, some times for some of them it is easier to be "... *a peece of the Continent, a part of the maine* ...." Geographical proximity with the mainland is probably the first candidate to explore in order to evaluate how connectedness with foreign countries can reduce the onus of islands bilateral trade costs.

One way to represent the relevance of spatial connectedness is that which changes the visual perspective of the adjacency between countries taking space (latitude and longitude) in the background. This perspective is offered by a network visualization of countries proximity, as represented in Figure 6.

The figure visualizes the spatial connectedness between countries, identified by their Iso3 codes. Countries that share a common land border are connected by a link;[25] light blue nodes are $\mathcal{I}$-countries, yellow nodes are mainland countries. The nodes within a thick black circle are $\mathcal{L}$-countries,

---

[22] All these geographical variables are drawn from the NASA Socioeconomic Data and Applications Center (2009) database.

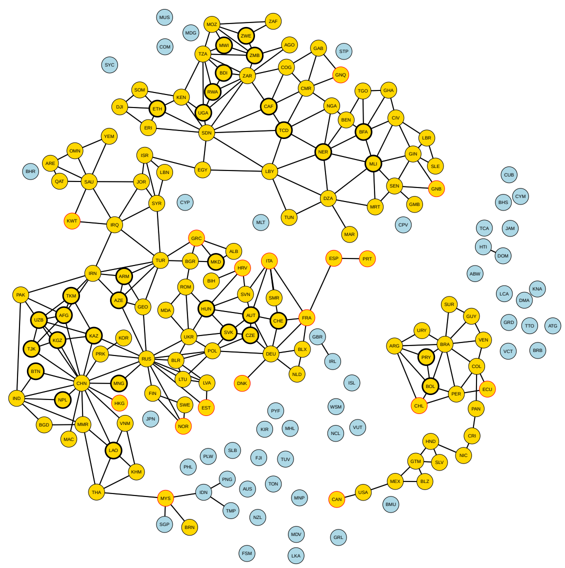
[23] The variable was originally coded by La Porta *et al.* (1997).

[24] According to the four Kappen-Geiger tropical climates (Nunn and Puga, 2012).

[25] The length of the link is endogenously determined by the visualization algorithm and has no specific meaning. The adjacency matrix that corresponds to the network in Figure 6 is a binary matrix. Only two alternatives are considered: sharing or not sharing a land border. Pakistan (PAK) and India (IND) share a common land border, as do India and Bangladesh (BGD). The difference in the length of the two different links is due to the general effect of global adjacency (e.g. Pakistan is linked also to Iran (IRN) that is not linked to India, and that drives India and Pakistan far apart).

Figure 6:

A network visualization of countries' spatial connectedness



Note: The figure represents spatial connectedness between countries. Links connect countries that share a common border; light blue nodes are Islands, yellow nodes are Mainland countries. The nodes within a thick black circle are Landlocked countries ($\mathcal{L}$), while nodes within a red circle are Partial islands ($\mathcal{P}$), as defined in Section 2.2. The position of each node depends on its relative connectedness as in the Kamada Kawai algorithm for network visualization. Islands are located within the closer country according to the intervals $\leq$300 kms, and $\leq$500 kms. Data are drawn from the CEPII database. Elaborations are our own.

while nodes within a red circle are $\mathcal{P}$-countries, as defined in Section 2.2. The position of each node depends on its relative spatial connectedness as obtained through the use of a "brute force algorithm" for network visualization.[26] Islands are located near the closest country according to the intervals $\leq$300 kms, and $\leq$500 kms.

The spatial network is characterized by a giant component of directly and indirectly connected nodes and by a second component consisting of countries of the Americas. The majority of $\mathcal{I}$ are isolated, while some of them (e.g the United Kingdom (GBR) and Ireland (IRL), or the Dominican Republic (DOM) and Haiti (HTI)) are locally connected.

[26] The algorithm used in the visualization in Figure 6 is the Kamada Kawai algorithm.

$\mathcal{L}$-countries are located in the inner part the network. Some of them show a high level of local centrality (e.g. Niger (NER) and Mali (MLI), or Hungary (HUN) and Austria (AUT)), measured by the number of neighboring countries. $\mathcal{P}$-countries play the role of gatekeepers and are located at the boundaries of the network's components. The position of $\mathcal{I}$-countries in the topology of the spatial network depends on nearby countries. The United Kingdom has a central role in the networks, while Micronesia (FSM) is quite isolated. But the topology of the spatial network reveals an important element of the relative spatial position of countries: true isolation is rare and countries, even islands, could be considered, at least in principle, as potential "... *parts of the maine*."

To capture spatial connectedness we define a nested index that progressively considers countries according to their level of spatial proximity. At the first level the index takes a value of one if the two countries i and j share a common land border; at the second level the index adds in the countries whose bilateral distance is below the limit of 300 kms; at the third level the index includes also those countries with bilateral distance of less than 500 kms. All remaining country pairs are given a value of zero.[27]

The spatial connectedness index is used in subsequent regressions to control for the role of the relational dimension in geography in influencing trade costs.

6 The empirical model

Our benchmark model is a simple anova model, where trade costs in logs are regressed against the insularity index previously described. Operationally, the model takes the expression in Equation 9 and is estimated using a least squares approach, under the classical assumption of linearity and additivity.

$$\tau_{ij} = \iota_{ij}\beta + \epsilon_{ij} \equiv \underbrace{\beta_0 + \beta_1\mathcal{I}_{ij} + \beta_2\mathcal{I}\mathcal{I}_{ij} + \beta_3\mathcal{L}_{ij} + \beta_4\mathcal{L}\mathcal{L}_{ij} + \beta_5\mathcal{P}_{ij} + \beta_6\mathcal{P}\mathcal{P}_{ij}}_{\text{Insularity Index}} + \epsilon_{ij} \tag{9}$$

[27] Alternative spatial contiguities have also been explored. Besides the ones included in the analysis reported in the paper (Nearest neighbor (300 km), and Nearest neighbor (500 km)), also Queen contiguities, Nearest neighbor (3 neighbors), Nearest neighbor (5 neighbors) have been considered. They are all depicted in the online Appendix.

The deterministic part of τ_{ij} depends only on the components of the insularity index, $\mathcal{I}$, $\mathcal{L}$, $\mathcal{P}$, $\mathcal{C}$, structured as mutually exclusive dummy variables that, given the dyadic nature of the data, take the following coding:

$$\mathcal{I}_{ij} = \begin{cases} 1 & \text{if one of the two countries in the pair, being } i \text{ or } j\text{, is an island} \\ 0 & \text{otherwise;} \end{cases}$$
$$\mathcal{II}_{ij} = \begin{cases} 1 & \text{if both the } i\text{th country and the } j\text{th country are islands} \\ 0 & \text{otherwise;} \end{cases}$$
$$\mathcal{L}_{ij} = \begin{cases} 1 & \text{if one of the two countries in the pair, being } i \text{ or } j\text{, is landlocked} \\ 0 & \text{otherwise;} \end{cases}$$
$$\mathcal{LL}_{ij} = \begin{cases} 1 & \text{if both the } i\text{th country and the } j\text{th country are landlocked} \\ 0 & \text{otherwise;} \end{cases}$$
$$\mathcal{P}_{ij} = \begin{cases} 1 & \text{if one of the two countries in the pair, being } i \text{ or } j\text{, is a partial island} \\ 0 & \text{otherwise;} \end{cases}$$
$$\mathcal{PP}_{ij} = \begin{cases} 1 & \text{if both the } i\text{th country and the } j\text{th country are partial islands} \\ 0 & \text{otherwise.} \end{cases}$$

Continental countries ($\mathcal{C}$) act as the reference category, τ_{ij} is the average level of trade costs (in logs) between 1995-2010 and ϵ_{ij} is an error term, clustered at the country-pair level. The model provides a measure of the heterogeneity of τ_{ij} across different levels of the insularity taxonomy, ι_{ij}: $\mathcal{I}$-states and $\mathcal{L}$-countries lie at the edges of the spectrum, while countries where a portion of the territory is insular are at an intermediate level.

Results from the estimation of this simple model are presented in Table 3. Standard errors have been computed so as to correct for the pair effect in all except for Model 1, where only the intercept is included and it is equivalent to the geometric mean of τ_{ij}. Our categorical variables capture the deviation in τ_{ij} when trade partners vary across groups. Coefficients in Model 2 give a measure of how much higher is τ_{ij} when one country in the pair is an island ($\mathcal{I}_{ij} = 1$) and when both are ($\mathcal{II}_{ij} = 1$). The presence of a second island in the ij pair does not increase trade costs when the reference category includes all other levels of the insularity taxonomy.[28] The specification including all categories (Model 3) measures the average deviation $\mathcal{L}$ (zero insularity), $\mathcal{P}$ (partial insularity) and $\mathcal{I}$ (full insularity) have with respect to the reference category, for which $\hat{\tau}_{ij} = 0.8521$.

[28] The coefficient of the variable $\mathcal{II}_{ij} = 1$ gives a measure of the differential effect associated with the case when both countries in the pair are Islands with respect to the case when only one is. Magnitude is small and not significantly different from zero.

Table 3: Trade Costs and the Insularity Index

	Model 1	Model 2	Model 3
	OLS	OLS	OLS
Intercept	1.0876***	0.9013***	0.8521***
	(0.0043)	(0.0086)	(0.0111)
$\mathcal{I}_{ij}$		0.3837***	0.4058***
		(0.0124)	(0.0124)
$\mathcal{II}_{ij}$		0.0199	0.0470**
		(0.0221)	(0.0225)
$\mathcal{L}_{ij}$			0.2985***
			(0.0128)
$\mathcal{LL}_{ij}$			0.1841***
			(0.0363)
$\mathcal{P}_{ij}$			-0.3477***
			(0.0163)
$\mathcal{PP}_{ij}$			-0.2737***
			(0.0629)
Standard errors	(plain)	(clustered, pair)	(clustered, pair)
R^2	0.0000	0.0582	0.1320
Adj. R^2	0.0000	0.0581	0.1319
Num. obs.	34364	34364	34364
RMSE	0.8015	0.7778	0.7468

Note: *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; and * indicates 10% level. Standard errors are clustered at the dyadic level. $\mathcal{I}_{ij}$ indicates that one of the two countries in the dyad ij is an Island; $\mathcal{II}_{ij}$ indicates that both countries in the dyad ij are Islands. The same applies to $\mathcal{L}_{ij}$ and $\mathcal{P}_{ij}$. The reference category is $\mathcal{C}_{ij}$, i.e. Continental country. The absence of subscript indicates that the covariate is at the monadic level.

The lowest level of trade costs is for $\mathcal{P}$: if the pair of countries ij is composed of a Continental country and a Partial island country, then $\hat{\tau}_{ij} = 0.8521$ - $0.3477 = 0.5044$; while where ij are two $\mathcal{P}$-countries, $\hat{\tau}_{ij} = 0.2307$. The insular condition is mitigated when Islands are administratively connected with the mainland (which in the majority of cases implies geographical proximity). Countries whose territories include islands seem to be characterized by lower trade costs with their partners, also with respect to the base group of $\mathcal{C}$-countries. All coefficients are significantly different from zero at 1%.

The regression coefficients indicate the level of trade costs for all possible country pairs: one Island and one Continental country, $\hat{\tau}_{ij}$ =1.2579; two islands, $\hat{\tau}_{ij}$ =1.3049; one Landlocked country and one Continental country,

$\hat{\tau}_{ij}$=1.1506; two Landlocked countries, $\hat{\tau}_{ij}$=1.3049. The same coefficients can be used to calculate different combinations of categories in the taxonomy: one Island and one Landlocked country, $\hat{\tau}_{ij}$=0.8521+0.4058+0.2985=1.5564, the worst combination in terms of trade costs; one Partial island and one Landlocked, $\hat{\tau}_{ij}$=0.8029; one Island and one Partial island, $\hat{\tau}_{ij}$=0.9102.

The percentage effect of switching from the reference category to $\mathcal{I}_{ij} = 1$ can be calculated following Kennedy *et al.* (1981) and Giles (2011) in interpreting a dummy variable's coefficient when the dependent variable has been log-transformed. Taking the estimated coefficient for the case where one country in the pair is an Island (Model 3) the effect can be calculated as $100 \times e^{(0.4058-0.5(0.0124^2))} - 1$, and is equivalent to 50.05%. The effect is asymmetric, switching from $\mathcal{I}_{ij} = 1$ to $\mathcal{C}_{ij} = 1$ leads to a decrease in trade costs of 33.36%.[29]

6.1 Trade costs and geography

Our first results need to be checked against further geographical attributes which may interplay with insularity.

$$\tau_{ij} = \iota_{ij}\beta + \mathcal{G}\gamma + \epsilon_{ij} \tag{10}$$

Equation 10 is the same as Equation 9 just with the addition of a set of geographical controls $\mathcal{G}$ which may vary for each country in the pair, either as exporter, i, or as importer, j, or at the dyadic level, ij. Firstly, as suggested by the empirical development literature, we distinguish the effect of insularity from that of size. Small islands developing countries (SIDCs) are a category of singular interest from a policy perspective (Mimura *et al.*, 2007), and the open question is whether and how much insularity adds to the condition of being a small economy (Easterly and Kraay, 2000).

A second geographical dimension which introduces elements of heterogeneity across trade relationships is bilateral distance between country i and country j. Contiguous or nearby countries tend to trade more than countries located far away.

Models in Table 4 introduce size and distance, in logs, as controls (Model 4) while we look at their interaction with insularity in Model 5. Following Ozer-Balli and Sørensen (2013) both size (the land area of country i) and

[29] This depends on the change in the sign of the base category: $100 \times e^{(-0.4058-0.5(0.0124^2))} - 1 = -33.36\%$. See Kennedy *et al.* (1981).

Table 4: Trade Costs, Insularity Index and Geography

	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	OLS	OLS	OLS	OLS	OLS	IV
Intercept	0.9902***	0.9991***	1.0383***	0.9762***	0.8211***	0.9678***
	(0.0105)	(0.0108)	(0.0093) [0.0084]	(0.1055)	(0.1055)	(0.4048)
$\mathcal{I}_{ij}$	0.1435***	0.1227***	0.1359***	0.0892***	0.1221***	0.2131***
	(0.0125)	(0.0129)	(0.0096) [0.0093]	(0.0128)	(0.0127)	(0.0105)
$\mathcal{L}_{ij}$	0.2945***	0.2825***	0.2734***	0.1695***	0.1953***	0.2689***
	(0.0119)	(0.0116)	(0.0084) [0.0088]	(0.0133)	(0.0135)	(0.0118)
$\mathcal{P}_{ij}$	-0.3601***	-0.3630***	-0.3676***	-0.1963***	-0.1871***	-0.1775***
	(0.0151)	(0.0149)	(0.0108) [0.0104]	(0.0144)	(0.0144)	(0.0089)
ln distance$_{ij}$	0.3262***	0.2603***	0.1994***	0.2536***	0.3046***	0.3131***
	(0.0079)	(0.0150)	(0.0119) [0.0107]	(0.0135)	(0.0077)	(0.0097)
ln land area$_i$	-0.0587***	-0.0823***	-0.0843***	-0.1252***	-0.0811***	-0.0727***
	(0.0014)	(0.0033)	(0.0033) [0.0031]	(0.0035)	(0.0018)	(0.0144)
$\mathcal{I}_{ij}\times$ln distance$_{ij}$		0.1256***	0.1683***	0.1088***		
		(0.0180)	(0.0135) [0.0125]	(0.0159)		
$\mathcal{L}_{ij}\times$ln distance$_{ij}$		0.0946***	0.1085***	0.1146***		
		(0.0180)	(0.0126) [0.0118]	(0.0271)		
$\mathcal{P}_{ij}\times$ln distance$_{ij}$		-0.0269***	-0.3676***	0.1146***		
		(0.0212)	(0.0108) [0.0131]	(0.0271)		
$\mathcal{I}_{ij}\times$ln land area$_i$		0.0246***	0.0171***	0.0514***		
		(0.0034)	(0.0037) [0.0035]	(0.0036)		
$\mathcal{L}_{ij}\times$ln land area$_i$		0.0375***	$0.0385***$	0.0706***		
		(0.0031)	(0.0030) [0.0034]	(0.0064)		
$\mathcal{P}_{ij}\times$ln land area$_i$		-0.0131***	$-0.0119***$	0.0706***		
		(0.0043)	(0.0043) [0.0041]	(0.0064)		
Other Insularity categories	YES	YES	YES	YES	YES	YES
Geographical controls[c]	NO	NO	NO	YES	YES	YES
Full interactions[a]	NO	YES	YES	YES	NO	NO
Quadratic terms[b]	NO	NO	YES	NO	NO	NO
Standard errors	(clustered, pair)	(clustered, pair)	(robust) [plain]	(clustered, pair)	(clustered, pair)	(bootstrap)
R^2	0.2521	0.2621	0.2658	0.3915	0.3805	
Adj. R^2	0.2520	0.2616	0.2653	0.3163	0.3163	
Num. obs.	34364	34364	34364	31130	31130	31130
RMSE	0.6932	0.6887	0.6870	0.6259	0.6314	$\varrho = 0.3192$

Note: The dependent variable is τ_{ij}. *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; and * indicates 10% level. Standard errors in parenthesis are clustered at the dyadic level (or robust standard errors in Model 6), the ones in square brackets are plain OLS standard errors. $\mathcal{I}_{ij}$ indicates that one of the two countries in the dyad ij is an Island; $\mathcal{II}_{ij}$ indicates that both countries in the dyad ij are Islands. The same applies to $\mathcal{L}_{ij}$ and $\mathcal{P}_{ij}$. The reference category is $\mathcal{C}_{ij}$, i.e. Continental country. (a) Full interactions means that the regression also includes the following terms: $\mathcal{L}_{ij}\times$ln distance$_{ij}$, $\mathcal{LL}_{ij}\times$ln distance$_{ij}$, $\mathcal{L}_{ij}\times$ln land area$_i$, $\mathcal{LL}_{ij}\times$ln land area$_i$, $\mathcal{P}_{ij}\times$ln distance$_{ij}$, $\mathcal{PP}_{ij}\times$ln distance$_{ij}$, $\mathcal{P}_{ij}\times$ln land area$_i$, $\mathcal{PP}_{ij}\times$ln land area$_i$. (b) Quadratic terms means that the regression also includes the following terms: ln distance$^2_{ij}$, ln land area^2_i. (c) Geographical controls include: ruggedness (Nunn and Puga (2012)), percentage of tropical territory, annual average temperature, average precipitation, distance from nearest ice-free coast (NASA Socioeconomic Data and Applications Center, 2009), distance from the equator (La Porta *et al.*, 1997). The presence of missing observations in the Geographical controls reduces the number of observations (31130) and the number of countries (177) in the dataset.

bilateral distance, being continuous variables, have been demeaned in order to clearly evaluate their interaction with our main category of interest.[30] As expected, distant countries have higher international trade costs while smaller country size is associated with higher costs.

Now the intercept represents the value of $\hat{\tau}_{ij}$ at the mean of the covariates: the trade costs for a $\mathcal{C}$-country pair at the average bilateral distance and at the average land area for country i is $\hat{\tau}_{ij}$=0.99. Controlling for size and bilateral distance shows the following ranking in trade costs: $\hat{\tau}_{\mathcal{L}} > \hat{\tau}_{\mathcal{I}} > \hat{\tau}_{\mathcal{C}} > \hat{\tau}_{\mathcal{P}}$.

Not only controlling for distance and size does not subtract significance from our categories of insularity, showing that trade costs are not simply a matter of bilateral distance, but above and beyond this important remark, results in Model 5 show how the effect of distance is exacerbated for $\mathcal{I}$. The interaction between distance and the Insularity index categories show heterogeneous effects. The cost of distance for island countries is higher than for landlocked ones, and is even smaller for partial islands than for continental ones. A quantification of this difference shows that, at an average level of land area, trade costs increase for $\mathcal{C}$-countries by (a level of) 0.2603, while the increase for $\mathcal{L}$-countries is (0.2603 + 0.0946)=0.3549, a little higher for $\mathcal{I}$-countries, 0.3859, and smaller for $\mathcal{P}$-countries, 0.2334.

The interaction effect of size shows how islands and landlocked countries are penalized more than the other categories, with a marginally higher cost for the latter group of countries.

The second-order expansion in Model 6 allows to control for a possible spurious relation in size and distance, in their interplay with the ι_{ij} variables and τ_{ij}. Including quadratic terms in the regression does not compromise the role of size and distance and the way their effect varies heterogeneously for the different categories of the insularity taxonomy. Performing a GIM test, as in King and Roberts (2015), based on comparison between classical and robust standard errors suggests that this more demanding model is over-specified. The model we should refer to for quantifying our results is therefore Model 4.

Before moving to the consideration of simple direct effects in Model 8 and Model 9, we augment Model 5 with other geographical characteristics at the country level, as reported and compared in Table 2. Estimates of the geography effects (not included in Table 4, but available in the online Ap-

[30] The mean of the logarithmic transformation of bilateral distance is 8.785874, while the mean of the logarithmic transformation of land area of country i is -.3327445.

pendix) are consistent with priors that extreme geography conditions (higher precipitation, shorter distance from the equator, more tropical territory) are in general associated with higher trade costs. All controls for different levels of insularity maintain their significance, τ_{ij} for $\mathcal{I}$-country or a $\mathcal{L}$-country is slightly reduced but still relevant (9% higher than $\mathcal{C}$-countries), as is the reduction associated with $\mathcal{P}$-countries. Other dimensions of 'bad geography' come into play in influencing trade costs, but the insularity dimension remains of relevance.

A similar result is obtained in Model 8 where geographical controls augment the specification of Model 4. In this less demanding specification, with geographic controls $\mathcal{G}$, the estimated coefficient for the $\mathcal{I}$-countries is quite similar to the one of Model 5 including full interactions: geography seems to reduce country heterogeneity more on the side of $\mathcal{L}$-countries and especially $\mathcal{C}$-countries and $\mathcal{P}$-countries.

In order to move toward a causal interpretation of the estimated coefficient we need to be sure that the selection on observed (as in Equation 10) and on unobserved variables - such as trade policy (including non-tariff barriers and measures of gray protectionism such as standards or market access obstacles) and cultural, linguistic or institutional characteristics at the country level, both in i and j - is correctly taken care of (Altonji *et al.*, 2005). Full control of unobservable heterogeneity would require the use of country fixed effects. This strategy is however impracticable in our case since i and j fixed effects would absorb all geographical covariates, and above all would impede the analysis of both spatial and historical connectedness, which is the main goal of this empirical exercise. The alternative of assuming a random intercept for every country imposes too restrictive conditions on the correlation between observables and unobservables.

One possible way out, which we use, is to rely on the properties of the Hausman-Taylor IV estimator (Hausman and Taylor, 1981, Wyhowski, 1994, Egger, 2005, Baltagi *et al.*, 2014) to obtain unbiased estimates. Classical fixed effects estimator assumes that explanatory variables may be correlated with omitted variables, and it eliminates this correlation by the within-transformation. Classical random effects estimator gives consistent estimates of the parameters only assuming that observables are uncorrelated with the omitted variables, i.e. excluding the possibility of any omitted factor. The Hausman-Taylor random effects estimator allows to fit models in which some of the covariates are correlated with the unobserved country-level random effect

The ultimate advantage of the Hausman-Taylor estimator is that it also helps to solve a further concern regarding the potential endogeneity of our insularity taxonomy. National borders are artificial, they are man-made. The fact that one island is an $\mathcal{I}$-country instead of being "*... a part of the maine*," included in a $\mathcal{P}$-country or even in a $\mathcal{C}$-country, is the result of historical political events, that may separate islands which are characterized by less favorable natural conditions than the one considered of greater importance.

The Hausman-Taylor estimator uses internal instruments to solve the potential endogeneity problem. The regression looks like:

$$\tau_{ij} = \iota_{ij}\beta + \mathcal{G}\gamma + u_i + v_j + \epsilon_{ij} \tag{11}$$

where the error component captures countries' heterogeneity along both i and j dimensions, $u_i + v_j$.

Results in last column of Table 4 correspond to the implementation of the Hausman-Taylor IV estimator. The specification includes all geographic controls as in Model 8. When controlling for a possible endogenous process in the insularity taxonomy the trade costs associated with $\mathcal{I}$ and $\mathcal{L}$ are larger. This would point to the need to account for unobserved factors (instrumented by our geographical controls) in order to provide an unbiased measure of the differences between the base category and Islands and Landlocked countries. The omitted variable bias is negative for $\mathcal{I}$-countries and $\mathcal{P}$-countries, increasing the wedge in terms of trade costs between islands not part of a mainland country and islands annexed to a dry land country. Given this, the Hausman-Taylor IV random effect estimator will be used in the following analysis of the role of spatial and historical connectedness.

6.2 Trade costs and connectedness

What makes islands different from other territories? Islands are surrounded by sea, which is the opposite of the crucial characteristic which makes the geography of landlocked countries highly problematic. But sharing a border brings advantages. Proximity favors exchanges through different channels largely created through historically established means of communication. The way we address proximity is directly linked to measures which capture the non-linear effect of distance on exchanges and human relations. In this part of the analysis we try to highlight these arguments by combining our insularity measure with measures of contiguity adjusted for the spatial discontinuity generated by the sea.

The regression takes the following form:

$$\tau_{ij} = \iota_{ij}\beta + \mathcal{G}\gamma + \xi\mathcal{S}_{ij} + u_i + v_j + \epsilon_{ij} \quad (12)$$

where $\mathcal{S}_{ij}$ is an index that, in a nested structure, progressively considers countries according to their level of spatial proximity. $\mathcal{S}_{ij}$ at the first level is equal to 1 if the two countries i and j in the pair share a common land border; at the second level the index adds in the countries whose bilateral distance is below the limit of 300 kms; at the third level the index also includes those countries whose bilateral distance is less than 500 kms. All remaining country pairs are given a value of zero.

All equations in Table 5 include controls for other geographical characteristics, as summarized in Table 2. As in previous regressions, distance is directly associated with trade costs, but does not entirely explain the variability in τ_{ij}. At the same time contiguity is inversely related to trade costs, and the direct effect is magnified by the increase in the number of neighbors considered in each group of contiguous countries. The interaction effect between contiguity $\mathcal{S}_{ij}$ - considered in subsequent steps as "sharing a land border", "being separated by a bilateral distance less than 300km", "being separated by a bilateral distance less than 500km" – and the categories of the insularity taxonomy in ι_{ij} shows an interesting pattern. Spatial connectedness reduces the benefits of the natural geographic position of $\mathcal{P}$-countries with respect to $\mathcal{C}$-countries: playing the role of bridges (recall the discussion of the content of Figure 6) between the continental inland and the outer territories no longer pays off anymore if spatial connectedness is considered as a general feature of all countries.

$\mathcal{L}$-countries immediately gain from spatial connectedness. The higher gains, in terms of a lower τ_{ij}, are obtained for the first step of contiguity considered, that of coming from a direct connection with neighbors with whom a land border is shared. Gains are diluted at higher levels of connectedness. On the contrary, $\mathcal{I}$-countries are unaffected by simple adjacency: the sign of the coefficient is negative, indicating a reduction in trade costs due to spatial connectedness but the coefficient is imprecisely estimated due to the limited number of cases of islands sharing a land border (e.g. Haiti and the Dominican Republic, Ireland and Great Britain, Indonesia and Papua New Guinea). The greatest benefits are obtained at intermediate levels of connectedness (below 300km). Short distance spatial connectedness pays off.[31]

[31] What we are saying is directly related to what the literature on the gravity model

Table 5: Trade costs and spatial connectedness

	Model 10 IV	Model 11 IV	Model 12 IV	Model 13 IV
$\mathcal{I}_{ij}$	0.213*** (0.009)	0.213*** (0.009)	0.214*** (0.009)	0.214*** (0.009)
ln distance$_{ij}$	0.313*** (0.004)	0.290*** (0.005)	0.291*** (0.05)	0.287*** (0.005)
$\mathcal{S}_{ij}$		-0.214*** (0.037)	-0.220*** (0.035)	-0.224*** (0.034)
$\mathcal{I}_{ij} \times \mathcal{S}_{ij}$		-0.168 (0.169)	-0.182** (0.085)	-0.129** (0.065)
$\mathcal{L}_{ij} \times \mathcal{S}_{ij}$		-0.248*** (0.048)	-0.238*** (0.048)	-0.222*** (0.045)
$\mathcal{P}_{ij} \times \mathcal{S}_{ij}$		0.152** (0.068)	0.187*** (0.064)	0.263*** (0.058)
Contiguity	Distance	Land border	Contiguity (<300Km)	Contiguity (<500Km)
Other Insularity categories	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES
Random effects	i,j	i,j	i,j	i,j
Num. obs.	31130	31130	31130	31130
Num. pairs	177	177	177	177
σ_u	0.36	0.36	0.36	0.36
σ_e	0.52	0.52	0.52	0.52
ρ	0.32	0.32	0.32	0.32

Note: The dependent variable is τ_{ij}. *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; and * indicates 10% level. Standard errors are clustered at the dyadic level. $\mathcal{I}_{ij}$ indicates that one of the two countries in the dyad ij is an Island; $\mathcal{II}_{ij}$ indicates that both countries in the dyad ij are Islands. The same applies to $\mathcal{L}_{ij}$ and $\mathcal{P}_{ij}$. The reference category is $\mathcal{C}_{ij}$, i.e. Continental country. Geographical controls (see Table 1) included.

Going back to John Donne, if every country is assumed to be "... *a peece of the Continent*" the associated reward in terms of trade costs is far from homogeneous. And this opens the door to the next (and last) question: how the role of geography can be modified by history, or in other terms how we

calls multilateral trade resistance (Anderson and van Wincoop, 2003). What is of relevance in determining bilateral trade is not only to have or not to have a nearby trade partner (absolute distance) but also the number of such alternatives (relative distance). The more the better and the higher the level of connectedness. This is the case of countries which have access to the sea but also share several borders with other countries which in turn share multiple borders.

move from First nature to Second nature (Krugman, 1991).

7 Elaborations on historical routes data

The idea we want to elaborate upon, in the specific context of islands, is that connectedness is related not only to immutable geographical positions but depends also on a network of contacts (a.k.a. relations) established over time, precisely as a consequence of the physical placement of spatial units in the world map. The information provided by colonial records of shipping routes in and out of Europe allows us to build a measure of the chance, due, in the case of islands, to random historical events, of connecting with other countries. Our elaborations on the CLIWOC 2.1 database (García-Herrera *et al.*, 2005, Konnen and Koek, 2005) allow us to produce two novel pieces of information, which we will exploit in the context of this analysis: (1) the places where ships had long stops, which were normally located in the colonial regions, linked to the mother country by special economic ties; (2) and the locations where ships had short stops during the journey, i.e stopovers normally motivated by conditions and technicalities related to the navigation strategy.

7.1 Long stops

Out of the 5227 total trips included in CLIWOC 2.1, we selected 5191 trips for which we have complete information on the location of voyage origin and destination. Locations in CLIWOC are recorded as found in logbooks. We, then, harmonized the information about the origin and destination, collecting the various names which were given to the same place and, using a simple name disambiguation procedure, singled out the locations and countries of relevance, the ones that we could geo-reference.[32]

We identified 383 locations of origin and 632 destinations. To each one we associated the current country name, the country ISO code and the frequency of recurrence in all 5191 trips. Table 6 lists the locations recorded at least 50 times in our data. It is notable that some countries appear a number of times

[32] For example the island of St. Helena was recorded by different naval officers in the archives as St Helena, St Helena via Table Bay, St Hellena, St. Helena, St Hellens Road (England), St Helens Road UK, Cape of Good Hope has been found as Kaap de Goede Hoop, Kaap de Goede Hoop via O. Indie, Cape of Good Hope, Cape Good Hope.

proportionately to the national composition of the archives, which depended on the number of logbooks available to the CLIWOC team of researchers.[33] Islands are well represented in the list of voyage origins and destinations, Indonesia, Cuba, Barbados, St Eustacius and the Falklands occupying the first places. When moving to the country level frequency records increase since some countries have more than one city or town as origin or destination of historical routes. The correlation, at the country level, between being an origin or a destination is 0.94, which implies that there is a balance between outbound and inbound voyages of European origin.[34]

7.2 Short stops

In order to identify short stops, different variables referring to the duration of the journey have been used. In fact trips were quite lengthy, depending first of all on the distance covered but also on the events happening during the journey. The number of days varies from 1 to 412. It is during navigation that islands had a special role, since they were among the most likely territories to be encountered by ships navigating the sea. As mentioned, CLIWOC presents the daily position of the vessel in terms of latitude and longitude.[35] We use 272,674 entries to process the information, which amounts to the total number of navigation days in CLIWOC 2.1 after cleaning the data for duplicates. Daily details are given by the CLIWOC research team on the procedure used to establish the daily position of the vessel: dead reckoning, from true navigation, interpolated manually or, what we focused on, inserted in the original database as the *actual position of ports or islands* (Russell

[33] CLIWOC is a dataset on navigation which does not concentrate on commercial journeys but includes information for all types of vessels including military ones. Data have been collected with the aim of producing climate information for each day in each ocean basin, except for the Pacific where the lack of daily data is reported. The fact that it is neither representative nor concentrated on journeys with a commercial purpose casts some doubts on the possibility of using the information provided by frequencies while instead vessels' positions provide a solid information base on the routes used during the age of sailing.

[34] Frequency data reveal a robust concentration both in terms of cities of departure and arrival and origin and destination countries. We found 134 (314) localities which appear as routes' origin (destination) only once, also when naming harmonization was completed. Only 83 (88) places appear more than 10 times as voyage origins (destinations).

[35] As reported in (Russell and Cohn, 2012), chapters 1 and 3, the CLIWOC team group harmonized the reference meridian among the several different logbooks. This check implied a strict verification of the vessels position along the route.

Table 6: Most frequent Origins and Destinations of Historical Trade Routes

Origin	Freq. in Dataset	Country ISO code	Country name	Destination	Freq. in Dataset	Country ISO code	Country name
La Coruña	738	724	Spain	Batavia	344	360	Indonesia
Coastal East India	489	699	India	La Coruña	230	724	Spain
Batavia	447	360	Indonesia	Spithead	202	826	UK
La Habana	254	192	Cuba	La Habana	162	192	Cuba
Cadiz	213	724	Spain	Montevideo	128	858	Uruguay
Hellevoetsluis	207	528	Netherlands	Cadiz	120	724	Spain
Falkland Islands	201	238	UK	Curacao	119	530	Brasil
Texel	190	528	Netherlands	Suriname	102	740	Suriname
Nederland	187	528	Netherlands	St Helena	97	826	UK
Barcelona	169	724	Spain	Nederland	93	528	Netherlands
Curacao	158	530	Brasil	Barbados	88	28	Barbados
Rochefort	100	251	France	Madras	87	699	India
Montevideo	98	858	Uruguay	Nieuwediep	83	528	Netherlands
Rotterdam	97	528	Netherlands	Hellevoetsluis	72	528	Netherlands
Galle	92	144	SriLanka	Downs	71	826	UK
Nieuwediep	87	528	Netherlands	Rotterdam	69	528	Netherlands
Amsterdam	83	528	Netherlands	Plymouth	68	826	UK
Iceland	80	352	Iceland	UK	68	826	UK
Suriname	77	740	Suriname	Paramaribo	64	740	Suriname
St Eustacius	72	530	Antilles (NL)	Table Bay	64	711	South Africa
Acapulco	71	484	Mexico	Bombay	64	699	India
Middelburg	70	528	Netherlands	St Eustacius	61	530	Antilles (NL)
Vlissingen	61	528	Netherlands	Halifax	61	124	Canada
Brest	57	251	France	Texel	59	528	Netherlands
Paramaribo	52	740	Suriname	Madeira	58	620	Portugal
				Kaap de Goede Hoop	56	711	South Africa

Note: Our elaborations from CLIWOC 2.1.

and Cohn, 2012) where vessels stopped during their voyage. We merged this information with other coding available in the original database: the detail on anchoring, with a specific note reporting when the *ship is at anchor or moored* (Russell and Cohn, 2012). We concentrate on where anchorage took place and by using these bits of information on the vessels' daily position from all trips we are able to identify a list of islands, which can be defined as the 'treated group', which includes countries where vessels used to either start their journey (long stops in the section above) or to stop during navigation. This group includes all countries in the first column in Table 7. Within this group we isolate those countries which share two common features: ships used to *anchor* inside country's boundaries and the vessel position was the *actual position of ports or islands*, i.e. the position was either that of a known port or an island recorded in navigation maps used at the time. These are the countries denoted by the subscript *a* in Table 7. The implication being that contact between locals and foreigners was highly probable in these cases, as the crew could disembark and meet with the local population to restock for the onward long voyages. This group includes 40 countries for which we have data on 1995-2010 trade costs; 18 out of 40 are islands. Some of them belong to the developed world (Australia, UK, Ireland, Japan) and on some journeys they appear as origins or final destinations of our voyages.

Table 7: Islands: Treated and Control Groups

Islands in Historical Navigation routes	Islands not involved
Antigua and Barbuda[a]	Aruba
Australia[a]	Bahrain
Bahamas[a]	Bermuda
Barbados[a]	Brunei
Cape Verde[a]	Cayman Islands
Comoros[a]	Cyprus
Cuba[a]	Dominican Republic
Dominica	Greenland
French Polinesia	Grenada
Great Britain[a]	East Timor
Haiti[a]	Fuji
Honduras	Iceland
Ireland[a]	Jamaica
Indonesia[a]	Kiribati
Japan[a]	Maldives
Madagascar[a]	Malta
Mauritius[a]	Marshall Islands
Micronesia	New Zealand
New Caledonia	Palau
Papua New Guinea	Samoa
Philippines[a]	Seychelles
Sri Lanka[a]	Singapore
Sao Tome and Prince[a]	Solomon Islands
St Kitts and Navis	St Lucia
Tonga	St Vincent and the Grenadines
Trinidad and Tobago[a]	Turks and Caicos Islands
	Tuvalu
	Vanuatu

Note: Our elaborations from CLIWOC2.1; [a]: CLIWOC variable LatInd and LonInd (Position Indicator, i.e. the origin of the given decimal latitude or longitude) equal to 5 (CLIWOC variable definition: 'is from atlas'); variable Anchored equal to 1.

7.3 Further remarks on long and short stops

Some considerations on the process behind the participation in historical routes, which we interpret to be different when looking at long and short stops, are worth noting. Reasoning on how the country involvement in routes has been created revolves, in fact, around the endogeneity issues related to the novel information extracted from the CLIWOC database.

The main purpose of historical voyages was to reach specific destinations whose main characteristics were favorable for providing Europe with new, high-value products. Origins (destinations) of the voyages, when in Europe, determined the nationality of the ship, while the location outside the old world identifies the first main trade-colonies outside Europe. Following this reasoning the association with lower trade costs from being an origin (destination) of routes has to be evaluated considering the endogenous creation of colonial linkages/trade routes.

A different reasoning has to be made when referring to locations where vessels could stop on the journey. Stop-overs were necessary because of the long distance between origin and destination. Having the possibility of stopping en route limited the high level of uncertainty which unforeseen events imposed on journeys, such as unfavorable weather conditions or the need to replenish any supplies needed for navigation. Possible stop-overs were known to captains since they were detailed in the maps used at the time. Islands were the most likely territories to be encountered during navigation but some of them were more likely than others, thanks to certain geographical features. In a number of cases vessels stopped and anchored in an open sea geo-location near the territory encountered during navigation (what we will call 'anchoring' in next section), in some others, when at anchor, the vessel position was taken as the *actual position of ports or islands* (what we will call "anchoring in port').

Going back to the navigation of the Seaford depicted in Figure 2, we can distinguish locations we call long stops from those we refer to as short stops. Ports with a red dot identify the former, while all locations in yellow identify locations where the ship, according to definitions in our original dataset, is *at anchor or moored* and the position of the vessel is taken as the *actual position of ports or islands*.[36]

Said differently, the peculiar geographical status of islands potentially gave them a special role (as a probable stop-over location) in the history of ocean navigation where the probability of this role being played, initially, depended largely on the interplay of wind patterns and land configuration offering vessels easy harbor. This point makes islands an interesting natural experiment. After controlling for the few cases of islands which were also origin of routes, our identification strategy rests on studying how participation in historical journeys had a differential impact for islands, because of the

[36] It is worth noting that our definition of long and short stops is not related to the number of days the ship stayed but only to the need to distinguish territories which were pertinent to an established colonial tie (origins) from those where stops were determined by factors related to the uncertainty and needs of the navigation.

element of causality involved in their inclusion in routes. Also, any further economic process linked to historical trade routes facilitating the access to the island, and therefore reducing trade costs, is related to the spatial distribution of trade costs two centuries ago, which is no longer in place owing to the various breakthroughs and advances in maritime navigation technology. Being an internal point in connecting the links of an historical trade route is determined by factors unrelated to current day navigation and trade costs.

8 Historical routes results

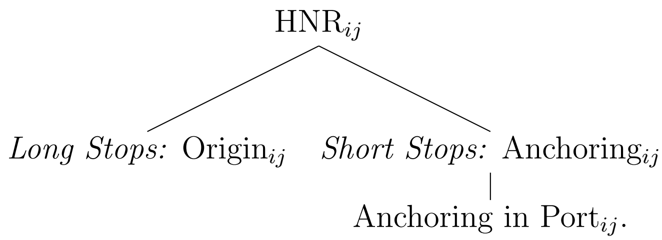
We will focus now on how trade costs vary with participating to historical routes and countries' geographical status. The expectation being that historically established connectedness are associated with lower trade costs in the most general case, i.e. for all countries with sea access. Any differential effect on islands allows a clearer interpretation of the causality link, since participation in trade routes, for islands, has been determined mainly by geographical characteristics and by factors external to navigation and trade costs today.

The estimated Equation 13 is now augmented with different measures of the involvement of countries in historical navigation routes.

$$\begin{aligned}\tau_{ij} = \iota_{ij}\beta + \delta_1\mathcal{I}_{ij} + \delta_2\text{HNR}_{ij} + \delta_3\mathcal{I}_{ij} \times \text{HNR}_{ij} + \delta_4\mathcal{II}_{ij} \times \text{HNR}_{ij} + \\ + \mathcal{G}\gamma + \xi\mathcal{S}_{ij} + u_i + v_j + \epsilon_{ij}\end{aligned} \tag{13}$$

where HNR identifies the involvement in historical navigation routes as a whole, which we also consider in the interaction with the condition of being an island, so to evaluate any differential effect that historical connectedness may have for the peculiar geographical state of being in the middle of an ocean. The base group is now $\mathcal{C}$-countries not involved in historical navigation, and our insularity index separates out $\mathcal{I}$, $\mathcal{L}$ and $\mathcal{P}$-countries. All estimated models include our spatial contiguity measures, $\mathcal{S}_{ij}$, our set of geographical controls, all included in matrix $\mathcal{G}$, and we run a regression model using the Hausman-Taylor estimator in order to control for the unobserved heterogeneity across different geographies.

The general measure, HNR, takes different specifications according to how the 'treatment' is conceived. Operationally, we constructed a set of categorical variables following the tree structure:



In the more general treatment, HNR isolates countries either as route origin, or where ships used to anchor (or both). In further specifications we consider the more specific treatment of anchoring, i.e. countries where ships used to stay *at anchor or moored*. Then we move into more detail, identifying where ships used to stay *at anchor or moored* and taking the position of the vessel as the *actual position of ports or islands*. All the variables are defined at the dyadic level, since the category takes value 1 when one country in the pair satisfies the condition of being either an origin-country or where ships used to stop and anchor, or, where vessels used to anchor in a location known in the maps. The categorical variable(s) taking value 1 may include all countries, except for landlocked states, while within all categories our interaction terms isolate island states.[37]

Results are presented in Table 8, where first we include our HNR variable (Model 14), then we specifically distinguish territories which were stop overs and where ships used to anchor (outside or inside the port) (Model 15) and then we isolate those where ships anchored inside the port (Model 16). In Model 17 we include controls also for territories where routes originated, in order to single out from the treated group those countries that were the main actors in the building process of navigation routes and colonial linkages.

All first four specifications suggest that participation in historical routes is associated with lower trade costs today. The effect is significant and remarkable; countries where routes originated or where vessels stopped during the long navigation register current trade costs 29% lower (Model 14). Islands have higher trade costs (in these specifications $\mathcal{I}$ isolates, from all 54 islands in the sample, those never involved in historical routes: they have trade costs between 30% (Model 14) and 39% (Model 16) higher than conti-

[37] Regarding islands, the more general measure follows the dichotomy between treated and non-treated presented in Table 7. Anchoring was reported in countries indicated in the first column of the same table, while only in countries *a* ships anchored inside the port.

nental countries.[38] All coefficients are significantly different from zero at 1%. There is a very small differential effect for islands, increasing slightly when trade occurs between two islands both involved in historical routes. Both coefficients are significantly different from zero at 10%. When we move to qualify the type of involvement in historical trade, in Models 15 and 16, we isolate territories where ships used to stop and anchor. After isolating anchoring, the association of our historical measures with a reduction in trade costs increases, for all types of countries. Now the differential effect associated with $\mathcal{I}$ is substantial (islands record a further 10% reduction in trade costs) and significant at 1%. When we move further, isolating the specific case of anchoring in a port or place known in the maps, the trade costs for the host countries are significantly lower (35%). The fact that anchoring took place inside a port is now more important for islands with respect to other territories. In Model 17 we add the specific control for being country of route origin in order to separate out those territories which contributed to the creation process of the navigation route. The sign and size of the effect associated with islands where ships used to stop and anchor in ports is unchanged while, as expected, being a route origin is associated with lower trade costs, although islands do not differ significantly from other types of territories. Islands which welcomed vessels in their ports exhibit a further 15% reduction in average trade costs. The effect increases when the pair consists of two islands participating in historical routes.

All specifications in Table 8 control for the fact that being accessible by historical routes was mainly due to their favorable geographic position, the fact of being nearer to the mainland, for example on routes connecting Europe to Asia, where the navigation path followed long stretches of the coasts. Coefficients associated with islands in historical routes are cleaned from the spatial connectedness effect ($\mathcal{S}_{ij}$), introduced in Equation 12), shown in the previous section to be associated with lower trade costs.

8.1 Robustness

Feyrer and Sacerdote (2009) examine the legacy of colonial history on modern income. With a random variation in the colonial experience, due to the random nature of discovery and colonization of the 80 islands of their

[38] As before, results (not included in the table) indicate that landlocked countries are associated with higher trade costs while partially insular countries exhibit lower ones.

Table 8: Trade costs and Historical Connectedness

	Model 14	Model 15	Model 16	Model 17	Model 18	Model 19	Model 20
	IV	IV	IV	IV	IV	IV	IV
$\mathcal{I}_{ij}$	0.264***	0.271***	0.334***	0.335***	0.329***	0.073**	-0.242***
	(0.017)	(0.011)	(0.010)	(0.014)	(0.014)	(0.024)	(0.025)
Historical Navigation Routes	-0.254***						
	(0.011)						
$\mathcal{I}_{ij}$ × Historical Navigation Routes	-0.035*						
	(0.018)						
$\mathcal{II}_{ij}$ × Historical Navigation Routes	-0.130*						
	(0.048)						
Anchoring		-0.198***					
		(0.009)					
$\mathcal{I}_{ij}$ ×Anchoring		-0.103***					
		(0.014)					
$\mathcal{II}_{ij}$ ×Anchoring		-0.228***					
		(0.032)					
Anchoring in Port			-0.299***	-0.220***	-0.161***	-0.165***	-0.175***
			(0.01)	(0.010)	(0.009)	(0.009)	(0.009)
$\mathcal{I}_{ij}$ × Anchoring in Port			-0.144***	-0.153***	-0.153***	-0.058***	-0.121***
			(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
$\mathcal{II}_{ij}$ × Anchoring in Port			-0.293***	-0.277***	-0.223***	-0.129***	-0.197***
			(0.033)	(0.032)	(0.030)	(0.035)	(0.030)
Origin of Route				-0.285***	-0.175***	-0.169***	-0.159***
				(0.010)	(0.009)	(0.009)	(0.009)
$\mathcal{I}_{ij}$ × Origin of Route				-0.010	-0.018	-0.017	0.084***
				(0.015)	(0.014)	(0.013)	(0.013)
Other Insularity categories	YES	YES	YES	YES	YES	YES	YES
Geographical Controls	YES	YES	YES	YES	YES	YES	YES
Spatial Connectedness Controls	YES	YES	YES	YES	YES	YES	YES
Main Colonizer Dummy	NO	NO	NO	NO	YES	YES	YES
Duration of Colonization	NO	NO	NO	NO	NO	YES	YES
Ocean Dummy	NO	NO	NO	NO	NO	NO	YES
Random effects	i,j	i,j	i,j	i,j	i,j	i,j	i,j
Intercept	1.256***	1.353***	1.405***	1.561***	0.928**	0.924**	1.237**
	(0.354)	(0.363)	(0.355)	(0.338)	(0.319)	(0.312)	(0.299)
Observations	31130	31130	31130	31130	31130	31130	31130
Number of groups	177	177	177	177	177	177	177
ς_u	0.342	0.351	0.344	0.327	0.283	0.276	0.265
ς_e	0.514	0.514	0.505	0.496	0.462	0.458	0.447
ϱ	0.307	0.319	0.317	0.303	0.272	0.267	0.261

Note: The dependent variable is τ_{ij}. *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; and * indicates 10% level. Standard errors are clustered at the dyadic level. $\mathcal{I}_{ij}$ indicates that one of the two countries in the dyad ij is an Island; $\mathcal{II}_{ij}$ indicates that both countries in the dyad ij are Islands. The same applies to $\mathcal{L}_{ij}$ and $\mathcal{P}_{ij}$. The reference category is $\mathcal{C}_{ij}$, i.e. Continental country. Geographical controls (see Table 1) included.

sample, they show that being colonized before 1700 - that is, before enlightenment values began to influence policy towards the indigenous inhabitants of colonies - has left no residual effect on modern income per capita. On the contrary, the time spent as a colony between the XVIII and the XX century is beneficial to modern income. Our measure of involvement in historical routes is clearly tangled with the colonial history of our units of analysis (Acemoglu *et al.*, 2001), in particularly with the colonial ties for the dyads in our sample. We therefore check against the possibility that measuring the involvement in historical routes is simply capturing differences in the colonization histories of the different countries. We tackled the issue in two different ways, and results are presented in the last two regressions reported in Table 8. In Model 18 we add to the specification of Model 17 a set of colonial ties dummies, referred to several nation-state empires.[39] Results on our variables of interest do not change: insularity is associated with higher trade costs; hosting ships in port is associated with lower trade costs for countries regardless of their geography. The size of the coefficient is now smaller, but the differential effect associated with islands holds. Also being the origin of a trade route is still associated with lower trade costs. The implication is that, as in Acemoglu *et al.* (2001), colonial links capture some of the heterogeneity associated with geography, but history holds a further role. Average trade costs for those islands which participated in historical routes are only 5% higher than to the reference category; while islands not treated by history now register trade costs 39% higher.[40]

Model 19 of Table 8 includes controls for the length of colonial history as in Feyrer and Sacerdote (2009) for all islands in our sample, as visualized in Figure 1.[41] Our coefficient $\mathcal{I}_{ij}$ is now associated with a different group of islands, the group of colonizers. Trade links of colonizers are characterized, on average, by higher trade costs nowadays, but the increase with respect to the $\mathcal{C}$-country group is only 7%. Including controls on the length of colonial history is very important when testing the robustness of our historical measures, since the involvement in historical routes is strongly linked to the age of discovery. Adding controls for the length of colonial history is likely to interact with the differential effect of participating in routes. Results give

[39] Historical records allow to select the following list of colonizers: Australian, Austrian, Belgian, German, Danish, Spanish, French, English, Italian, Japanese, Dutch, New Zealander, Portuguese, Russian, Turkish, American and Yugoslavian.

[40] For the treated group the formula for calculating the total effect is: $100 \times e^{(0.329-0.5(0.014^2))-(0.161-0.5(0.009^2))-(0.153-0.5(0.014^2))} - 1$.

[41] We differentiated islands according to the century when their colonial history began: 1500-1600-1700-1800-1900- countries which are still under a colonial legacy.

some evidence of this. Countries where vessels used to anchor in port exhibit lower trade costs; the differential effect associated with islands is still there, though to a lesser extent but still significant at 1%.

In the last specification (Model 20) we include a control for separating out islands according to their ocean location. Now $\mathcal{I}$ isolates Pacific-ocean islands which exhibit higher trade costs than all the remaining island categories. Participation in historical routes is still associated with lower trade costs, for all types of countries. Also the differential effect associated with $\mathcal{I}$ is significantly in place.

8.2 Transmission mechanisms and the *culture of openness*

Three issues still warrant discussion. The first concerns the mechanism that links the historical randomness of being part of a navigation route in the XVIII century and low trade costs today. As shown by Pascali (2017), after 1850 navigation technology changed radically. The mastering of longitude due to the adoption of Harrison's chronometer or its reinventions modified the way navigation routes were planned. In 1819 the SS Savannah hybrid sailing ship crossed the Atlantic Ocean, from the coast of Georgia, US, to Liverpool, England. In fifty years steamships became the norm in transatlantic navigation, as wooden vessels were substituted by metal steamers. Sea lanes changed accordingly. They increasingly approximated straight lines across the oceans and were discussed and codified in several international conferences between 1866 and 1891 (Collinder, 1955). The point is, navigation routes in the XVIII and XIX century were different from those that prevail nowadays. The islands that were involved in historical navigations routes are not used anymore as stopping places by modern freighters, which require specific port and logistic technologies.

The evidence that our empirical analysis is asserting is neither related to the persistence of the positive effect of transportation routes (Dell, 2010, Banerjee *et al.*, 2012, Faber, 2014), nor to the investment in infrastructure built in the past, is still giving positive returns after two-hundred and fifty years (Donaldson, 2016, Donaldson and Hornbeck, 2016). Our explanation of why islands that were touched by HNR have lower τ_{ij} than similar but untreated islands, is based on different and independent pieces of evidence. One is the substantial absence of correlation between the HNR dummy and variables that could be related to positive economic outcomes: the quality of institutions, the level of democracy, the level of infant mortality, the participation of women in the labor force, the share of the population having a secondary and tertiary education, the specificity of product specialization and several other variables that we considered as possible clues for the detection of one unique transmission mechanism from HNR to low τ_{ij} failed miserably. We could not find strong evidence in favor of a single and unique explanation. The conclusion that we are inclined to consider is perfectly in line with the previously shown negative evidence: there is not a single and unique transmission mechanism between HNR and a low τ_{ij}. The islands involved in HNR took different economic trajectories during the second half of the XX century. Some of them specialized in tourism, others developed a modern financial sector, still others heavily rely on international economic support and the remittances of a large migration diaspora.

Some of the variables listed before nicely fit a subset of HNR islands, but not all of them. On the other hand, HNR islands share a common feature. In spite of being so heterogenous along many dimensions, treated islands distinguish themselves by the large number of international trade partners. Substantially larger than what similar but untreated islands have. HNR islands seems to be characterized by what we call a *culture of openness*: the propensity to exploit the opportunities offered by international relations. The last piece of evidence lies in history. In spite of the very different economic trajectories taken by the different HNR islands, the culture of openness that originated from being part of an international navigation system, by interacting with foreigners and - during the long stops of vessels anchored in the island's port – by exchanging goods, experiences, ideas and traditions with those foreigners, gave those islands an advantageous position that each one exploited in a different way (e.g. through the development of a tourism industry or a financial sector) when the third wave of globalization increased the opportunities to be captured having a well established *culture of openness*.

The second issue that is worth discussing is more technical and is related to the possibility offered by using the HNR dummy as an instrument for an endogenous variable correlated with τ_{ij}, e.g. modern infrastructure. Two elements discourage this possible empirical strategy. The first is that the HNR dummy directly affects the potential outcome τ_{ij}; the second is even more important: given the strong heterogeneity of the HNR islands and the evidence of absence of a unique transmission mechanism, attributing the level of τ_{ij} to a single specific variable (even if instrumented by the HNR dummy) would be unreasonable.

The navigation routes and the role of islands as stopping places changed substantially in the XIX century, reducing the scope of physical infrastructure. What remained after those changes is necessarily less tangible, something admittedly fuzzy as the notion of the *culture of openness*, that was of value in the XVIII century and in the XX century as well, and not directly associable with a single channel of transmission (or variable to be instrumented) linking the past to the present.

The third issue regards the possible interpretations of the results. The effect of the past on future outcomes warrants further discussion. Policies that are targeted to open up countries to global markets and to reduce international trade costs often disregard the conditioning effect of the past. In this paper we demonstrate how islands involved in HNR are characterized by lower trade costs today. From a policy perspective, this evidence highlights that islands that did not develop a culture of openness in the past will require extra political effort in the present in order to recover from 'bad geography'.

9 Concluding remarks

This paper provides new evidence on how spatial and, especially, historical connectedness can influence, modify and overturn natural conditions. In a world of 191 economies we focused on islands versus other types of geographical entities because of their unique group-characteristic of entering modern history through a fluke, a random discovery process (Feyrer and Sacerdote, 2009, Diamond and Robinson, 2010). After deriving a generalized measure of international trade costs (Novy, 2013, Jacks *et al.*, 2008, Chen and Novy, 2011), we investigated the spatial distribution of bilateral trade costs across the different categories of a novel ordered Index of Insularity. Our inferential evidence reveals that on average Islands (and Landlocked countries) currently have higher costs, while Partial islands show lower costs with respect to the base category, Continental countries. Results are robust to the inclusion of several geographical controls, including bilateral distance and the size of the countries, and after taking into account unobservable heterogeneity through the use of a Hausman-Taylor IV estimator.

The inclusion in the regression of several measures of spatial connectedness shows that international trade costs are substantially reduced, and results are particularly strong for islands, indicating that a substantial part of the cost of being an island is its position with respect to its potential trade partners. What is of relevance in determining bilateral trade is not only to have or not to have a nearby trade partner (absolute distance) but also to the number of such alternatives (relative distance). The more the better.

The evaluation of history's role in shaping geography through the ability of a country to be connected with other territories, or with the rest "... *of the maine*", using John Donne's words, is accomplished re-elaborating the information contained in the Climatological World's Oceans 1662-1855 database (García-Herrera *et al.*, 2005, Konnen and Koek, 2005), on British, Dutch, French and Spanish navigation from and to Europe and the New World. The database covers the 'age of discoveries by sailing,' when navigation was marked by the inability to precisely establish the longitudinal position of vessels.

Islands were discovered by chance, and we exploit the randomness of the discovery process in our identification strategy in order to properly isolate the effect of history on geography and, subsequently, on economic outcomes. Our data provide two main pieces of information: they identify places where ships had long stops, which were normally located in the colonial regions, linked to the mother-country by special economic ties. Secondly they identify locations where ships had short stops, i.e stopovers normally motivated by conditions and technicalities related to the navigation strategy which, given the type of technology used were largely associated with favorable winds, currents and other geographical characteristics, such as land configuration offering vessels easy harbor. By using the information on the daily position of the vessels, we are able to distinguish route origin countries from those territories where vessels used to anchor during navigation, where also the latter information comes from logbook notes. By merging information on the way the position of the vessel was taken, particularly when the position was recorded as 'actual position of ports or islands' we also isolate when ships anchored in a port.

Our empirical exercise gives a novel account on how participation in historical navigation routes, driven initially by geographic conditions, is associated with lower international trade costs between 1995 and 2010, particularly for those territories where ships anchored during navigation. The reduction is stronger for those countries, particularly islands, where the vessels' position was recorded in a port or island known in maps used at the time. Average trade costs for those islands which participated in historical routes are only 5% higher than countries having access to the sea but not islands; while islands not treated by history register trade costs 39% higher. The interesting part of the results lies in the fact that this occurs regardless of whether those territories are not included in today's global navigation routes, since their renovation by subsequent changes both in the technology of navigation (motor engines) and after the strategic opening of canals (Suez and Panama, to indicate the most important) which shortened the distance between continents.

The process of building trade relations had a role in creating something which is likely to work also after centuries: a culture to openness, an aspect which is fundamental if a country aims to better position its economy in the international market. Our results would suggest that this dimension, the formation and evolution of a culture of openness as a fundamental ingredient of countries connectedness, warrants further research.

References

Acemoglu D., Johnson S. and Robinson J.A. (2001) The Colonial Origins of Comparative Development: An Empirical Investigation, *The American Economic Review*, 91, 5, 1369–1401.

Acemoglu D., Johnson S. and Robinson J.A. (2002) Reversal of fortune: Geography and institutions in the making of the modern world income distribution, *Quarterly Journal of Economics*, 1231–1294.

Almaty Program of Action (2003) Report of the International Ministerial Conference of Landlocked and Transit Developing Countries and Donor Countries and International Financial and Development Institutions on Transit Transport Cooperation, Almaty, Kazakhstan, 28 and 29 August 2003 (A/CONF.202/3, annex I), [New Vienna online info; accessed 25-July-2016].

Altonji J.G., Elder T.E. and Taber C.R. (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy*, 113, 1, 151–184.

Anderson J.E. and van Wincoop E. (2003) Gravity with Gravitas: A Solution to the Border Puzzle, *American Economic Review*, 93, 1, 170–192.

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

Anderson J.E. and Yotov Y. (2012) Gold standard gravity, *NBER Working Paper*, w17835.

Anderson J.E. and Yotov Y.V. (2010) The changing incidence of geography, *American Economic Review*, 100, 5, 2157–86.

Armington P.S. (1969) A Theory of Demand for Products Distinguished by Place of Production, *Staff Papers-International Monetary Fund*, 159–178.

Arvis J.F., Duval Y., Shepherd B. and Utoktham C. (2013) Trade Costs in the Developing World: 1995-2010, *World Bank Policy Research Working Paper*, 6309.

Ashraf Q. and Galor O. (2013) The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development, *The American Economic Review*, 103, 1, 1–46.

Baltagi B.H., Egger P. and Pfaffermayr M. (2014) Panel Data Gravity Models of International Trade, in: *The Oxford Handbook of Panel Data*, Baltagi B.H., ed., Oxford University Press, chapter 20.

Banerjee A., Duflo E. and Qian N. (2012) On the road: Access to transportation infrastructure and economic growth in China, Technical Report 17897, National Bureau of Economic Research.

Becker C. (2012) Small Island States in the Pacific: the Tyranny of Distance, *IMF Working Paper*, 223, 12.

Bertho F., Borchert I. and Mattoo A. (2016) The trade-reducing effects of restrictions on liner shipping, *Journal of Comparative Economics*, 44, 2, 231–242.

Briguglio L. (1995) Small island developing states and their economic vulnerabilities, *World Development*, 23, 9, 1615–1632.

Briguglio L. and Kaminarides J. (1993) *Islands and small states: Issues and policies*, Pergamon Press.

Chaney T. (2008) Distorted Gravity: The Intensive and Extensive Margins of International Trade, *American Economic Review*, 98, 4, 1707–1721.

Chen N. and Novy D. (2011) Gravity, trade integration, and heterogeneity across industries, *Journal of International Economics*, 85, 2, 206–221.

Chen N. and Novy D. (2012) On the measurement of trade costs: direct vs. indirect approaches to quantifying standards and technical regulations, *World Trade Review*, 11, 3, 401–414.

Collinder P.A. (1955) *A History of Marine Navigation*, St. Martin's Press.

Conference on Landlocked Developing Countries (2014) Report of the Second United Nations Conference on Landlocked Developing Countries), [New Vienna online info; accessed 25-July-2016].

De Benedictis L. and Taglioni D. (2011) The Gravity Model of International Trade, in: *The Trade Impact of European Union Preferential policies: An Analysisthrough Gravity Models*, De Benedictis L. and Salvatici L., eds., Springer, chapter 4.

Dell M. (2010) The Persistent Effects of Peru's Mining Mita, *Econometrica*, 78, 6, 1863–1903.

Diamond J. (1999) *Guns, Germs, and Steel: The Fates of Human Societies*, WW Norton & Company.

Diamond J. and Robinson J.A. (2010) *Natural Experiments of History*, Harvard University Press.

Donaldson D. (2016) Railroads of the Raj: Estimating the Impact of Transportation Infrastructure, *American Economic Review*, forthcoming.

Donaldson D. and Hornbeck R. (2016) Railroads and American economic growth: A "market access" approach, *The Quarterly Journal of Economics*, 131, 2, 799–858.

Easterly W. and Kraay A. (2000) Small states, small problems? Income, growth, and volatility in small states, *World Development*, 28, 11, 2013–2027.

Eaton J. and Kortum S. (2002) Technology, Geography, and Trade, *Econometrica*, 70, 5, 1741–1779.

Egger P. (2005) Alternative Techniques for Estimation of Cross-Section Gravity Models, *Review of International Economics*, 13, 5, 881–891.

ESCAP and World Bank (2013) *International Trade Costs*, World Bank, Washington.

EU Posei Program (2006) Programme d'Options Spècifiques à l'Èloignement et l'Insularitè, [New EU online info; accessed 25-July-2016].

Faber B. (2014) Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System, *The Review of Economic Studies*, 81, 1, 1046–1070.

Fajgelbaum P. and Redding S.J. (2014) External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914, Technical Report 20217, National Bureau of Economic Research.

Fally T. (2015) Structural gravity and fixed effects, *Journal of International Economics*, 97, 1, 76–85.

Faye M.L., McArthur J.W., Sachs J.D. and Snow T. (2004) The challenges facing landlocked developing countries, *Journal of Human Development*, 5, 1, 31–68.

Feenstra R.C. (2003) *Advanced International Trade: Theory and Evidence*, Princeton: Princeton University Press.

Feyrer J. and Sacerdote B. (2009) Colonialism and modern income: islands as natural experiments, *The Review of Economics and Statistics*, 91, 2, 245–262.

Gallup J.L., Sachs J.D. and Mellinger A.D. (1999) Geography and economic development, *International Regional Science Review*, 22, 2, 179–232.

García-Herrera R., Können G.P., Wheeler D., Prieto M.R., Jones P.D. and Koek F.B. (2005) CLIWOC: A climatological database for the world's oceans 1750–1854, *Climatic Change*, 73, 1-2, 1–12.

Gaulier G. and Zignago S. (2010) Baci: International trade database at the product-level, *CEPII Working Paper*, 23.

Giles D.E. (2011) Interpreting dummy variables in semi-logarithmic regression models: Exact distributional results, *University of Victoria, Department of Economics, Econometrics Working Paper*, EWP1101.

Hausman J.A. and Taylor W.E. (1981) Panel data and unobservable individual effects, *Econometrica*, 1377–1398.

Head K. and Mayer T. (2014) Gravity Equations: Workhorse, Toolkit, and Cookbook, in: *Handbook of International Economics*, Gopinath G., Helpman E. and Rogoff K., eds., Elsevier Science, Handbooks in Economics.

Helpman E. (2009) *The Mystery of Economic Growth*, Harvard University Press.

Helpman E., Melitz M. and Rubinstein Y. (2008) Estimating Trade Flows: Trading Partners and Trading Volumes, *The Quarterly Journal of Economics*, 123, 2, 441–487.

Hibbs J.D.A. and Olsson O. (2004) Geography, Biogeography, and Why Some Countries are Rich and Others are Poor, *Proceedings of the National Academy of Science*, 101, 10, 3715–3720.

Hoekman B. and Nicita A. (2008) Trade Policy, Trade Costs, and Developing Country Trade, *World Bank Policy Research Working Paper Series*, 4797.

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

Imbens G.W. and Rubin D.B. (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Jacks D.S., Meissner C.M. and Novy D. (2008) Trade Costs, 1870-2000, *The American Economic Review*, 529–534.

Jedwab R., Kerby E. and Moradi A. (2014) History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya, Technical Report 4, Centre for the Study of African Economies, University of Oxford.

Kennedy P.E. *et al.* (1981) Estimation with correctly interpreted dummy variables in semilogarithmic equations, *American Economic Review*, 71, 4.

King G. and Roberts M.E. (2015) How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It, *Political Analysis*, 23, 159–179.

Konnen G. and Koek F. (2005) Description of the CLIWOC database, *Climatic Change*, 73, 117–130.

Krugman P.R. (1991) *Geography and Trade*, MIT press.

La Porta R., Lopez-De-Silanes F., Shleifer A. and Vishny R.W. (1997) Legal Determinants of External Finance, *Journal of Finance*, 1131–1150.

Licio V.M. and Pinna A.M. (2013) Measuring insularity as a state of nature. Is there a case of bad geography?, *Crenos working papers*, 22.

Limao N. and Venables A.J. (2001) Infrastructure, geographical disadvantage, transport costs, and trade, *The World Bank Economic Review*, 15, 3, 451–479.

Melitz M.J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica*, 71, 6, 1695–1725.

Milner C. and Zgovu E.K. (2003) Export Response to Trade Liberalization in the Presence of High Trade Costs: Evidence for a Landlocked African Economy, *University of Nottingham CREDIT Working Paper*, 03/04.

Mimura N., Nurse L., McLean R., Agard J., Briguglio L., Lefale P., Payet R. and Sem G. (2007) Small islands, *Climate change*, 687–716.

NASA Socioeconomic Data and Applications Center (2009) National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 2 (PLACE II), Technical report, Center for International Earth Science Information Network - CIESIN - Columbia University. Palisades, NY.

Novy D. (2013) Gravity Redux: Measuring International Trade Costs with Panel Data, *Economic Inquiry*, 51, 1, 101–121.

Nunn N. (2009) The Importance of History for Economic Development, *Annual Review of Econonomics*, 1, 1, 65–92.

Nunn N. and Puga D. (2012) Ruggedness: The blessing of bad geography in Africa, *Review of Economics and Statistics*, 94, 1, 20–36.

O'Rourke K.H. and Williamson J.G. (2001) *Globalization and history: the evolution of a nineteenth-century Atlantic economy*, Mit Press.

Ozer-Balli H. and Sørensen B.E. (2013) Interaction effects in econometrics, *Empirical Economics*, 45, 1, 583–603.

Pascali L. (2017) The wind of change: Maritime technology, trade and economic development, *American Economic Review*, forthcoming.

Riley S.J., DeGloria S.D. and Elliot R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity, *Intermountain Journal of Sciences*, 5, 1-4, 23–27.

Rodrik D., Subramanian A. and Trebbi F. (2004) Institutions rule: the primacy of institutions over geography and integration in economic development, *Journal of Economic Growth*, 9, 2, 131–165.

Russell J. and Cohn R. (2012) *Cliwoc*, Book on Demand.

Spolaore E. and Wacziarg R. (2013) How Deep Are the Roots of Economic Development?, *Journal of Economic Literature*, 51, 2, 325–69.

Volpe Martincus C., Carballo J. and Cusolito A. (2012) Routes, Exports, and Employment in Developing Countries: Following the Trace of the Inca Roads, *Inter-American Development Bank*, mimeographed.

Wei S.J. (1996) Intra-national versus international trade: how stubborn are nations in global integration?, Technical Report wp 5531, National Bureau of Economic Research.

Wolf H.C. (2000) Intranational home bias in trade, *Review of Economics and Statistics*, 82, 4, 555–563.

Wyhowski D.J. (1994) Estimation of a panel data model in the presence of correlation between regressors and a two-way error component, *Econometric Theory*, 10, 01, 130–139.