Navigating the Geography of Regional Disparities: Market Access and the Core-Periphery Divide

Gabrielle Gambuli*
CY Cergy Paris Université
ESSEC Business School
THEMA

September 26, 2023

Link to last version available

Abstract

This paper investigates the impact of market proximity on subnational development worldwide, considering the heterogeneous effects on core and peripheral regions, as well as on countries with different income levels. A gravity-based market potential index is revised to accurately assess distances for land and maritime trips to better capture geographic limitations. Estimations are performed in cross-section with country-fixed effects, by addressing endogeneity issues with instrumental variables. Robustness checks are also conducted with panel data on a smaller sample. The findings reveal that regions with better access to markets and port experience higher regional income per capita, with the effect being higher for wealthier regions. Peripheral regions consistently exhibit a 2 percentage point lower elasticity to market potential compared to core regions. The paper also highlights the potential negative impact of proximity to foreign markets on peripheral regions, especially if they are central to foreign core markets without free trade agreement in place between the respective countries. Results suggest that policies which aim at improving the connectivity of peripheral regions to core domestic markets and develop trade agreements could help mitigate the adverse effects of trade barriers and reduce regional disparities within countries.

Keywords: Market Potential, Economic Geography, Regional Development, Core-Periphery
JEL Classification: F15, O18, R11

*Contact: `gabrielle.gambuli@cyu.fr` - **Acknowledgements:** I am particularly grateful to Sara Biancini and Rodrigo Paillacar for their invaluable guidance and support throughout this research project. Comments received from participants to the following conferences are gratefully acknowledged: 22nd INFER Annual Conference, ESSEC PhD Poster Session, 20th Doctoral Meetings RIEF, 22nd ETSG Annual Conference, 31th ITFA Conference, 18th ACDD Doctoral Day and 20th RSEP Conference. Finally, I would like to thank my colleagues who provided helpful feedback during internal seminar presentations and the writing process: Margaux Suteau, Thomas Rowley, Mélanie Marten and Lucas Javaudin.

1 Introduction

The core-periphery divide has significant consequences, giving rise to welfare concerns related to unequal access to resources such as education and healthcare, as well as limited employment opportunities. These disparities can extend to the rise of protest movements and voting decisions cleavage, ultimately shaping a nation's political landscape. Among recent prominent examples are Brexit in the United Kingdom, the election of Trump in the United States, and the rise of far-right votes in various European countries. They have shed light on the deep-seated frustrations arising from regional disparities and have highlighted the sense of marginalization and resentment felt by individuals who perceive that urban elites are favored over citizens in rural and peripheral regions (Abreu and Öner, 2020; Dijkstra et al., 2020; Rodríguez-Pose et al., 2020).

Understanding the causes of regional development disparities is crucial to reduce inequalities within countries and foster social cohesion. Among these factors, proximity to markets has been suggested as a key factor of development. Proximity to demand enables firms to reduce transportation costs when shipping their goods. As a result, they can offer higher wages to their workers, leading to the development of the firm's location as the population becomes wealthier. This concept is captured in the *international trade wage equation* developed by Fujita et al. (1999). However, this framework assumes labor immobility, which may be less applicable when examining within countries disparties, where labor mobility barriers are typically lower than at the international level.

Another crucial factor in regional development is the presence of agglomeration economies. When considering spatial equilibrium models that account for people's mobility within countries, agglomeration economies are identified as a centripetal force that benefits wealthier areas. However, they can have adverse effects on less affluent regions. This is because firms, industries, and workers tend to cluster in close proximity to each other, where economic activity is concentrated (Rosenthal and Strange, 2004). This clustering leads to productivity gains, which translate into higher wages in these agglomerated areas.

A third critical factor in regional development is education, as evidenced by the Lucas-Lucas model introduced by Gennaioli et al. (2013). In this model, a worker's productivity is shaped by both the average skills and education of the workforce in their region and their individual skills and education. Higher levels of education among the local workforce in a region correspond to high marginal productivity of labor in that particular region. This, in turn, translates into higher wage levels. The persistence of population in less productive areas is attributed to certain mobility frictions, such as congestion forces stemming from factors like housing and land prices.

Institutional quality has also been designated as an important development factor, as underscored by Acemoglu et al. (2005). However, within countries, institutional differences may be in general limited to certain exceptions. Notably, Gennaioli et al. (2013) find that regional variations in institutional quality, measured by factors such as ease of starting a business, fail to account for disparities in regional Gross Domestic Product per capita. Additionally, cultural factors, including the number of ethnicities, do not exhibit any significant influence. Consequently, the present paper does not delve into an investigation of these factors.

(How much) Does proximity to markets matter in explaining intranational regional development disparities conditional on regional productivity? This paper contributes to the ongoing debate in the New Economic Geography literature by addressing broader questions. It explores the consistency of this effect across countries, the robustness of different measures of market access, and the differential impact on core and peripheral regions. Importantly, the study accounts for productivity-related factors that may also have an influence on regional disparities as discussed above, including education and population density among others. The study also investigates the distinct effects of centrality to domestic and foreign markets on core and peripheral regions.

Despite the growing concern for peripheral areas being left behind, literature has shown limited exploration of the distinct effects of market access on the development disparities between core and peripheral regions (Brülhart et al., 2004; Brülhart, 2006; Baum-Snow et al., 2020; Brülhart et al., 2020). A notable subset of this research is dedicated to the concept of the *border shadow*, which characterizes the lower levels of development experienced by regions located in close proximity to international borders, as discussed in studies such as Adam et al. (2023) and Bonadio et al. (2023). These regions often contend with trade barriers, limited market access, and reduced economic opportunities due to their location near borders. This paper aims to contribute to the discussion by conducting an analysis of the core-periphery divide and the differential impact of market access.

These questions are crucial in the context of increasing integration and interdependence among countries and regions, where foreign markets and internal transportation infrastructures play a substantial role in shaping subnational development. This integration and interdependence have been pivotal in recent events, including the Brexit referendum and voting decisions in the United States and various European countries. By examining the relationship between market proximity and regional development, policymakers can develop more targeted policies to foster sustainable and inclusive growth.

Empirical tests of the international trade wage equation of Fujita et al. (1999) emerged in the last two decades, originally focusing on inequality across countries (Redding and Venables, 2004; Fingleton, 2008; Head and Mayer, 2011; Jacks and Novy, 2018). Proximity to wealthy countries is consistently related to their development level. However, considering a country as a market is a substantial simplification. The vast literature of firm location theory together with the core-periphery model developed by Krugman (1991) justify this argument. The concentration of economic activity is often observed in specific locations, commonly referred to as the "core." Typically, core regions are home to a megapolis, and experience vastly different conditions compared to other regions within the same country (Baldwin and Martin, 2004; S. J. Redding, 2022).

In the last decade, the focus of the falsification test of the international trade wage equation has switched from cross-countries inequality to within-country inequality. The literature on subnational development emphasizes the diversity of regions within a country and recognizes that subnational approaches can be more effective than national approaches for promoting development. Regions can also be a more relevant unit of analysis for representing markets and understanding the impact of market proximity. For these reasons, the present paper adopt a regional-level position. Due to data availability of development measures and covariates at the regional level, the literature has focused on single-country analysis. Papers have studied the case of the United States, European countries, and more recently, emerging and developing countries. The wider geographical scope of analysis in the literature is found to be at the European level using the rich data produced by European institutions (Niebuhr, 2006; Head and Mayer, 2004; Brülhart et al., 2004; Breinlich, 2006; Head and Mayer, 2006; Brakman et al., 2009; Bruna et al., 2016).

The generalization of the statement to the entire world is hindered by the scarcity of data on regional development measures and covariates, as well as the lack of data required for computing sophisticated market access indexes. The scarcity of data on physical infrastructure for each country and the underutilization of the gravity trade literature, which examines the impact of different distance measures on trade, contribute to these limitations despite their significance as fundamental components for market access measures. This paper aims to address these gaps by making four significant contributions to the literature.

The first contribution of the paper is to provide a falsification test of the international trade wage equation at the regional level within countries and generalize the test to the whole world using a complete set of countries. For robustness, I control for regional development covariates which proxy for regional productivity, such as population density, education, natural resources, and climatic conditions. The study relies on the large regional dataset built by Gennaioli et al. (2013), which gathers information on Gross Domestic Product (GDP) and covariates for more than 1,500 subnational regions in 107 countries. The dataset covers about 70% of the world's surface and more than 90% of its GDP for the year 2005. For robustness checks, I use a regional panel dataset of about 1,000 regions in 72 countries from Gennaioli et al. (2014) with similar information.

The second contribution is the computation of the index of proximity to markets, the *market potential*,[1] which summarizes the market capabilities of all regions, weighted by their geographical proximity. An effort has been made to fully consider international interactions from trade building a gravity-based index. The proximity function takes into account factors such as common language, national contiguity, colonial ties, common currency, and regional trade agreement as determinants that facilitate closer trade relations between countries and regions, even when they are physically distant. Additionally, since "around 80 per cent of global trade by volume and over 70 per cent of global trade by value are carried by sea and are handled by ports worldwide" as reported by (The Review of Maritime Transport 2018, UNCTAD), I consider international distances between regions through land and maritime shortest paths passing by world ports.[2] Considering the predominant role of transportation infrastructures and geographic limitations allows to recover the travel path of commodities as closely as possible.

The third contribution is on the investigation of the differences in the elasticity coefficients to market potential between core and peripheral regions and how differently they are affected by proximity to foreign markets.[3] To this matter, a classification of regions is done using k-means clustering conditional on the regional economic production. Each country is split into 3 groups of subnational regions: the core, the semi-periphery and the periphery. While core regions benefit from a better access to larger markets, peripheral regions face higher transportation costs, making it more difficult for them to attract investment and develop competitive industries. By employing this classification of regions, the analysis is better equipped to consider the regional variation within countries and gain a better understanding of the specific challenges faced by peripheral regions, specifically according to proximity to domestic and foreign core regions.

The fourth and final contribution lies in a comprehensive examination of the mitigation effects of proximity to markets, specifically focusing on centrality to core markets, distinguishing between domestic and foreign contexts. Integrating centrality measures into the regression analysis not only amplifies the variability in the explanatory variable but also allows for a meaningful contribution to the border shadow literature (Adam et al., 2023; Bonadio et al., 2023). Additionally, differentiating between centrality to core foreign markets with and without free trade agreements underscores their role in reducing trade barriers, which can be particularly important for regions in the periphery of the economic activity. The effect of improved centrality to foreign cores with a trade agreement was investigated using the panel sample.

The analysis is first performed in cross-section to gather as many observations worldwide as

[1] The concepts of market potential and market access are used interchangeably throughout the paper.

[2] The title *Navigating the Geography of Regional Disparities* is justified by this aspect of the paper. The term "navigating" is used to refer to the maritime routes taken by ships between world ports, which serves as a metaphor for the investigation of the determinants of subnational disparities.

[3] The initial endeavor in this area was conducted by Baum-Snow et al. (2020). They undertook a classification of prefecture regions based on population size, subsequently identifying primate prefectures. They analyzed the differences in market access elasticity coefficients on GDP between primate and non-primate prefectures.

possible. The country fixed-effects allow to control for national unobservable variables, as well as to compare the regional observations with their national average for each country. I address endogeneity issues with instrumental variables using a centrality index and a spatial lagged market potential as instruments. The former summarizes the central position to others, and the latter the proximity to foreign markets. Finally, to further test the robustness of the results, the analysis is conducted in panel on a smaller sample of countries.

Results show that proximity to markets is a strong and robust determinant of subnational development all around the world. A 1% increase in the regional market potential is associated with an increase in GDP per capita about 0.1% with respect to countries' average. This elasticity coefficient is similar to those found in single-country analysis (Brakman et al., 2004; Pires, 2006; Mion and Naticchioni, 2009; Fally et al., 2010; Hering and Poncet, 2010; Kosfeld and Eckey, 2010; Paredes, 2013), as well as in European regional analysis (Niebuhr, 2006; Head and Mayer, 2006; Brakman et al., 2009). Conversely, a lower coefficient was identified in comparison to those obtained through country-level analysis (Head and Mayer, 2011), which can attributed to the relatively lower variation in development levels within countries when compared to the variation across countries, or to a more effective control for omitted variables bias. Additionally, the access to ports is revealing to be a powerful determinant of both higher market potential and higher income per capita.

Upon further investigation, heterogeneous elasticity coefficients of regional development to market potential are observed, revealing distinct impacts based on regions' GDP levels. Core regions benefit more from market proximity compared to the (semi-)periphery, with peripheral regions exhibiting a coefficient that is 0.02 percentage points lower than that of core regions. While the proximity to foreign regions has no significant effect on core regions, it can adversely affect regions in the (semi-)periphery. Notably, among peripheral regions, those with higher centrality to foreign cores experience lower levels of development within the same country. This finding aligns with Adam et al. (2023) and Bonadio et al. (2023), which demonstrates that subnational regions located at international borders, tend to have lower income per capita levels due to their low centrality to domestic markets and international trade barriers. It becomes evident that not only border regions but also regions in the periphery of national economic activity can be adversely impacted by close proximity to foreign markets.

Results show that proximity to domestic cores exhibits positive effects on regional development, while proximity to foreign cores without a trade agreement has a negative impact on peripheral regions. A discussion on firms selection attempting to explain the latter result is provided. However, the effect of proximity to foreign cores with a free trade agreement is interestingly not significantly different from zero. This suggests that the existence of a FTA can help alleviate the negative impact of close proximity to foreign core markets, as supported by the findings of Adam et al. (2023) and Bonadio et al. (2023). The panel data analysis reveals that when countries establish a trade agreement and enhance their connections to foreign cores through FTAs, there is an expected increase in a region's GDP per capita, although the statistical significance is limited.

Results highlight the concern for the development divide between core and periphery regions, where the latter tend to have lower income per capita, market potential, and growth rate. To reduce regional disparities and bridge the development gap between core and periphery regions, one possible approach following the present paper's results is to enhance subnational integration by improving national connectivity through transportation infrastructure. Ensuring better access to core national regions may prevent the periphery from being left behind and being excessively impacted by a too close proximity to foreign cores. However, this approach may not be suitable for low-income countries, as indicated by the findings across various country samples. It may pose a risk of regional economic activity and population shifting away from the periphery towards core regions.

Conversely, policies geared towards enhancing trade relations and forging free trade agreements with foreign nations can yield advantages for peripheral regions. Simultaneoulsy, policymakers should also work on boosting economic activity in these peripheral regions to prevent economic activity from moving to the core areas.

The paper is organized as follows. Section 2 provides a review of the existing literature on regional development and market potential and highlights the gaps in the literature that this paper aims to fill. Section 3 introduces the regional market potential indexes used for the analysis. Section 4 presents the data. Section 5 presents the descriptive statistics. Section 6 develops the empirical method. Section 7 shows and interprets the results. Finally, Section 8 concludes.

2 Literature

2.1 Wage Equation and Regional Development

Since the emergence of the New Economic Geography literature following the eminent works of Krugman (1980) and Krugman (1991), accessibility to markets has been showed to have a strong impact on the spatial distribution of economic activity. Fujita et al. (1999) have developed a full general equilibrium model with international trade and monopolistic competition, composed by agricultural and manufacturing sectors, which relies on labor immobility. Firms operate under increasing returns to scale and produce differentiated products. From profit maximization, the model express the increasing relationship between factor incomes and market access. This relationship is referred as the *international trade wage equation*.

In particular, a special focus is given on the model developed by Redding and Venables (2004). Redding and Venables' model, which extends the Fujita model by including transport frictions in trade and intermediate goods in production, shows that geographic location affects per capita income via flows of commodities, production factors, and ideas. Distant countries incur trade barriers such as transport costs and cultural differences, leading to reduced market access for exports and imports and lower maximum wage firms can afford to pay due to their zero-profit condition.

Head and Mayer (2011) developed a similar intuition based on the gravity equation for bilateral trade flows and market clearing conditions. The focus has been made on differences in wages, as labor is assumed to be the immobile factor in these theoretical models. However, the wage equation can be generalized to all immobile inputs. Both model imply that distance to markets affects the equilibrium factor prices and so, rises cross-countries income inequality.

The empirical regional-level analysis has widely been developed in the last decade. The literature has focused on single-country analysis due to data availability, such as the United States, European countries, and more recently, emerging and developing countries.[4] The wider geographical scope of analysis in the literature is found to be at the European level (Niebuhr, 2006; Head and Mayer, 2004; Breinlich, 2006; Head and Mayer, 2006; Brakman et al., 2009; Bruna et al., 2016). The

[4]Single-country studies of the effect of market access on regional development have been conducted for the United States (Hanson, 2005; Fallah et al., 2009; Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021), Italy (Mion, 2004; Mion and Naticchioni, 2009; A'Hearn and Venables, 2013; Daniele et al., 2018), Spain (Pires, 2006; López-Rodrıéguez et al., 2008), Germany (Brakman et al., 2004; Kosfeld and Eckey, 2010) or Finland (Ottaviano and Pinelli, 2006). Other studies have focused on emerging and developing countries, such as Mexico (Chiquiar, 2008), Chile (Paredes, 2013), Brazil (Fally et al., 2010; Hering and Paillacar, 2016), Turkey (Karahasan et al., 2016; Özgüzel, 2022), China (Hering and Poncet, 2010; Baum-Snow et al., 2020; Zou et al., 2021), India (Donaldson, 2018; Bonadio, 2022).

statement has not yet been generalized to the whole world, which is the intention of the present paper.

The literature on subnational development distinguishes between two types of geography that characterize regions and cities: first-nature and second-nature. The former refers to regional natural endowment, such as proximity to water, altitude or climate. The latter refers to the interactions a region can entertain with other regions, depending on its location with respect to them. The literature has argued that the attractiveness of a region is best described by intra-regional characteristics (P. P. Combes et al., 2005; Brakman et al., 2009).[5] The present paper is going to question this statement.

In particular, my analysis was inspired by Gennaioli et al. (2013)'s, who investigate the determinants of regional development according to human capital and institutional quality, controlling for first-nature geography. They attest the primary importance of human capital for both regional and national development, while the institutional quality is only significant at the country level. While an effort has been made regarding the first-nature geography dimension, they have omitted the spatial interdependence of regions. Nevertheless, they considered proximity to the coast, which can be used as a proxy for proximity to foreign markets. The variable is showed to have a highly positive significant impact on the regional development levels.

The present paper reevaluates the determinants of regional development with a proper measure of accessibility to markets, referred as market potential, controlling for first-nature geography and education.

2.2 Market Potential

Market potential measures the proximity to demand from a given location and represents the attractiveness of a region based on its geographic position and that of its potential trading partners. It reflects a region's potential to trade with others and access different markets. It is calculated by adding the economic size of all other regions, weighted by the distance that separates them. Regions with high market potential are those that are close to wealthy regions with high demand and market capacities, which allows for low transport costs and higher trade volumes. In contrast, periphery regions, far from the demand, experience low market potential.

The internal market of a region also determines its level of market potential. The local market potential depends on its own market capacity and its internal distance as a proxy for internal transport costs. The wealthier the market, the higher the region's own demand, and the higher its market potential. But the larger its land area, the higher the transport costs faced by firms to sell their commodities on average, and the lower the local market potential.

Three different computation methods of the index are developed in the literature: (1) the simple index, as expressed in Harris (1954), (2) the gravity index and (3) the infrastructure index. The first method is a simple function summing the income of the geographical entities, weighted by the inverse of the distance between them (Mion, 2004; Hanson, 2005; Mion and Naticchioni, 2009; López-Rodrıéguez and Andrés Faıéña, 2006; Fallah et al., 2009). Despite the relevant effort of computing market potential indexes which may better translate market accessibility forces, the predictive power of this index has been shown to be as high as other more sophisticated indexes (Daniele et al., 2018).

[5]Studying the case of Chile, Paredes (2013) claims that spatial wage inequality may be better explained by amenities than by market access in low income countries. However, some papers demonstrate that the wage equation can holds also in developing and emerging economies (Chiquiar, 2008; Fally et al., 2010; Karahasan et al., 2016; Hering and Poncet, 2010).

The second computation method relies on the estimation of the international trade gravity equation and uses other variables of proximity to weight the markets' income, as well as price indexes obtained from exporter and importer fixed effects. This method was adopted by Redding and Venables (2004) and Head and Mayer (2011) for computing indexes at the national level. At the regional level, price indexes has hardly been computed from gravity-based fixed effects since intra-national trade data between sub-national entities is not commonly observable. Some authors have done so in the case of Brazil (Fally et al., 2010; Hering and Paillacar, 2016) and China (Hering and Poncet, 2010). In addition to physical distance between regions or cities, they include other type of proximity measures such as contiguity, common language and regional trade agreement.

The third method for market potential index computation usually relies on the expression of the simple market potential index developed by Harris (1954), using distance measures considering physical infrastructures. A first attempt was made by Hanson (2005), considering a "hub-and-spoke" geodesic distance function, which assumes that goods transported from one county to another must pass through a transportation hub located in the home state of the originating county. Since, road and rail distance has been taken into account (Pires, 2006; Ottaviano and Pinelli, 2006; Paredes, 2013; A'Hearn and Venables, 2013; Karahasan et al., 2016), as well as travel time by car (Brakman et al., 2004; Niebuhr, 2006; Baum-Snow et al., 2020) or train (Zou et al., 2021). Other studies have incurred freight transportation costs as well. For example, Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021) compute average price per ton of transported goods by available transportation routes, using rail and waterways distance. They also use different estimates of elasticity to distance rather than the -1 coefficient used in the simple market potential index.

Few papers integrate the role of international markets in their single-country analysis. An even smallest amount considers the predominant role of ports and maritime transport networks (Ducruet, 2020). In the case of China, Baum-Snow et al. (2020) considers road travel time from each prefecture city to the nearest of the nine largest international ports by value of shipments, as well as the average cost of shipment. In the case of India, Bonadio (2022) estimates trade costs related to ports' access via roads as well as ports' quality, in order to evaluate the relative importance of both infrastructures in shaping international market access of regions. The worldwide scope of my analysis does not allow me to integrate ports' quality, neither inland transportation networks in the market potential indexes. However, an effort has been made to give structure to the function of physical distance.

In this paper, the gravity-based approach is used, with an effort towards the infrastructure approach. I compute two indexes with different specifications of the distance function. Both indices integrates cultural, historical and economical proximity, such as done in the literature using the simple and the gravity index, but different measures of physical distance. The first index uses the haversine distance function, while for the second index, the distance function is provided with more structure, considering the inland shortest path between regions and ports, as well as the overseas shortest path between ports. This distance function considers the world's geography by integrating the delimitation of land and sea spaces and the predominant role of ports in international trade. Next subsection presents the regional market potential indexes.

3 Regional Market Potential

To compute the potential to trade of each region, I assume a complete network of regions, where they are supposed to be all interlinked. The intensity of their potential interactions is proportional to their market size weighted by their proximity. The market potential index is expressed as follows:

$$\text{MP}_i = \sum_{j \neq i} y_j \frac{\tilde{y}_j}{\tilde{y}_{\max,c_j}} \tau_{ij} + b \times y_i \frac{\tilde{y}_i}{\tilde{y}_{\max,c_i}} \tau_{ii} \tag{1}$$

where y_j is the GDP of region j, $\tilde{y}_j$ is the GDP per capita of region j, $\tilde{y}_{\max,c_j} = \max_{1\leq k\leq n} \tilde{y}_k$ is the maximum regional GDP per capita among observations within country c_j, τ_{ij} represents the trade costs between regions i and j, τ_{ii} is the internal transport costs in region i, and b is the regional border effect.

The market potential index is composed by two components: the local market potential and the non-local market potential. The former is the second term on the formula's right-hand side, composed by the income information of region i and the weight ranking his hierarchy position among the richest regions in term of GDP. The regional border effect b is giving weight to the internal demand of region i[6]. The intra-regional transportation costs are expressed as follows[7]:

$$\tau_{ii} = \left[\frac{2}{3} \sqrt{\frac{\text{area}_i}{\pi}} \right]^{-1} \tag{2}$$

The non-local market potential NLMP_i - first term on the right-hand side of equation 1 - corresponds to the sum of the GDP of all regions j different to i, which is multiplied by a weight indicating how well is ranked its income per capita compared to the richest region, i.e. the one with the highest level of GDP per capita.[8] Hence, if j is the richest region of the sample, the coefficient will be equal to 1. Otherwise, it belongs to]0;1[. This ratio gives more weight to richer regions and illustrates their competition power. The closest to 1 the ratio, the higher region's j market power. This component gives the market capacity of region j and is reduced by the trade costs, τ_{ij}, that it faces to import commodities from region i.

For the latter variable, I consider two specifications. The first one is expressed as follows:

$$\tau_{ij}^{(1)} = \text{dist}_{ij}^{(\text{haversine})\hat{\beta}_1} \times \left[\mathbb{1}_{\{c_i \neq c_j\}} e^{\hat{\beta}_2 \mathbb{1}_{\text{language}_{ij}} + \hat{\beta}_3 \mathbb{1}_{\text{contiguity}_{ij}} + \hat{\beta}_4 \mathbb{1}_{\text{colony}_{ij}} + \hat{\beta}_5 \mathbb{1}_{\text{rta}_{ij}} + \hat{\beta}_6 \mathbb{1}_{\text{currency}_{ij}}} + \mathbb{1}_{\{c_i = c_j\}} e^{\hat{\beta}_7} \right] \tag{3}$$

[6]The regional border effect is very rarely estimated in the literature due to the lack of trade flows data between regions of a same country. To proxy for this border effect, I compute b as the average ratio between the maximum domestic market potential on the maximum local market potential among regions in each country, with domestic market potential being the regional market potential with respect to all regions in the same country and the local market potential being the very own market potential of a region, excluding other regions. This ratio forces the local market for a region to weight in the total domestic market, since local demand is also important. I find that $b = 1.5$ for $\text{MP}^{(h)}$ index, and $b = 1.8$ for $\text{MP}^{(s)}$ index. As a comparison, Coughlin and Novy (2021) have estimated states' border effect to be about 1.5 on average in the case of the United States of America for balanced sample over the years 1993, 1997, 2002, and 2000. Hoping that future research will attend to estimate the proper regional border effect within countries for a wider set of countries.

[7]Regions are assumed to be circle-shaped, with the capital city at their center. Then, the average distance from the center of the region and its borders is calculated as a proportion of its radius, with the radius computed as the square root of the land area divided by π. The radius is weighted by 2/3, assuming that the economic activity is relatively concentrated toward the center (Head and Mayer, 2011). Hence, this distance function is an approximation of the average distance between every two cities in region i.

[8]This ratio is included in the index in order to integrate regional competitiveness. To compute the index with a gravity-based approach, importer and exporter fixed effects are usually included. However, the international trade data used in this paper rely on total bilateral trade flows. Thus, trade flows are not directed - it is not possible to identify an importer and an exporter - and are not at the regional level. Thus, including importer and exporter fixed effects is not relevant.

with $\text{dist}_{ij}^{(\text{haversine})}$ the physical distance between regions i and j, computed as the haversine distance, determined by the great-circle distance between two geocodes, which proxies for transportation costs. $\mathbb{1}_{\{c_i=c_j\}}$ and $\mathbb{1}_{\{c_i\neq c_j\}}$ dummies indicating whether region j belongs to the same country than region i. $\mathbb{1}_{\{c_i=c_j\}}e^{\hat{\beta}_7}$ corresponds to the national border effect.[9] It reflects the fact that a pair of regions belonging to the same country is more prone to trade together than with a region abroad, due to more similar consumption preferences, easier communication and the absence of tariffs and customs.

The set of dummy variables equal interacted with the dummy $\mathbb{1}_{\{c_i=c_j\}}$ are equal to one if countries c_i and c_j of regions i and j respectively share a common border, colonial ties, a regional trade agreement and the same currency, as well as if regions i and j share a common language predominantly spoken. These latter variables reflect the cultural and economic proximity between two regions from different countries, which alleviates the effect of distance between them. When $\mathbb{1}_{\text{language}_{ij}}$, $\mathbb{1}_{\text{contiguity}_{ij}}$ and $\mathbb{1}_{\text{colony}_{ij}}$ dummies equal one, they represent lower communication costs, more similar preferences and long term trade, historical and political relationships, which are supposed to increase trade volumes between two regions. $\mathbb{1}_{\text{rta}_{ij}}$ dummy indicates whether the two countries have a trade agreement that lowers trade costs, and $\mathbb{1}_{\text{currency}_{ij}}$ represents the ease for two countries trading with the same currency, as they avoid buying foreign exchange.

The distance between two countries is usually computed as the haversine distance. However, it is a basic measure of distance. This paper provides a better measure of geographical distance, which aims at following as closely as possible the travel undergone by goods from a territory to another, distinguishing the world's areas covered by land and ocean[10]. The new distance measure is built considering port access[11] for trade that might be undertaken overseas, as well as the shape of land masses, so the path between two regions can get around bodies of water.

To do so, I first collect data on world ports geolocation, and use high-resolution geography data from Wessel and Smith (1996), which provides boundaries between land masses and ocean. From the geographical polygon, I give different values for land masses and ocean and rasterize the map in order to compute shortest paths between each pair of points. The shortest path is computed following the grids of the maps between these points. I first compute the shortest path between each pair of regions by land, then between each pair of region-port by land, and finally, the shortest path between pairs of ports by sea (see figure 8 in the appendix for a visual example of the computation). The dataset results in distances computed between each pair of regions, with their associated closest port[12], combining distance by land and sea.

Then, the shipment distance function is computed as follows:

$$\text{dist}_{ij}^{(\text{shipment})} = \begin{cases} \kappa_{io}^{\gamma_1}\kappa_{od}^{\gamma_2}\kappa_{dj}^{\gamma_3} & \text{, if } \kappa_{od} \neq 0 \land \frac{\kappa_{io}+\kappa_{dj}+\kappa_{od}}{\kappa_{ij}} < 2 \\ \kappa_{ij}^{\gamma_4} & \text{, otherwise} \end{cases} \tag{4}$$

where κ_{od} corresponds to a proxy for the ocean shipping costs. It is the shortest path between

[9] The literature has estimated $\hat{\beta}_7 = 1.96$ on average (Head and Mayer, 2014).

[10] In 2002, sea transport accounted for 44% of goods, measured in value, traded between the EU and the rest of the world. Measured in volume, the share was about 78%. Regarding road and rail transportation, the share was about 30% in value and 13% in volume (Source: Statistics Explained - Eurostat, 2021). As air transportation can be approximated by the haversine distance function, and that it represents a lower share of traded goods as measured both in value and volume, it is not particularly considered it in this paper.

[11] The importance of ports on trade performance has been highlighted in the literature on trade facilitation and logistics (Limão and Venables, 2001; Blonigen and W. W. Wilson, 2008; Portugal-Perez and J. S. Wilson, 2012).

[12] Bonadio (2022) shows however that firms do not always use the closest port, but incur additional internal costs to reach a port of higher quality.

the origin port o and the destination port d. Variable κ_{io} is the distance from the region center of region i to its closest port o, and κ_{dj} is the distance from the region center of region j to its closest port d. Hence, κ_{io} and κ_{dj} is the distance traveled by land, and κ_{od} is the distance traveled by sea. I denote by κ_{ij} the distance between regions i and j by land, without passing by any port. If regions i and j are not connected by land, I define $\kappa_{ij} = \infty$. Coefficients γ_1, γ_2, γ_3 and γ_4 are expected to be negative.

The maritime distance function structure $\kappa_{io}^{\gamma_1}\kappa_{od}^{\gamma_2}\kappa_{dj}^{\gamma_3}$ is similar to the one of Bonadio (2022), but he rather uses a port quality measure instead of the maritime distance between ports. Port quality is an important factor of trade costs, however I do not have enough information to estimate it. Moreover, considering maritime distance between ports is relevant as transportation costs also increase with distance on the sea.

The inland and overseas shortest path between regions and ports are computed as straight lines and curves, assuming a continuous space with continental borders constraints. I do not consider physical transportation routes and real costs in the measure as it has been done by some papers in the literature presented in the previous subsection. To do so with a worldwide scope is not possible since this data is not available for most of the countries. The method used in this paper still made progress in considering the actual world geography.

Equation 4 integrates a condition of convenience with respect to sea freight. It considers if it is impossible to travel from i to j by land due to non-continuous land area, and if the whole distance between i and j by sea is not larger than twice the land distance between the two, if and only if the maritime distance does not equal zero. Indeed, two regions or countries will not find any interest to reach the same port to trade between each other. The ratio of distances lower than 2 assume regions to use land transportation if the distance travelled by sea is too important with respect to the land distance between two regions, even if maritime transportation is known to display scale economies in the whole transportation costs.

I define $\mathbb{1}_{\text{maritime route}}$ as the convenience condition to ship goods overseas. It equals one if the condition is respected, i.e. if it is convenient for exporters to send its good overseas, zero otherwise. Then, the alternative measure of trade costs becomes:

$$\begin{aligned}\tau_{ij}^{(2)} \quad = \quad & \left[\kappa_{io}^{\hat{\gamma}_1}\kappa_{od}^{\hat{\gamma}_2}\kappa_{dj}^{\hat{\gamma}_3}\ \mathbb{1}_{\text{maritime route}} + \kappa_{ij}^{\hat{\gamma}_4}\left(1-\mathbb{1}_{\text{maritime route}}\right)\right] \times \left[\mathbb{1}_{\{c_i=c_j\}}e^{\hat{\beta}_7}\right. \\ & \left. +\mathbb{1}_{\{c_i\neq c_j\}}e^{\hat{\beta}_2\mathbb{1}_{\text{language}_{ij}}+\hat{\beta}_3\mathbb{1}_{\text{contiguity}_{ij}}+\ \hat{\beta}_4\mathbb{1}_{\text{colony}_{ij}}+\hat{\beta}_5\mathbb{1}_{\text{rta}_{ij}}+\hat{\beta}_6\mathbb{1}_{\text{currency}_{ij}}}\right]\end{aligned} \tag{5}$$

with $\tau_{ij}^{(1)}$ the trade costs using the haversine distance as a proxy for transportation costs, $\tau_{ij}^{(2)}$ the trade costs using the shipment distance function which considers land and sea areas, as well as ports location.

As described above, two indexes of regional market potential are computed with two different measures of geographical distance, the first being computed as the haversine distance $\text{dist}_{ij}^{(h)}$, and the other as the overseas shipping distance $\text{dist}_{ij}^{(s)}$. Hence, the analysis will be conducted using the two different indexes of market potential $\text{MP}_i^{(h)}$ and $\text{MP}_i^{(s)}$. Section A presents the methodology for the computation of the indexes. Coefficients of equations 3 and 5 are estimated to be: $\left[\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\beta}_4,\hat{\beta}_5,\hat{\beta}_6,\hat{\beta}_7,\hat{\gamma}_1,\hat{\gamma}_2,\hat{\gamma}_3,\hat{\gamma}_4\right] = [-1.18, 0.66, 1.12, 1.37, 0.47, 0.79, 1.96, -0.07, -0.96, -0.06, -1.00]$

Next section describes the regional data used for the computation of the indexes and the analysis.

4 Data

The regional data used in this paper mainly come from the dataset built by Gennaioli et al. (2013), with more than 1,500 subnational regions, from 107 different countries, out of the 195 recognized around the world, in the year 2005[13]. This substantially large dataset considers a broad range of geographical, educational, cultural and institutional variables to disentangle the determinants of development at the regional level all over the world.

The regional data of Gennaioli et al. (2013) have been collected at the highest administrative level available for each country: regions, provinces, counties and municipalities, and then aggregated the data to regions and provinces. Since Gennaioli et al. (2013) focus on regions, and aim at identifying the determinants of development within countries, they do not include countries with no subnational administrative divisions in the sample.

The dataset contains the Gross Domestic Product (GDP) per capita of regions on the sample, measured in current purchasing-power-parity (PPP) dollars[14]. This variable is used to proxy for regional development as the dependant variable in the analysis. I also use the GDP as one of the main component of the indexes of market potential. Figure 6 in the appendix shows the world map with the regional data available.

The dataset also includes three measures of geography and natural resources which proxy for the first nature geography: temperature, oil production per capita and the inverse distance to the coast. The first two are gathered from the WorldClim database. The temperature in 2005 is proxied by the average temperature during the period 1950-2000.

Then, two proxy variables of the regional human capital are the population size, which comes from the City Population database, and the average years of education of the 15-year-old and older population, collected from UNESCO database. Zero year of school are allocated for the pre-primary level, six additional school years for the primary education and twelve for those who have completed secondary school. For each region, average years of schooling is computed as "the weighted sum of the years of school required to achieve each educational level". The choice of these variables to explain the development of regions and countries is deeply justified in the development and economic geography literature. Thus, they are not discussed in this paper.

The variable of interest, the regional market potential, is based on the GDP of each region, which proxies for the size of their demand size. To get the most realistic market potential indexes as possible, it is best to consider a complete network of every region around the world. However, there are missing data in the data set of Gennaioli et al. (2013) as they did not include countries for which there were no subnational division data. To solve for this problem in the indexes computation, the GDP information of 68 countries is added from the gravity database developed by CEPII. This extrapolation allows to include a hundred more observation of world's markets into the market potential indexes computation.

In order to compute the distance between each region, which proxies for transport costs, the

[13]The dataset is available on the website of Rafael La Porta.

[14]The dataset built by Gennaioli et al. (2013) gathers regional income data for 107 countries in 2005, drawn from sources including national statistics offices and other government agencies (42 countries), Human Development Reports (36 countries), OECDStats (26 countries), the World Bank Living Standards Measurement Survey (Ghana and Kazakhstan), and IPUMS (Israel). Their measure of regional income per capita is based on value added but they used data on income (6 countries), expenditure (8 countries), wages (3 countries), gross value added (2 countries), and consumption, investment, and government expenditure (1 country) to fill in missing values. They measured regional income in current PPP dollars because they lacked data on regional price indexes. To ensure consistency with the national GDP figures reported by World Development Indicators, they adjusted regional income values so that, when weighted by population, they total the GDP at the country level.

latitude and longitude are needed for each region. To do so, regions are geocoded on R software, by using a geocoding API key with Google Maps Plateform. After having collected latitude and longitude for each region, an expanded dataset that associates every pair of regions is created, and the distance between each of them is computed using the haversine formula for the measure $\text{dist}_{ij}^{(h)}$, and as the shortest path by land and sea for the measure $\text{dist}_{ij}^{(s)}$ presented in section 3.

In order to compute the average internal distance of each region, the land area information is used from Statoids.[15] In order to relate for cultural distances between regions, information on the languages spoken in each region is collected from the same source. Data on countries contiguity and colonial ties is collected from CEPII gravity database (Gaulier and Zignago, 2010). Moreover, information on regional trade agreement and currencies at the country level are collected from the country-level dataset by de Sousa (2012)[16]. The dataset covers 199 countries for the time period 1958-2015. The CEPII database also gathers information on trade for all country pairs between 1996 and 2020.[17] I use this data to estimate the trade elasticity coefficients to resistance and facilitator variables from the gravity equation.

Finally, geodata on global ports from WPI database are collected in order to compute the shipment distance functions.[18] The database contains the location and physical characteristics of 3,630 ports worldwide, as well as the facilities and services that they offer. I select coastal ports of a substantial size,[19] excluding the smallest ports that may not be used for trade. I also exclude river ports,[20] since rivers' recognition make considerably heavier the shortest path algorithm work. Then, I match to each region its closest port. By doing so, 463 ports remain. Figure 7 in Appendix B shows the matched ports on a map.

To evaluate if the relationship between the market potential and regional development is robust and persistent over time, I gather panel data from Gennaioli et al. (2014) for the same variables present in their cross-section version for the period 1976-2010.[21] Since there is a substantial amount of missing observations in the early years, I restrict the panel to the period from 1995 to 2010, with 5 years interval and adjust some observations of years around years multiples of 5 in case of missing data.[22] The panel data gather observation for 1,064 subnational regions from 72 countries.

[15]Visit the Statoids website to find the database - Gwillim Law. Administrative Divisions of Countries. McFarland & Company, Jefferson, North Carolina, October 1999.

[16]Visit the website of the author to find the databases.

[17]The transnational trade panel data comes from BACI.

[18]Visit the WPI website to find the database.

[19]WPI has classified ports' size in four groups according to their total area, wharf space and facilities, i.e. large, medium, small and very small. I exclude the very small ports from the dataset.

[20]Frensch et al. (2023) found that in the case of Europe, international rivers have a modest impact on trade.

[21]The original dataset from Gennaioli et al. (2014) contains different levels of definition of administrative regions for 13 countries compared to the first one from Gennaioli et al. (2013). In particular, the second dataset contains smaller regions. Thus, an effort has been made to combine these two datasets. To do so, I identify and gather regions defined at a lower administrative scale in a larger region, as defined in the regional Gennaioli et al. (2013) data. Among them, I sum the GDP, the population size and the production of oil, and I compute the average amount of education years and the average temperature.

[22]For example, if a country has observations for the year 1994-2001-2005-2010, they are reported to the years 1995-2000-2005-2010 so to have homogeneous time periods between observations.

5 Descriptive statistics

5.1 Regional Income Disparities

This section presents the state of regional inequalities within countries that can be observed in the data. Figure 1 displays the Gini coefficient which compares 2005's regional GDP per capita within each country in the sample. The coefficient ranges from 0 to 1, with zero-value referring to completely equal distribution of income per capita across regions, and unit-value referring to the extreme situation of one region holding the total national income, and the rest having no income at all. In other words, the greater the Gini coefficient, the greater development disparities in terms of income per capita distribution across regions. On the map, the darker a country, the greater its territorial inequality.

Kenya exhibits the greatest disparity in regional income per capita, with a Gini index of 0.45, followed by 11 other countries, including Indonesia, Thailand, South Africa, Argentina, Panama, Mozambique, Iran, Russia, Peru, China and Vietnam, all located mainly in South America and Asia. These 12 countries have a Gini index of over 0.3. Conversely, Egypt, Azerbaijan, Pakistan, Syria, France, Malawi and Israel show the least disparity between regions, with a Gini index of less than 0.07. This result is partly due to the fact that some countries have few recognized regions, with Egypt, Israel, Malawi and Pakistan having only 2 to 6 regions in countries with areas ranging from 20,000 to 1 million square kilometers.

The calculations here display different results than usual analysis of inequality of income, since the focus is given on territorial income per capita inequality rather than household income inequality. Thus, the map in Figure 1 is different than the one available on the World Bank website. Their Gini index ranges from 0.25 to 0.65 in 2005, suggesting that household income inequality is, in general, higher than regional income per capita inequality within countries. The most inequal countries according to their calculations were South Africa and Latin American countries, while the most equal countries were in Western Europe. However, we find similar Gini coefficients for Asian countries, around 0.4.

Figure 2 depicts the disparity in development between regions within countries, indicated by the ratio of maximum to minimum income per capita. Typically, countries with high Gini indices have a larger ratio of maximum to minimum income per capita. However, the relationship between the two measures of inequality is non-linear, despite a significant correlation of approximately 0.71. Consequently, the ranking of inequality shifted.

Russia has the greatest level of inequality, with its richest region, Tyumen, being 43 times wealthier than the poorest region, Republic of Ingushetia. This disparity is not surprising, as the country's economic activity is concentrated in a few regions, while others have lower levels due to geography and climate. Indonesia and Kenya follow with a ratio of 20. Argentina, Iran, the Democratic Republic of the Congo, Peru, India, Colombia, and China follow with ratios decreasing towards 9. The least unequal countries have ratios between 1 and 1.5, except for France where the wealthiest region is twice as wealthy as the poorest in terms of GDP per capita. While both measures of inequality are imperfect, they provide insight into existing regional disparities within countries.

5.2 Regional Market Potential Disparities

As there are great disparities in regional income per capita, there may be so in the regional market potential. Figure 4 displays the maximum to lowest market potential ratio within countries, considering the index $\mathrm{MP}^{(s)}$, which includes the shortest path distance by land and sea between

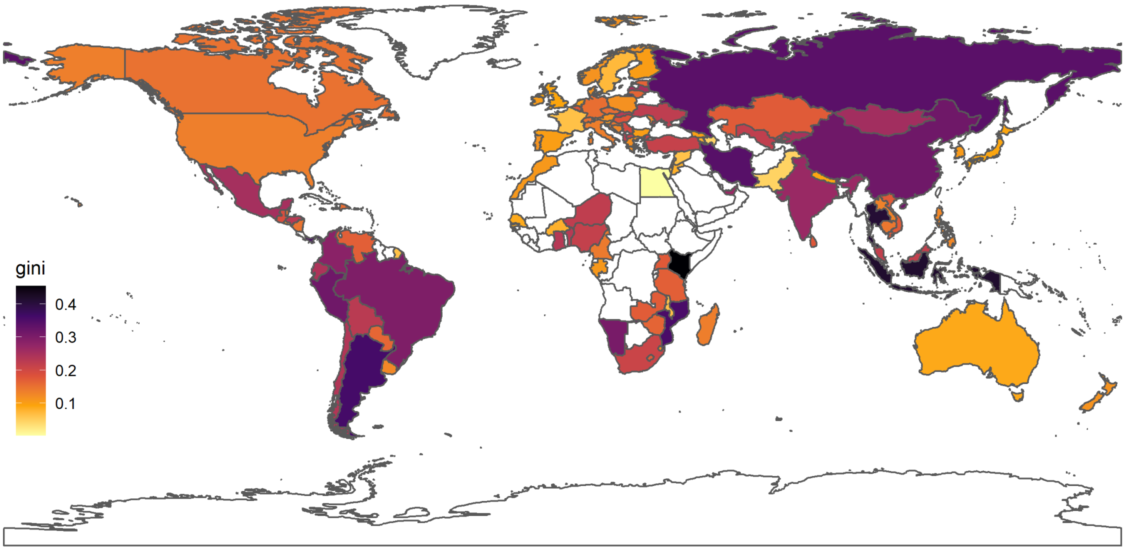

Figure 1: Gini index with respect to regional income per capita in 2005

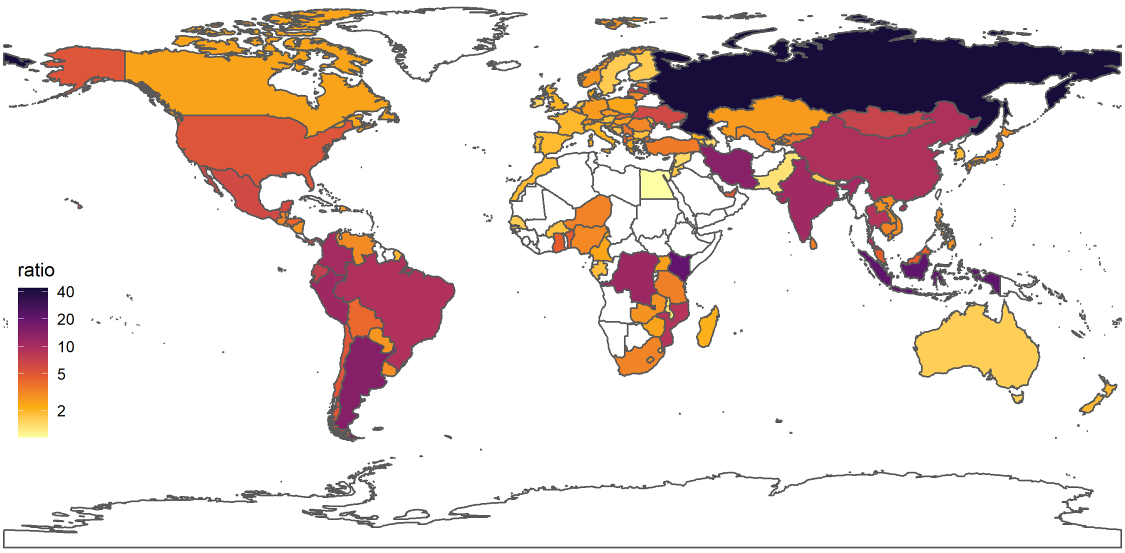

Figure 2: Ratio highest/lowest regional income per capita in 2005

regions. The income per capita is significantly correlated with the market potential ratios, with a correlation coefficient about 0.7. In other words, higher disparities in regional development should go in hand with higher disparities in trade opportunities.

Russia has the highest ratio of 90, followed by Argentina with a ratio of 32, and Japan with a ratio of 20. These countries have even greater disparities in terms of market potential than in terms of GDP per capita. In Europe, while regional development disparities appear higher in the East than the West, the opposite trend is observed for market potential. Higher ratios of highest-to-lowest market potential indicate significant spatial agglomeration. The wealthiest regions have wealthier neighboring regions, while the poorest regions are located further away. A high ratio indicates strong wealth agglomeration over space. However, a high ratio can also be due to larger regional area delimitation, resulting in greater distances between regions. This explains the significant disparities between regions in Australia and the United States (where regions are defined as states). African countries display particularly low differences in market potential due to similar foreign market potential indices. Their distance from the world's wealthiest regions is similar.

Table 14 presents the descriptive statistics of the variables used in the analysis for the whole 2005's sample, as well as by country income group following the World Bank classification: *High income*, *Upper middle income*, *Lower middle income* and *Low income*. The table indicates that market potential indexes using different distance functions are similar. Additionally, the richer the country, the higher the average regional GDP per capita, education level and market potential indexes. Standard deviations of GDP per capita and market potential are equal in the overall sample, but among high and upper-middle-income countries, the standard deviation of GDP per capita is lower than that of market potential. The opposite is observed in low-income countries. Notably, higher-income groups exhibit greater standard deviations in market potential.

5.3 Core and Periphery Divide

To examine regional heterogeneity and economic agglomeration, I categorize each country's regions into three groups based on their Gross Domestic Product: (1) core regions, the wealthiest regions, (2) periphery, the poorest regions, and (3) semi-periphery, those in between. I use k-means clustering to assign regions to their respective group. Note that countries have varying numbers of regions in each group. For countries with fewer than three regions, such as Ireland, no cluster classification is provided.

Statistics on these clusters are presented in table C. The core cluster counts between 1 and 6 regions, i.e. between 2% and 38% of the total amount of regions within countries. The semi-periphery counts between 1 and 12 regions, i.e. 3% and 67% in the total amount of regions within countries. Finally, the periphery counts between 1 and 66 regions in each country, which represents from 17% to 93% of the whole national territory. Therefore, the world counts much more peripheral regions than core regions. On average, the GDP per capita is higher in core regions compared to semi-peripheral and peripheral regions. This observation also holds true for population density and years of education.

To observe regional disparities of income between core and peripheral regions within countries, I run regressions of income per capita on the different clusters' dummies, including country fixed effects. Table 1 show the results. The average income per capita in semi-peripheral and peripheral regions are found to be respectively 39% and 62% lower than that of core regions. When investigating differences between core and peripheral regions in the different countries' income groups, results show that the core-periphery divide is greater for upper-middle and low income countries

	(1)	(2)
$\mathbb{1}_{\text{semi-periphery}}$	-0.39^{***}	-0.33^{***}
	(0.04)	(0.05)
$\mathbb{1}_{\text{periphery}}$	-0.62^{***}	-0.43^{***}
	(0.05)	(0.06)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{Upper-middle income country}}$		-0.06
		(0.09)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{Lower-middle income country}}$		-0.01
		(0.09)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{Low income country}}$		-0.32^{**}
		(0.14)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{Upper-middle income country}}$		-0.32^{***}
		(0.11)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{Lower-middle income country}}$		-0.17
		(0.12)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{Low income country}}$		-0.45^{**}
		(0.18)
Num. obs.	1516	1516
Num. groups: code	104	104
Adj. R^2 (full model)	0.93	0.93
Adj. R^2 (proj model)	0.23	0.25

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The dependant variable is the regional income per capita (in log). Robust standard errors adjusted for clustering on each country are in parentheses. Regressions include country fixed effects.

Table 1: Regional Development, the Core and Periphery Divide

than high-income and low-middle income countries.

Tables 16 and 17 in the appendix show the results using education, density and market potential indexes as dependant variables. (Semi-)peripheral regions are shown to be less educated, less inhabited and further from markets than core regions. Additionally, results show that the lower the income group of countries, the higher the disparities in education levels between core and peripheral regions.

Now that we have showed evidence of the existence of strong disparities within countries, it is also possible to study their evolution using panel data. Figure 3 shows the average growth in the Gini coefficient within countries from 1995 to 2010. We observe that 38% of the countries saw the dispersion of income per capita to decrease, with a Gini coefficient average growth ranging from 0 to -10% (from yellow to green areas on the map), while the others saw their disparities between regions to increase, with a Gini coefficient average growth ranging from 0 to 30% (from yellow to red areas on the map). The observed increase in regional disparities suggests a growing tendency towards divergence in the levels of development between different regions.

To investigate differences in development trends experienced by core and peripheral regions, I regress the regional income per capita on the year and interacts it with dummies for semi-peripheral and peripheral regions with country fixed effects. Results are displayed in table 2. Conditional on countries' unobservables characteristics, findings reveal that core regions witness annual average growth rate in income per capita about $(\exp^{0.0254} - 1) \times 100 = 2.57$ % on average, while the growth rate experienced by (semi-)peripheral regions is 0.02 percentage points lower. Adding interactions

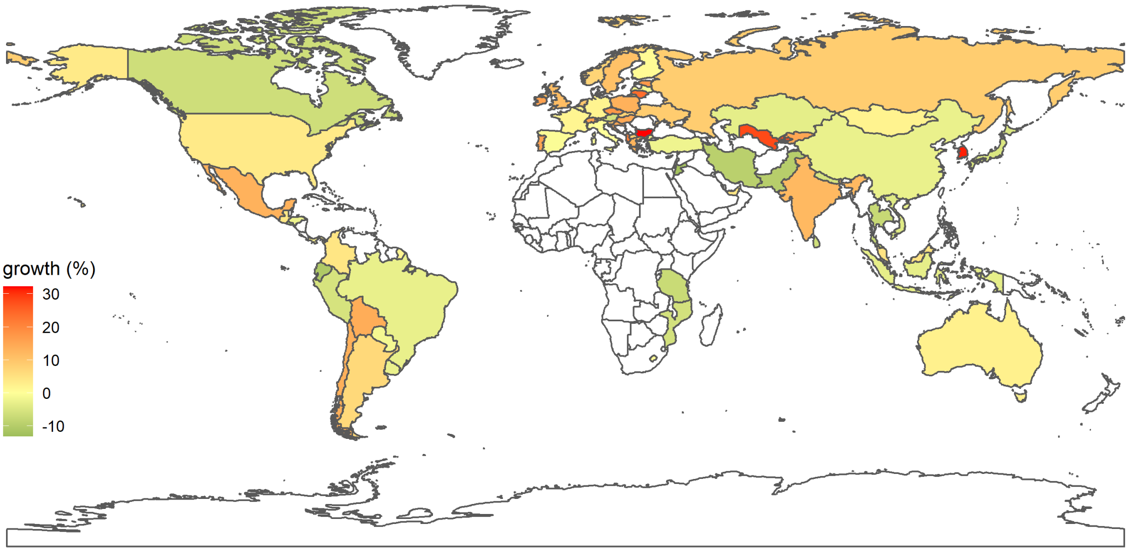

Figure 3: Gini coefficient average growth from 1995 to 2010

terms with dummies of countries' income group, it is found that the lower the countries' income, the higher the growth rate of their core regions, but also the lower the growth rate of their peripheral regions. These results suggest a possible convergence among core regions worldwide, but also an increasing divergence in development levels between core and periphery regions within countries. Wealthier regions experience higher growth than poorer regions, which widens the difference in their levels of development.

This increase in regional disparities between core and peripheral regions raises serious equity concerns. While regional disparities can be natural to a certain extent, because of different natural conditions and resources' availability, the rise to persistent and wider disparities highlight a concentration process of economic activity, leaving peripheral areas behind. Indeed, markets favor some places over others. Favored places and firms located there may gain in terms of productivity from the benefits of agglomeration economies, and further grow over time. On the contrary, remote and low productive areas become further remote and less productive, under-utilizing their potential and experiencing retrained growth. This paper aims at testing this assumption, using the market potential indexes and other regional development covariates.

Next section presents the empirical model for the study of subnational development determinants used as a falsification test for the international trade wage equation at the regional level within countries.

6 Empirical Model

The effect of market potential on wages can be estimated by transforming the wage equation in the logarithm form. The coefficient estimated will be interpreted as an elasticity. In other words, I am looking for the percentage change in wages according to a 1% change in market potential. Following Redding and Venables (2004) and Head and Mayer (2011), who empirically tested the international trade wage equation at the country level, income per capita is used as a proxy for

	Model 1	Model 2
year	0.0254***	0.0179***
	(0.0029)	(0.0035)
year $\times \mathbb{1}_{\text{semi-periphery}}$	−0.0002***	−0.0001***
	(0.0000)	(0.0000)
year $\times \mathbb{1}_{\text{periphery}}$	−0.0002***	−0.0002***
	(0.0000)	(0.0000)
year $\times \mathbb{1}_{\text{upper-middle income country}}$		0.0090*
		(0.0047)
year $\times \mathbb{1}_{\text{lower-middle income country}}$		0.0099
		(0.0094)
year $\times \mathbb{1}_{\text{low income country}}$		0.0288***
		(0.0062)
year $\times \mathbb{1}_{\text{upper-middle income country}} \times \mathbb{1}_{\text{semi-periphery}}$		−0.0000
		(0.0001)
year $\times \mathbb{1}_{\text{upper-middle income country}} \times \mathbb{1}_{\text{periphery}}$		−0.0001**
		(0.0001)
year $\times \mathbb{1}_{\text{lower-middle income country}} \times \mathbb{1}_{\text{semi-periphery}}$		−0.0001
		(0.0001)
year $\times \mathbb{1}_{\text{lower-middle income country}} \times \mathbb{1}_{\text{periphery}}$		−0.0001
		(0.0001)
year $\times \mathbb{1}_{\text{low income country}} \times \mathbb{1}_{\text{semi-periphery}}$		−0.0003***
		(0.0001)
year $\times \mathbb{1}_{\text{low income country}} \times \mathbb{1}_{\text{periphery}}$		−0.0003***
		(0.0001)
Num. obs.	3566	2946
Num. groups: code	58	49
Adj. R^2 (full model)	0.8985	0.8837
Adj. R^2 (proj model)	0.2204	0.2350

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The dependant variable is the GDP per capita (in log). Robust standard errors adjusted for clustering on each country are in parentheses. Estimations include country fixed-effects.

Table 2: Regional Development, Core and Periphery Divide

wages. Following this strategy, I can enlarge the interpretation of the usual wage equation and read the estimated coefficient of market potential as the elasticity of regional development with respect to market potential.

Assessing development by income per capita is largely adopted in the literature and allows to overcome the lack of available data regarding wages at the regional level worldwide, even if it is generally reasonable to criticise this strategy. Indeed, GDP per capita is not the only measure of development. There exist a large variety of indexes taking in consideration level of education, institutions, health care, etc. However, GDP per capita has a general impact on all of these variables. Indeed, richer countries tend to display higher educational and health care attainments, which essentially participate to the welfare of societies. Thus, GDP per capita may be reasonable to consider as a proxy for development.[23] This paper is therefore moving away from the traditional

[23]Further research can be conducted using other proxies for development. A difficulty in doing so may be in the

international trade wage equation and attempts to study what determines regional development disparities.

6.1 Baseline estimation

To conduct the analysis, the regional income per capita is regressed on market potential and covariates that are known to have a significant impact on regional development within countries. The covariates are geographical variables and human capital proxies, presented in section **??**. I expect the errors terms of GDP per capita regressions to be correlated with some specific group effect. Each region in a single group may be correlated but independent across groups, i.e. $Cov(u_{c_i,i}, u_{c_i,j}) = \zeta_{ij}^{c_i}$ and $Cov(u_{c_i,i}, u_{c_j,j}) = 0$ for any region $i \neq j$ belonging to countries c_i and c_j.[24] Country fixed effects are included in order to control for the long-run and historical factors that have shaped the economic development of countries and their institutions, aiming at reducing the omitted variable bias.

Unlike Gennaioli et al. (2013), I do not use their institutional quality index at the regional level because it is not found to affect GDP per capita and is limited to too few regions. I assume that institutional quality is uniform within a country and accounted for by using country fixed effects. Using country fixed effects eliminates the between-country variation by demeaning the data within countries. Hence, the results will be interpreted as average effects with respect to countries' average. The following regression is estimated:

$$\ln \mathrm{GDPpc}_i = \alpha_1 \ln \mathrm{MP}_i^{(h,s)} + \sum_{k=1}^{5} \delta_k X_i^{(k)} + \zeta_c + u_i \tag{6}$$

where α_1 is the coefficient of interest, $\mathrm{MP}_i^{(h,s)}$ the market potential index computed either with the haversine distance $\mathrm{dist}_{ij}^{(h)}$, either with the shipment distance function $\mathrm{dist}_{ij}^{(s)}$ defined in equation 4, ζ_c the country fixed-effect. Component $X_i^{(k)}$ represents the GDP per capita covariates, with k referring to each of the five following variables: temperature, inverse distance to port, oil production per capita, average years of education, and population density.

6.2 Robustness

6.2.1 Proxies of market potential

The market potential variables are suspected to be endogenous since their local component is expressed as the local GDP_i. It induces a problem of reverse causality between the regressor and the regressand since income and market potential are simultaneously determined. It possibly creates biased OLS coefficients for market potential. To first solve for this problem, I replace the market potential by the non-local market potential, i.e. NLMP_i.[25] Doing so deletes the endogenous local part of the market potential.

absence of a database with as many regional observations worldwide. For the same reason, Gross Value Added (GVA) per capita is not used as a proxy for wages. This is not a serious limitation as studies that use GVA per capita (Breinlich, 2006; Bruna et al., 2016) find similar results than those using GDP per capita (Head and Mayer, 2006) in the case of Europe. Moreover, in some countries, wages are set at the national level for many production sectors.

[24] The average number of observations per country is about 14.2, with the minimum about only 2 regions (Egypt and Ireland), and the maximum about 80 regions (Russia).

[25] Non-local market potential is also often referred as external market potential (Harris, 1954).

However, proxying the market potential MP_i by the non-local market potential NLMP_i may not be sufficient to erase the problem of endogeneity, and the coefficients of non-local market potential indexes could still be biased. As the first Law of Geography by Tobler (1969) states: "Everything is related to everything else, but near things are more related than distant things". To further delete the near information in market potential indexes, market potential indexes are also proxied by the foreign market potential FMP_i. It deletes information on domestic regions. The correlation between market potential and foreign market potential indexes are about 0.55 and 0.64 respectively for the indexes including distance measures expressed as $\text{dist}_{ij}^{(h)}$ and $\text{dist}_{ij}^{(s)}$.

The non-local market potential $\text{NLMP}^{(h)}$ represents about 96% (98% for the $\text{NLMP}^{(s)}$ index) of the market potential on average, while the foreign market potential, about 65% (69% resp.). Thus, the omission of the local and domestic part of the market potential is substantial. Regressing the GDP per capita on market potential and covariates, and each of the variables NLMP_i and FMP_i allow us to test whether proxies are good if they are redundant, i.e. unsignificant as the proxied variable is present in the regression. The foreign market potential is found to be a better proxy for the market potential index, as its elasticity coefficients are unsignificant.

6.2.2 2SLS estimations

We saw that the local market potential represents a substantial part of the market potential. Erasing information on the local market may not be the best solution to solve for endogeneity as it induces a problem of omitted variable. There may also be the problem of spatial endogeneity. In order to better solve the endogeneity problem and to keep the local and domestic information of the market potential at the same time, Two-Stage-Least-Squares (2SLS) estimations are conducted.

Redding and Venables (2004) chose the log distance to the three biggest international markets as instruments for their market access index, which are Japan, Western Europe, with Belgium at its center, and the USA. Head and Mayer (2011) chose to use a centrality index, which is computed as the sum of the distances between every countries. The higher their centrality measure, the closer a country to all the other markets regardless of their market size, and the lower its transportation costs.

Using this latter IV strategy to regions, the instruments are found not to be exogenous. The arising problem comes from the fact that the size of regions has a direct impact on proximity to other markets, as well as on GDP per capita. A smaller region in terms of land area tend to be closer to other markets, since its centroid is naturally closer to others than a vaster region. Thus, its centrality index will be higher, since the sum of the inverse of distances is naturally higher. However, administrative regions that have been designed to be smaller in terms of land area are often the regions with the highest level of GDP and GDP per capita. Centrality is a good instrument for market potential considering countries, but it is not when considering sub-national regions, which are defined for administrative and political purposes.

The IV strategy I am using in this paper is halfway between Head and Mayer's and Redding and Venables'. I define centrality_i, the harmonic centrality of each region with respect to the richest market of each country in the sample. The index provides information on the proximity to these markets such as the higher the index, the closer the region to the richest national markets. It may help better capturing the regional heterogeneity in proximity to rich regional markets. The instrument is computed as follows:

$$\text{centrality}_i = \sum_{j(r)} \frac{1}{\text{dist}_{ij(r)}^{(h,s)}} \tag{7}$$

with r the richest region in terms of GDP of country j, and $\text{dist}_{ij(r)}^{(h,s)}$ the distance between i and r for each country j, computed as the haversine distance and as the shipment distance function defined in equation 4. More particularly, $\text{centrality}_i^{(h)}$ is going to depend on $\text{dist}_{ij(r)}^{(h)}$, and $\text{centrality}_i^{(s)}$ on $\text{dist}_{ij(r)}^{(s)}$. The first stage regression is expressed as follows:

$$\ln \hat{\text{MP}}_i^{(h,s)} = \lambda_1 \ln \text{centrality}_i^{(h,s)} + \sum_{k=2}^{6} \lambda_k X_i^{(k)} + \zeta_c + \epsilon_i \tag{8}$$

Another instrument used for market potential is the foreign market potential. Indeed, it is found to not have a significant impact on the regional income per capita, and it is redundant when including it in equation 6, i.e. its coefficient is not significant - see the results in section 7.2.2. Hence, another version of the first stage regression is:

$$\ln \hat{\text{MP}}_i^{(h,s)} = \theta_1 \ln \text{FMP}_i^{(h,s)} + \sum_{k=2}^{6} \lambda_k X_i^{(k)} + \zeta_c + \epsilon_i \tag{9}$$

6.2.3 Panel estimation

To test whether regional development is robustly explained by market potential, I estimate the model with the panel data, and interact the country fixed effects with years, and denote it as ζ_{ct}. Doing so allows to control for heterogeneity of countries at different points in time.[26] In particular, these country-year fixed effects allows to control for possible change in institutional quality over time at the country level. Coefficient α_1 gives the average effect of a regional market potential deviation from country-year average.

6.3 The Core and Periphery Divide

The heterogeneous effect of market potential on regional development is also investigated for different groups of regions identified within each country according to their GDP levels using k-means clustering. The resulting groups of regions are classified as core, semi-periphery and periphery regions. They reflect the results of agglomeration effects described by Krugman (1991). I estimate the following specification:

$$\ln \text{GDPpc}_i = \alpha_1 \ln \text{MP}_i^{(h,s)} + \sum_{g=2}^{3} \alpha_g \ln \text{MP}_i^{(h,s)} \times \mathbb{1}(\Gamma_i = \gamma_g) + \sum_{k=1}^{5} \delta_k X_i^{(k)} + \zeta_c + u_i \tag{10}$$

with Γ_i the classification of region i, with $\gamma_g = \{\textit{core, semi-periphery, periphery}\}$ the different values of Γ_i. α_1 is the average percentage change of a 1% increase in market potential index from the national average for a core region, ceteris paribus. α_2 and α_3 are the effect of the two interaction terms of the market potential variable and dummies, which are equal to 1 if region i is a semi-periphery or a periphery region respectively, zero otherwise. In other words, they represent the

[26] Note that the only source of temporal variation for the market potential indices is the GDP information. Literature has shown that ocean shipping costs are falling over time for all countries as improved technologies reduce port time and speed sea travel and larger containers transport larger volumes of commodities. In particular, Hummels (2007) reports that from the 80s to 2000s, shipping costs have been declining steadily. However, the shipment distance function used in the computation of market potential indexes do not vary in this analysis. This omission can be translated in a decrease in the elasticity coefficients to market potential over time. We do not find that the coefficient change (see results from the cross-sectional estimations for 1995, 2000 and 2005 in table 26).

difference in percentage points of the elasticity to market potential between (semi-) peripheral and core regions. In order to investigate potential heterogeneity effects accross countries, I split the sample into four groups based on the level of income of each country: high-income, upper-middle-income, lower-middle-income, and low-income countries. It allows to determine the different degrees of sensitivity to market access for core and periphery regions across different countries' income levels.

6.3.1 Centrality to core markets

Additionally, the study delves deeper into the impact of proximity to domestic and foreign core regions on (semi-)peripheral regions. In order to do this, other measures of centrality are used, which are based on the inverse haversine distance between a region and the national or foreign core regions. Note that based on the k-means clustering of regions according to their GDP levels, countries in the sample contain between one to three core regions. The inclusion of centrality measures in the regression analysis helps provide a clearer insight into the influence of proximity to affluent markets by addressing certain limitations associated with market potential indexes. These indexes tend to aggregate a multitude of values, leading to limited variation between regions within countries. By introducing centrality measures, a more precise understanding of the connection between market access to core domestic and foreign regions and regional development disparities between core and peripheral regions can be attained, thanks to the additional variation they introduce into the regression estimations.

I distinguish centrality to foreign core regions between centrality to foreign core with a free trade agreement and without. Being in close proximity to foreign cores without a trade agreement can impede trade due to national barriers such as tariffs, quotas, and regulations. These barriers restrict the movement of goods, services, and investments, increasing costs and hindering trade for peripheral regions. Despite their geographical closeness, the absence of a trade agreement limits the potential economic benefits that could arise from proximity to foreign cores, resulting in missed trade opportunities and reduced growth prospects.

I will contrast the findings with those of Adam et al. (2023) and Bonadio et al. (2023) who find that that subnational regions near international borders often exhibit lower per capita income levels and that trade agreements at these borders can mitigate this negative impact by facilitating trade and reducing barriers. Although the present papers' specification regarding proximity to foreign cores without a trade agreement may not directly align with Adam et al. (2023) and Bonadio et al. (2023)'s results, they provide additional insights into the compensatory role of trade agreements in mitigating the adverse consequences of geographical proximity to foreign cores.

The four centrality measures are:

$$\text{centrality}_i^{\text{domestic cores}} = \sum_j \frac{1}{\text{dist}_{ij}} \times \mathbb{1}(\Gamma_j = \text{core}) \times \mathbb{1}(c_i = c_j) \tag{11}$$

$$\text{centrality}_i^{\text{foreign cores}} = \sum_j \frac{1}{\text{dist}_{ij}} \times \mathbb{1}(\Gamma_j = \text{core}) \times \mathbb{1}(c_i \neq c_j) \tag{12}$$

$$\text{centrality}_i^{\text{foreign cores, FTA}} = \sum_j \frac{1}{\text{dist}_{ij}} \times \mathbb{1}(\Gamma_j = \text{core}) \times \mathbb{1}(c_i \neq c_j) \times \mathbb{1}(\text{rta}_{c_i,c_j} = 1) \tag{13}$$

$$\text{centrality}_i^{\text{foreign cores, no FTA}} = \sum_j \frac{1}{\text{dist}_{ij}} \times \mathbb{1}(\Gamma_j = \text{core}) \times \mathbb{1}(c_i = c_j) \times \mathbb{1}(\text{rta}_{c_i,c_j} = 0) \tag{14}$$

where $\mathbb{1}(\Gamma_j = \text{core})$ correspond to a dummy stating that region j is a core region in country c_j. $\mathbb{1}(c_i = c_j)$ refers to domestic relationship and indicates that countries of regions i and j are the same if the dummy equals 1, zero otherwise. $\mathbb{1}(c_i \neq c_j)$ refers to foreign relationship where countries of regions i and j are different. If they are different, the dummy equals 1, zero otherwise. Among those foreign relationships, a distinction is done between $\mathbb{1}(\text{rta}_{c_i,c_j} = 1)$ and $\mathbb{1}(\text{rta}_{c_i,c_j} = 0)$, which indicate whether or not the countries of regions i and j have a trade agreement in 2005.

I estimate the following specifications:

$$\ln \text{GDPpc}_i = \sum_{w=1}^{W} \alpha \ln \text{centrality}_i^{(w)} + \sum_{k=1}^{5} \delta_k X_i^{(k)} + \zeta_c + u_i \tag{15}$$

$$\ln \text{GDPpc}_i = \sum_{g=1}^{3} \sum_{w=1}^{W} \alpha_{wg} \ln \text{centrality}_i^{(w)} \times \mathbb{1}(\Gamma_i = \gamma_g) + \sum_{k=1}^{5} \delta_k X_i^{(k)} + \zeta_c + \gamma_g + u_i \tag{16}$$

with the first one investigating the effect of the centrality measures that contain different weights which indicate whether it relates to centrality to domestic or foreign cores, with and without free trade agreement. The second specification estimate the coefficients for each cluster group of regions, i.e. the core and the (semi-)periphery.

I conduct robustness checks using panel data to further assess the effects of centrality on foreign core regions that have a trade agreement with the region's country of reference. Specifically, I evaluate the impact of signing a trade agreement by computing the centrality index for foreign core regions with a trade agreement for each year of the panel sample (1995, 2000, 2005, 2010) and explore how variations in the index influence regional development. When a country signs a trade agreement with another, the centrality index of its regions increases as I add the sum of the inverse distances to the core regions of the second country. I estimate the following regressions:

$$\ln \text{GDPpc}_{it} = \alpha \ln \text{centrality}_{it}^{\text{foreign cores, FTA}} + \sum_{k=1}^{5} \delta_k X_{it}^{(k)} + \zeta_{ct} + u_{it} \tag{17}$$

$$\ln \text{GDPpc}_{it} = \sum_{g=1}^{3} \alpha_g \ln \text{centrality}_{it}^{\text{foreign cores, FTA}} \times \mathbb{1}(\Gamma_i = \gamma_g) + \sum_{k=1}^{5} \delta_k X_{it}^{(k)} + \zeta_{ct} + \zeta_i \times \gamma_g + u_{it} \tag{18}$$

where α is the average percentage effect of a 1% increase in the variable $\text{centrality}_{it}^{\text{foreign cores, FTA}}$ due to trade agreement with other countries, and α_g is the specific effect for each group of regions, the core and the (semi-)periphery. ζ_{ct} is a country-year fixed effect and $\zeta_i \times \gamma_g$ a region-cluster group fixed effect. Results are displayed and discussed in next section.

7 Results

In this section, evidence is presented on the determinants of regional development, first in cross-section in 2005, then in panel estimations for the years $t = \{1995; 2000; 2005; 2010\}$. A particular focus on the effect of market potential is given. Regressions allow to test the international trade wage equation and to compare coefficients with the structural parameter from the theory.[27]

[27]Coefficients may be equal to $1/\beta\sigma$, with σ the elasticity of substitution between varieties and β the income labor share. Following a methodology derived from a model of international trade in differentiated products and monopolistic competition, the elasticity of substitution between varieties has been estimated in the literature. At the

7.1 Baseline estimations

Univariate regressions of GDP per capita on market potential and its proxies are presented in table 18 in the appendix. Results show positive and highly significant coefficients for the estimations with the different distance function. In particular, if the average region in a country experience a 1% increase in its market potential, it is expected to increase its GDP per capita by 0.2%. The R-squared statistics indicate that variations in regional $\ln \text{GDPpc}$ are explained by between 4% and 6% of the variations in $\ln \text{MP}$ within countries. R-squared statistics are lower when considering the non-local market potential indexes, i.e. 0.02. The foreign market potential does not seem to explain regional development. $\text{MP}^{(h)}$ and $\text{MP}^{(s)}$ display similar coefficients among the different specifications.

It is important to note that the elasticity coefficient should theoretically equate to $1/(\beta\sigma)$, where β represents the income labor share, and σ signifies the elasticity of substitution between varieties. Regarding the income labor share, Reshef and Santoni (2023) conduct estimations and found that it ranged from 0.3 to 0.7, depending on the country, as of 2007.[28] Furthermore, Fontagné et al. (2022) estimate the elasticity coefficient of substitution between varieties at the product level and found values between 5 and 20. The expected α_0 coefficient that we need to find should fall between 0.07 and 0.7. The coefficients revealed in the present paper fall within the expected range, which is reassuring both for validating the theoretical framework and the worldwide falsification test.

Table 3 shows the baseline results in cross-section for the year 2005. The estimations are controlled for other geographical variables and human capital proxies, such as *temperature*, known to affect labor productivity, *inverse distance to the closest port*, which is used as another measure for the impact of international trade activities as it can proxy for foreign markets proximity, *production of oil per capita*, as it is a particular endowment that affect development in specific ways, and finally *average years of education* and *population density*, which both play a role as knowledge resources and production means. The covariates tend to reduce the biases of the univariate regressions.

Conditional on time-invariant characteristics of countries, the elasticity coefficient of regional development to market potential is found to be equal to 0.10, and 0.05 for the non-local market potential. In other words, a 1% higher market potential index is associated to a 0.1% higher GDP per capita. These coefficients correspond to high values of elasticity of substitution between varieties. Notice that the non-local market potential as measured by port accessibility and ocean shipment distance function give a relatively lower confidence interval, but the p-value is still about 0.057. The foreign market potential is not found to have a significant impact on the GDP per capita differences within countries. It is also the case when we exclude *inverse distance to the closest port* from the regression. The lack of significance in the foreign market potential effect might be attributed to the limited variation captured by the index within countries, as indicated by the regression results of the index against (semi-)periphery dummy variables - refer to Table 17.

Despite the non-significance of the foreign market potential coefficient, accessibility to foreign markets as proxied by the inverse distance to ports is found to be a highly significant determinant of GDP per capita. The coefficient is robustly estimated to be equal to 0.14. The oil production per capita and the average years of education are also found to explain regional differences in devel-

aggregate level, it has been found to be around 6 and 11 - see Head and Ries (2001), Lai and Trefler (2002), Romalis (2007). Depending on the industry field, Ahmad and Riker (2020) find that the elasticity of substitution between varieties can range from 1 to 13, with the median around 2 and 3. On the other hand, the income labor share is generally estimated to be between 0.5 and 0.75 - see Guerriero (2019). Assuming these ranges for σ and β, we can expect the coefficient of interest of the analysis to be in the interval $[0.10; 0.33]$, or even in the interval $[0.44; 2[$ for values in the lower bound of σ.

[28]For detailed country-level labor shares in 1995, 2007, and 2014, refer to Table A2 in their appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
market potential	0.11^{***}	0.09^{***}	0.06^{**}	0.05^{*}	0.07	0.00
	(0.03)	(0.03)	(0.02)	(0.03)	(0.07)	(0.09)
inv. dist. port	0.14^{**}	0.13^{**}	0.14^{**}	0.13^{**}	0.14^{**}	0.13^{**}
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
years education	0.28^{***}	0.28^{***}	0.28^{***}	0.28^{***}	0.28^{***}	0.28^{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
population density	-0.00	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
temperature	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
oil per cap.	0.19^{***}	0.19^{***}	0.19^{***}	0.19^{***}	0.19^{***}	0.19^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Num. obs.	1,464	1,464	1,464	1,464	1,464	1,464
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Num. groups: code	103	103	103	103	103	103
Adj. R^2 (proj model)	0.43	0.42	0.42	0.42	0.42	0.42
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 3: Regional Development and Market Potential

opment levels within countries. In particular, an additional year in the average level of education in a region with respect to the national average is found to increase the GDP per capita by 28%. This coefficient is lower than the one obtained by Gennaioli et al. (2013) (see table IV, page 134).

It is worth to notice that the effect of education on regional development keeps being much larger than the micro estimates, which are generally of the 0.06-0.10 range from the Mincerian literature. It can be due to the fact that the educational variable is endogenous[29]. According to Gennaioli et al. (2013), one way to proxy this variable is by considering the education level of the elderly population, specifically those aged 65 years and older. While this proxy may not be ideal due to potential long-term growth effects, the authors argue that it may suffice since they do not directly intervene in the current production function. Using this proxy yields coefficients that closely align with those obtained using other human capital proxies and enhances the coefficients associated with the various market potential indexes. However, it has the drawback of significantly reducing the number of observations. The results are presented in Table 19.

I compare the results with the ones found by Head and Mayer (2011). They use measures of market potential considering countries, with a similar formulation than $MP^{(h)}$ index. Their index takes into account distance, contiguity, common language, colonial links, and dummies for common

[29] Acemoglu et al. (2014) instrumented the educational variable by the allocation of protestant missionaries in the early twentieth century

membership of a regional trade agreement (RTA), a currency union and WTO membership. Net of the effect of education levels, they find significant coefficients about 0.41 for the year 1995 without country fixed effects, and 0.55 for the period 1965-2005 with country fixed effects. This is sensitively higher than what we find in the similar regressions considering regions in 2005. It suggests that regional development differences within countries are less sensitive to market potential than national development differences worldwide.[30] This observation can be due to the fact that there exist less variation of GDP per capita and market potential within countries than between countries.

Single-country analysis support this statement. The effect of a 1% increase in the regional market potential is found to increase wages by below than 0.2% in general. This results hold for developing countries such as Brazil (Fally et al., 2010), China (Baum-Snow et al., 2020), Chile (Paredes, 2013), as well as European regions from European analysis (Niebuhr, 2006; Head and Mayer, 2006; Brakman et al., 2009), and for developed countries such as Germany (Brakman et al., 2004; Kosfeld and Eckey, 2010), Italy (Mion and Naticchioni, 2009), and Spain (Pires, 2006).

7.2 Robustness checks

This section presents the robustness checks undergone to verify the validity of the results first within each country, second using Two-Stages-Least-Square estimations on the cross-sectional sample, and finally on a panel, but smaller, sample.

7.2.1 Countries' regressions

Tables 20, 21, 22 and 23 in the appendix display the coefficients per country. These coefficients have been estimated following equation 6 for each country. Most of the estimated coefficients are found insignificant. Some coefficients of market potential elasticity are even found to be negative - in the case of Portugal (7 observations) or Czech Republic (8 observations). This may be due to the level of aggregation of the data. Regions may be too large and too few per country - we report the amount of regions per country in the table.[31] Hence, comparing both observations in the within country analysis gives a non-meaningful and highly negative coefficient. However, it could also testify the strong agglomeration forces within countries, where agglomeration of economic activity is stronger. Rich and large regions might be harmful for surrounding regions, since their mass may attract their population and investments.

We find similar coefficients of market potential among countries that have been studied in the literature. Among high income countries, Italy's coefficient is about 0.4, same as found by A'Hearn and Venables (2013), as showed by Table 20. In particular, results show that Italy is highly sensitive to its foreign market potential, considering the simple haversine distance function as well as the shipment distance function. Spain's coefficient is found to be equal to 0.1, although not significant, with Pires (2006) and López-Rodrıéguez et al. (2008) finding coefficients between 0.1 and 0.2. This range of value are the ones found for European countries in general (Niebuhr, 2006; Head and Mayer, 2006; Brakman et al., 2009). I find that Denmark and Hungary display significant coefficients about 0.5 for $\text{MP}^{(h)}$ and 0.7 for $\text{MP}^{(s)}$.

Among upper middle income countries, China displays a coefficient about 0.7, as displayed in Table 21. This elasticity coefficient is higher than the 0.1 coefficient found by Hering and Poncet (2010) and Baum-Snow et al. (2020). Brazil and Turkey display a similar coefficient than Fally et al.

[30] Another explanation for smaller coefficients is that I control for more geographic variable than Head and Mayer (2011) do, reducing the potential bias in the estimators. With panel data and controlling for auto-correlation, Boulhol and De Serres (2010) find coefficients between 0.07 and 0.1, as I find in table 3.

[31] Note that significant coefficients are associated to countries with at least 20 observations.

(2010) and Karahasan et al. (2016) respectively, but a negative coefficient for Chile, although not significant, contrary to what is found by Paredes (2013). In addition, I find significant coefficients about 0.6-0.9 for Columbia and 0.3 for Russia. In both countries, regions are found to be very sensitive to their foreign market access, in particular to the shipment foreign market access for Columbia - the elasticity is estimated about 4.7. Among lower middle income countries, only Indonesia's and India's regions are found to depend on their proximity to markets - as measured with both haversine and shipment distance functions. Their coefficient elasticity are respectively about 0.9 and 0.2. The lack of information for low income countries do not allow me to conclude.

7.2.2 2SLS estimations

Two different instruments are used in the Two-Stage-Least Squares estimation. A first one is a centrality index which summarizes the inverse distance to the wealthiest region of each country in the sample, presented in equation 7. The second is the foreign market potential, which is not fount to have an effect on GDP per capita, as showed by results above.

First, market potential indexes are regressed on the instrument chosen and the regional development covariates, following equation 8. Then, the second stage consists in regressing the GDP per capita on the fitted market potential values, as well as the regional development covariates. Column 1 and 2 in table 4 show the results of the first and second stage equations, i.e. equations 8 and 6 respectively. Column 3 and 4 show the results for specifications with the first stage expressed in equation 9, using the foreign market potential as instrument.

First stage results show that the higher the centrality of regions in terms of proximity to the wealthier regions from each country, the higher their market potential. More particularly, I find that a region with a 1% higher centrality index than the average region in the same country is predicted to have a 0.08% higher market potential. This result is significant at the 1% significance level. Density is found to be an important determinant of market potential. The larger the population with respect to the regional area, the higher the potential demand and thus, the higher the market potential. Second stage display considerably higher coefficients for market potential indexes than what has been estimated with OLS, suggesting that the effect of regional market potential on development has been under-estimated. Indeed, coefficients are between 0.8 and 0.9 points greater.

In order to check for the instrumentation's relevance, I conduct different tests: (1) endogeneity test, (2) under-identification test[32], and (3) weak identification test. Because it is assumed that the error terms may be correlated at the national level, justifying the use of fixed effects and standard errors clustering at the country level, a first test relies on the Durbin–Wu–Hausman statistic, a second on the Kleibergen-Paap rk LM statistic, and a third on the Kleibergen-Paap rk Wald F statistic. This three tests are robust to heteroskedasticity.

The first test evaluates whether the OLS and the 2SLS estimates are statistically different, i.e. $H_0 : \alpha_1^{(OLS)} \neq \alpha_1^{(2SLS)}$. In other words, it checks for the endogeneity of the market potential variable. The Durbin–Wu–Hausman statistics in column 1 and 2 suggest that the two estimates from the OLS and the IV regressions are statistically different, so that the OLS estimate is inconsistent. However, the second and third IV tests reveal statistics lower than 10[33]. Thus, we can not reject that the instrument is weak, neither that it is redundant. In other words, the instrument seems not to be relevant and the estimated coefficient may be biased.

[32] With one instrument, the underidenfication test is equivalent to the IV redundancy test.

[33] The Kleibergen-Paap rk Wald F statistic from the weak identification test is compared to the Stock-Yogo weak ID test critical values for a 10% maximal IV size, which is about 16.38. Hence, the null hypothesis is not rejected.

	(1)	(2)	(3)	(4)
Second Stage				
temperature	−0.01	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)
inv. dist. port	0.18	0.14**	0.14**	0.13**
	(0.14)	(0.09)	(0.06)	(0.06)
oil per cap.	0.24***	0.23***	0.19***	0.19***
	(0.04)	(0.04)	(0.04)	(0.04)
population density	−0.13**	−0.08**	0.00	0.01
	(0.06)	(0.04)	(0.01)	(0.01)
years education	0.26***	0.27***	0.28***	0.28***
	(0.03)	(0.03)	(0.02)	(0.02)
market potential	1.04**	0.87**	0.08	0.00
	(0.51)	(0.42)	(0.08)	(0.09)
Num. obs.	1,464	1,464	1,464	1,464
Country Fixed Effects	Yes	Yes	Yes	Yes
Num. groups: country	103	103	103	103
Adj. R^2 (proj model)	−0.32	0.06	0.43	0.42
Regressor (endogenous variable)	$\ln \mathrm{MP}^{(h)}$	$\ln \mathrm{MP}^{(s)}$	$\ln \mathrm{MP}^{(h)}$	$\ln \mathrm{MP}^{(s)}$
First Stage				
temperature	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
inv. dist. port	−0.04	−0.00	0.00	−0.08
	(0.14)	(0.11)	(0.15)	(0.13)
oil per cap.	−0.05	−0.06*	−0.02	−0.02
	(0.04)	(0.03)	(0.04)	(0.03)
population density	0.12***	0.09***	0.12***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)
years education	0.00	−0.01	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.01)
centrality	0.05**	0.07**		
	(0.03)	(0.03)		
foreign market potential			0.98***	1.02***
			(0.13)	(0.16)
Endogeneity test				
F stat	16.35	13.89	0.23	1.11
	[0.00]	[0.00]	[0.63]	[0.29]
Under-identification test				
Kleibergen-Paap rk LM stat	4.23	5.23	11.12	10.71
	[0.04]	[0.02]	[0.00]	[0.00]
Weak identification test				
Kleibergen-Paap rk Wald F stat	4.07	5.18	55.29	38.77

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. $\mathrm{MP}^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $\mathrm{MP}^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. The variables centrality$^{(h)}$ and $\mathrm{FMP}^{(h)}$ (centrality$^{(s)}$ and $\mathrm{FMP}^{(s)}$ respectively), are used as instruments for $\mathrm{MP}^{(h)}$ ($\mathrm{MP}^{(s)}$ respectively).

Table 4: IV results

Since the OLS coefficients for the effect of foreign market potential indexes on regional GDP per capita are not significant, I choose to use them as instruments[34]. The first stage shows highly significant elasticity coefficients around unity, with Kleibergen-Paap statistics higher than 10 and 16.38 for the under-identification and weak identification tests respectively. However, the endogeneity test reveals that the effect of the fitted market potential is not different from the OLS specification, with non-significant IV coefficients. This result suggests to believe in the coefficients about 0.1 estimated in column 1 and 2 from table 3. However, 2SLS coefficients are found to be not significant. It could come from the fact that market potential impact regional inequality differently depending on the income group of the country, as explained in section 7.3.2.

Head and Mayer (2006), who estimate market potential indexes for European regions with a similar method, find coefficients about 0.12 in their OLS estimation, and about 0.07 in their IV estimation. These coefficients are similar to those found from the OLS specification - see Table 3.

7.2.3 Panel estimations

To test the robustness of the cross-sectional results, we estimate regression 6 with the panel data available for the years 1995, 2000, 2005 and 2010. The panel sample comprises 1,064 subnational regions in 72 countries, although the dataset is unbalanced. On average, each region appears in the sample for 2.9 years, with a minimum of one year and a maximum of four years. However, due to a lack of information on education for some regions during the years 2005 and 2010, a significant number of observations will need to be removed from the estimations. To address this issue, I merge the panel dataset with the 2005 cross-sectional dataset to complete missing observations on education for that year. This will help to improve the completeness of the sample and ensure that the estimations are based on a more representative dataset. First, cross-section regressions with country fixed effets are estimated for each year of the panel sample. Then, panel estimations are performed with fixed effects at the country-year level, ζ_{ct}, to control for country-time varying unobservables. Results are displayed in table 26 in the appendix.

Elasticity coefficients to market potential from the different cross-sectional estimations are similar than those found in table 3 - about 0.1 - for the years 1995, 2000 and 2005. Foreign market potential elasticity coefficient is not significant in 2005, as found in table 3, but is significantly equal to 0.1 in 1995 and 0.2 in 2000. However, results for the year 2010 show unsignificant coefficients for market potential indices computed with the haversine distance, while indices computed with the shipment function are found to have a significant negative effect on income per capita. Results may be biased as a result of a lack of observations, i.e. only 20 countries are considered.

Results for the complete panel estimations support that a region with a market potential that is 1% higher than the average region within country-year groups is predicted to have a 0.1% higher income per capita. Coefficients display a higher significance level when excluding the year 2010. It is also found that proximity to foreign markets do matter even more significantly than the non-local market potential, while the reverse has been found in cross-section.

Table 27 in the appendix presents the results for the investigation of heterogenous effect of market potential regarding the core-periphery structure of regions within countries. Results still show that peripheral regions robustly display elasticity coefficients lower by 0.01 to 0.02 percentage points lower than core regions.

[34] Note that the foreign market potential is redundant when adding the variable in equation 6 to be estimated.

7.3 The Core and Periphery Divide

7.3.1 Baseline regressions

The heterogeneous effect of market potential on regional development is investigated with respect to the income position of regions within countries, classified into three different groups: the *core*, the *semi-periphery* and the *periphery*. This specification is motivated by the Krugman (1991)'s Core-Periphery model which explains how the economic activity can concentrates in specific regions, leaving the periphery behind. The lower the transportation and trade costs to a wealthy region, the higher the agglomeration forces towards this wealthy regions. The poorer a market, the higher the agglomeration forces towards wealthy regions. Following this statement, we can wonder if market potential can have a negative effect on development for peripheral regions.

Table 5 shows the results, with the interacted effect of market potential with the regional group dummies. The coefficient of market potential gives the effect of market potential on core regions - the wealthiest in terms of GDP. A 1% increase in the market potential in core regions is expected to increase their GDP per capita by 0.1%. If the region is categorized as a periphery region, the expected effect of market potential on regional development is 0.02 percentage points lower, but still positive. The same result holds for non-local market potential. While proximity to wealthy foreign regions is found to have no effect on core regions, (semi-)peripheral regions are found to have an elasticity coefficient significantly lower by (0.01) 0.02 percentage points. This result suggest that the (semi-)periphery could suffer from too much proximity to foreign markets. A 1% increase in the foreign market potential in peripheral regions can be associated to a decrease by 0.02% in the GDP per capita.

The elasticity coefficient should theoretically equate to $1/(\beta\sigma)$ as mentioned earlier, where β

	(1)	(2)	(3)	(4)	(5)	(6)
market potential	0.12^{***}	0.10^{***}	0.08^{***}	0.07^{***}	0.09	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)	(0.08)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. obs.	1460	1460	1460	1460	1460	1460
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Num. groups: code	101	101	101	101	101	101
Adj. R^2 (proj model)	0.47	0.47	0.47	0.47	0.46	0.46
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 5: Regional Development and Market Potential - Core and Periphery

	(1)	(2)	(3)	(4)	(5)	(6)
High income countries						
market potential	0.04	0.04	0.02	0.03	0.14	0.09
	(0.03)	(0.03)	(0.02)	(0.03)	(0.10)	(0.09)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. obs.	441	441	441	441	441	441
Num. groups: code	29	29	29	29	29	29
Adj. R^2 (proj model)	0.42	0.42	0.42	0.42	0.43	0.42
Upper-middle income countries						
market potential	0.16^{***}	0.13^{**}	0.12^{**}	0.10	0.14	-0.06
	(0.05)	(0.06)	(0.05)	(0.06)	(0.13)	(0.17)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.01	-0.01	-0.01^{*}	-0.01^{*}	-0.01^{*}	-0.01^{*}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Num. obs.	519	519	519	519	519	519
Num. groups: code	29	29	29	29	29	29
Adj. R^2 (proj model)	0.52	0.51	0.51	0.51	0.50	0.50
Lower-middle income countries						
market potential	0.21^{**}	0.25^{**}	0.14^{***}	0.18^{***}	-0.00	0.05
	(0.09)	(0.10)	(0.04)	(0.06)	(0.15)	(0.23)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{**}	-0.01^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Num. obs.	368	368	368	368	368	368
Num. groups: code	29	29	29	29	29	29
Adj. R^2 (proj model)	0.51	0.51	0.50	0.50	0.49	0.49
Low income countries						
market potential	0.27	-0.01	0.14	-0.09	1.21	0.30
	(0.52)	(0.65)	(0.50)	(0.62)	(1.04)	(1.03)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{*}	-0.01^{*}	-0.01^{*}	-0.01^{*}	-0.01^{*}	-0.01^{*}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.02^{**}	-0.02^{**}	-0.02^{**}	-0.02^{**}	-0.02^{**}	-0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Num. obs.	132	132	132	132	132	132
Num. groups: code	14	14	14	14	14	14
Adj. R^2 (proj model)	0.55	0.54	0.54	0.54	0.56	0.54
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i \tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j \neq i} y_j \tau_{ij}$ with i and j both in country c.

Table 6: Regional Development and Market Potential - by countries' income group (2005)

represents the income labor share and σ signifies the elasticity of substitution between varieties. The observed lower coefficient linked to the periphery, as compared to the coefficient associated with the core, could arise from either a higher income labor share or a greater elasticity of substitution between varieties. The former explanation is plausible in regions where agriculture serves as the predominant local economic activity. In such areas, which often contend with elevated transportation and trade costs, market potential may hold less relevance. A higher elasticity of substitution between varieties might depend on the quality of the goods produced and the industry (Fontagné et al., 2022).

7.3.2 Countries' income group regressions

Table 6 displays the estimated coefficients of the effect of market potential and its proxies on regional development by countries' income group following the World Bank classification: *High income*, *Upper middle income*, *Lower middle income* and *Low income*. Results show an heterogeneous effect of market potential. Especially, the split of the sample do not give significant estimate of the effect of market potential for high income and low income countries. Moreover, differences in market potential explains higher regional development differences in lower middle income countries than in upper middle countries. The former display elasticity coefficients about 0.21, 0.25 for $\text{MP}^{(s)}$, while the latter about is 0.16, 0.13 for $\text{MP}^{(s)}$. This result is in line with Boulhol and De Serres (2010) who find that, in the sample of Redding and Venables (2004), higher income OECD countries display a lower coefficient elasticity of GDP per capita to market access than for the whole sample of OECD countries, which is also lower than coefficients for the whole sample. Hence, they show that pooling developed and developing countries in country-level analysis lead to upward biased estimators (Redding and Venables, 2004; Head and Mayer, 2011).

The current findings align with the research conducted by Brülhart et al. (2020), which delves into agglomeration and dispersion effects on economic activity based on a country's developmental stage. Their model features one central and two peripheral locations, exploring the influence of decreasing trade costs on employment location. Initially, when trade costs are exceedingly high, a location's connectivity offers no discernible advantage, resulting in an even distribution of employment across all three locations. However, as trade costs decrease, the central location gains employment at the expense of the peripheries due to enhanced access. Subsequently, as trade costs continue to diminish, the central location loses its proximity advantage, rendering peripheral locations attractive once more due to reduced congestion and improved market access.

Brülhart et al. (2020)'s model and results provide insight into the absence of significant market potential influence on regional development disparities within high-income countries, often characterized by high connectivity and low transportation cost. In such contexts, the periphery is expected to converge gradually toward the level of development seen in core regions. Conversely, market potential plays a substantial role in driving development disparities within middle-income economies, where trade costs remain substantial, the economic activity tends to concentrate in areas with high market potential, often at the expense of the periphery. Furthermore, these insights help clarify the lack of significant outcomes observed in low-income economies, where transportation infrastructure remains undeveloped, and the proximity advantage has little influence on the location of employment and GDP per capita, both of which may be widely dispersed across regions.

Insignificant results may also stem from the dataset containing regions that are larger in size than what we would ideally prefer for examining economic activity concentration and disparities. In the regression results from table 5, we note that there are 1,464 observations included in the estimation, spanning across 103 countries, which leads to an average of roughly 14 regions per country — a relatively small number for analysis. In particular, there are an average of 15 regions per high-income country, 18 per upper-middle income countries, 13 per lower-middle income countries and 9 per low-income countries. Additionally, Table 17 reveals limited variation in market potential values between core and peripheral regions within countries, with the coefficient of determination ranging from approximately 5% to 7% for market potential, 3% for non-local market potential, and 0% to 1% for foreign market potential.

Summing up a multitude of values in the market potential indexes poses a challenge in capturing the regional disparities that distinguish core from periphery regions. This challenge is exacerbated by the limited differentiation in market potential values between these regions within countries. To delve deeper into the influence of proximity to core markets and tackle the constraints imposed by market potential indexes, I incorporate measures of centrality for core regions into the regression analysis. This approach seeks to introduce additional variation among regions within countries, based on their proximity to significant demand pools. The following section presents the results of this effort to address the aforementioned limitation.

7.3.3 Centrality to domestic and foreign cores

Results above have suggested that proximity to foreign markets is not necessarily beneficial for development, especially for the (semi-)periphery. It is natural to think that, within countries, regions with the strongest foreign market potential are located at the national borders. Hence, it is possible to relate this result to the trade border literature introduced by McCallum (1995). International trade is found to be more costly than intra-national trade among regions within a country. In addition, while the discussion is still vivid, there is a consensus that international trade increases income (Frankel and Romer, 1999; Anderson et al., 2020). As a result, if international borders decrease trade between regions, and if trade leads to income growth, regions at the border may display lower income levels than the national average since their location places them further to central national regions and closer to foreign regions. This statement is supported by Adam et al. (2023) and Bonadio et al. (2023), where they find that subnational regions at the international borders have lower income per capita levels and night lights intensity, while trade agreements at borders compensate this negative border effect and ultimately reduce regional disparities. On another hand, Brülhart (2006) finds that both peripheral and core regions of countries experienced growth in the context of European Union integration.[35]

Another explanation for a potential negative impact of a close proximity to foreign markets is spatial sorting of firms within countries and market selection. P.-P. Combes et al. (2012) highlights how denser cities foster competition, allowing only the most productive firms to survive. As a result, firms are spatially sorted based on their productivity, with peripheral regions typically accommodating the least productive firms. On another hand, Melitz and Ottaviano (2008) argues that easier international trade increases competition due to good access to foreign goods, leading to the demise of less productive firms and an overall higher productivity.

Among core domestic regions, firms selection effect may result in higher levels of productivity and higher development. In contrast, peripheral regions, which tend to host less productive firms, can be hurt from proximity to core foreign markets if firms cannot face the international competition. Based on this assumption, the impact of proximity to core markets with a free trade agreement is expected to be negative for peripheral regions, and higher in magnitude than proximity to core markets without free trade agreement.

[35] In particular, Brülhart (2006) finds that countries' peripheral regions grew in terms of manufacturing employment, while core regions grew in terms of service employment.

In order to expand upon the previous assertion, I examine the impact of proximity to markets within clusters and examine the effects of centrality to core domestic and foreign markets. I regress regional per capita income on development covariates with fixed effects at the country level, as well as at the cluster level, and include the centrality indexes as independent variables. To account for the impact of free trade agreements and for comparability with the results found by Adam et al. (2023) and Bonadio et al. (2023), I also differentiate between centrality to foreign cores with and without such agreements. Results are presented in tables 7 and 28, which respectively excludes and contains the market potential indexes as controls. Table 8 examines the relationships within four distinct country samples, categorized based on their income levels.

Table 7 shows that overall centrality to cores have no effect on regional development. However, when decomposing the index of centrality to cores into its domestic and foreign components, it is found that centrality to domestic cores has a positive impact on regional development, while centrality to foreign cores has a negative impact, especially for peripheral regions (see columns 3 and 4). The negative impact is primarily driven by proximity to foreign cores that have no free trade agreement with the region's country (see columns 5 and 6). It tends to approve the hypothesis that periphery regions closer to the border face higher trade costs since they are far from domestic markets in general, and close to foreign countries that apply tariffs. Proximity to foreign core regions with free trade agreements does not significantly contribute to explaining regional disparities within countries. This result prompts questions regarding the validity of the hypothesis suggesting that proximity to foreign core markets has detrimental effects on peripheral regions due to firm selection. However, I refrain from making definitive conclusions on this matter since my available data does not allow for a conclusive assessment.

Table 28 in the appendix presents results with market potential control. Including the market potential index in the regression reveals a negative effect of centrality to domestic cores for core regions, although this effect is not statistically significant (columns 7 and 8). This finding underscores the spatial division that arises from economic agglomeration, where cores are located at a certain distance from each other, and peripheral regions fill the space between them. Nevertheless, the positive and significant effect of market potential still supports the idea that regions closer to domestic cores tend to be wealthier. On the other hand, centrality to foreign cores remains a significant predictor of regional development, even with the market potential control. The closer regions are to foreign cores, the poorer they tend to be, with this relationship particularly pronounced for peripheral regions (columns 7 and 8), everything else being equal.

Table 8 displays the regression estimation derived from column 6 of Table 7, by splitting the sample into the four countries groups, categorized based on their income levels. The results consistently demonstrate a positive effect of proximity to domestic core markets, with the exception being low-income countries. This discovery may appear to contradict the findings of Baum-Snow et al. (2020), who observed that connectivity to national core markets in China had an adverse impact on regional income and population growth in the hinterlands. However, if both regional income and population size decrease, it is possible for regional income per capita to remain stable or even improve, maintaining or increasing the level of development. It is worth highlighting that China is categorized as an upper-middle-income country within the dataset, and this specific group of countries displays the most notable positive coefficient for peripheral regions.

Proximity to foreign cores without a trade agreement consistently exhibits a negative impact, with more pronounced effects observed in peripheral regions. This negative effect is highly significant in both high- and low-income countries. In contrast, proximity to foreign cores with a trade agreement is found to have a positive and statistically significant impact in lower-middle-income countries, especially for core and semi-periphery regions, which advocates for the positive impact

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{centrality}^{\text{cores}}$	-0.03					
	(0.05)					
$\text{centrality}^{\text{domestic cores}}$			0.03^{*}		0.04	
			(0.02)		(0.02)	
$\text{centrality}^{\text{foreign cores}}$			-0.31^{**}			
			(0.14)			
$\text{centrality}^{\text{foreign cores, no FTA}}$					-0.31^{*}	
					(0.16)	
$\text{centrality}^{\text{foreign cores, FTA}}$					-0.11	
					(0.12)	
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{core})$		0.33^{**}				
		(0.14)				
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$		-0.05				
		(0.05)				
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$		-0.02				
		(0.07)				
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{core})$				0.31^{***}		0.31^{**}
				(0.11)		(0.15)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$				0.02		0.02
				(0.02)		(0.02)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$				0.05^{**}		0.06^{*}
				(0.02)		(0.03)
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{core})$				-0.18		
				(0.13)		
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$				-0.32^{**}		
				(0.14)		
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$				-0.33^{**}		
				(0.15)		
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{core})$						-0.23
						(0.25)
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$						-0.30
						(0.19)
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{periphery})$						-0.30^{**}
						(0.15)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{core})$						-0.06
						(0.12)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$						-0.11
						(0.11)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{periphery})$						-0.12
						(0.12)
Num. obs.	1460	1460	1460	1460	1392	1392
Num. groups: code	101	101	101	101	97	97
Num. groups: cluster id	3	3	3	3	3	3
Adj. R^2 (proj model)	0.30	0.30	0.31	0.32	0.32	0.33

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. The market potential and centrality variables are entered with the logarithm in the regressions. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures.

Table 7: Regional Development, the Core and Periphery, and Centrality to cores

	(1)	(2)	(3)	(4)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{core})$	−0.33	0.21	14.64	−264.63***
	(0.39)	(0.27)	(18.39)	(36.35)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$	−0.02	0.05*	−0.01	0.00
	(0.03)	(0.03)	(0.02)	(0.06)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$	0.00	0.12**	0.07	−0.09**
	(0.02)	(0.05)	(0.05)	(0.03)
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{core})$	−0.08	0.05	0.03	−2.97***
	(0.36)	(0.45)	(0.27)	(0.59)
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$	−0.53**	0.11	−0.25	−3.24***
	(0.23)	(0.27)	(0.30)	(0.47)
$\text{centrality}^{\text{foreign cores, no FTA}} \times \mathbb{1}(\gamma_g = \text{periphery})$	−0.58**	−0.01	−0.15	−3.59***
	(0.23)	(0.20)	(0.43)	(0.60)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{core})$	0.13	−0.20	0.19***	0.34
	(0.13)	(0.20)	(0.06)	(0.25)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$	0.13	−0.24	0.09*	0.29
	(0.14)	(0.15)	(0.05)	(0.20)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{periphery})$	0.11	−0.24	0.03	0.08
	(0.14)	(0.15)	(0.05)	(0.16)
Countries' income sample	high	upper-midde	lower-middle	low
Num. obs.	441	489	346	116
Num. groups: code	29	28	28	12
Num. groups: cluster id	3	3	3	3
Adj. R^2 (full model)	0.84	0.68	0.79	0.97
Adj. R^2 (proj model)	0.25	0.41	0.34	0.45

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. The market potential and centrality variables are entered with the logarithm in the regressions. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures.

Table 8: Regional Development, the Core and Periphery, and Centrality to cores by countries' income group

of trade liberalization. However, the coefficient for the periphery is not significant.

To gain further insights into the impact of free trade agreements, I employ panel data to examine whether the act of signing a free trade agreement contributes to regional development, with a specific emphasis on the periphery, which is the primary subject of concern. Table 9 shows the results. I find that a 1% increase in the centrality to foreign cores with a trade agreement is expected to lead to a 0.07% increase in a region's GDP per capita (see column 1). This effect is approximately 10% for core regions, while there is no significant difference in estimates for the semi-periphery and the periphery (column 2). The estimated coefficient for each cluster group is presented in column 3, where all coefficients are approximately 0.1. This suggests that a 1% increase in the centrality to foreign cores resulting from the signing of an agreement is associated with a 0.1% increase in a region's GDP per capita. However, this effect is only statistically significant at the 10% level and becomes insignificant when accounting for development covariates.

While this study provides valuable insights, further research is needed to fully understand the dynamics of proximity to foreign markets, market selection, and regional development. Future

	(1)	(2)	(3)	(4)
$\text{centrality}^{\text{foreign cores, FTA}}$	0.07	0.10^{*}		
	(0.05)	(0.05)		
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$		0.03	0.12^{*}	0.07
		(0.03)	(0.07)	(0.05)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{periphery})$		−0.01	0.09^{*}	0.08
		(0.03)	(0.05)	(0.06)
$\text{centrality}^{\text{foreign cores, FTA}} \times \mathbb{1}(\gamma_g = \text{core})$			0.10^{*}	0.05
			(0.05)	(0.04)
Num. obs.	2676	2395	2395	2049
region FE	1105	890	890	882
cluster group FE		3	3	3
country × year FE	162	143	143	127
Covariates	No	No	No	Yes
Adj. R^2 (full model)	0.98	0.98	0.98	0.98
Adj. R^2 (proj model)	−0.00	−0.00	−0.00	0.24

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 9: Regional Development and Trade Agreement - Panel

studies should employ micro-founded analyses to explore the specific mechanisms of firms' selection, consider the heterogeneous impacts of trade on different regions, and account for local economic factors. By delving deeper into these aspects, researchers can validate and refine the assumptions made in this study, ultimately providing a more comprehensive understanding on challenges face by peripheral areas.

8 Conclusion

The main objective of this paper is to explore the key factors contributing to regional development disparities within countries and investigate the role played by proximity to markets on the core-periphery divide. This paper contributes to provide a falsification test to the international trade wage equation developed by Fujita et al. (1999) at the regional level with a worldwide scope. By analyzing these factors across a wide range of regions and countries, this research aims to provide a more comprehensive understanding of the international trade wage equation and its implications for regional disparities.

I use an extended regional dataset which includes more than 1,500 subnational regions, from 107 different countries, out of the 195 recognized around the world. This dataset contains information on regional economic activity, geography and education, to control the estimations for these regional development covariates. Hence, market potential indexes are built at the regional level, following a gravity-based approach. One of the contributions of the paper lies in the computation of a more accurate measure of distance between markets, which considers land masses and ocean areas in the market potential distance function. I compute the inland shortest path between regions and ports, as well as the overseas shortest path between ports.

Results provide evidence that access to markets has a significant impact on regional development for both specification of market potential presented. In particular, proximity to demand is associated to a more intense economic activity and a higher income per capita. In other words, regions surrounded by rich markets may be richer than regions with poor neighbours, due to lower trade costs as stated by the wage equation, but also to positive externalities. This statement comes the fact that the market potential index computed on the basis of the haversine distance function seems equally robust than the one constructed with the shipment distance function. It suggests that other things than goods may flow from a region to another, which does not rely on port connection, such as knowledge and people.

Connecting regions between each other seems to be an important challenge to tackle in order to develop and increase living conditions everywhere. However, heterogeneous effects exist conditional on the national income group, as well as on the regional core-periphery structure within countries. On one hand, the effect of market potential is found to be particularly strong in both upper- and lower-middle-income countries. On the other hand, (semi-)peripheral regions are found to exhibit lower elasticity coefficients to market potential compared to core regions. Additionally, results show that being in close proximity to foreign core markets can be detrimental to regions located in the (semi-)periphery. Peripheral regions tend to have lower per capita income and lower market potential compared to core regions. Additionally, peripheral regions experience slower growth rates over time, which further widens the income gap between them and core regions. Moreover, due to their lower development elasticity coefficient to market potential, peripheral regions may be less responsive to the growth of surrounding markets, which could leave them even more disadvantaged in the long run.

Further investigation highlights that the impact of proximity to core markets varies depending on whether they are domestic or foreign, as well as whether a region's country has a free trade agreement with others. Proximity to domestic cores is associated with positive effects on regional development, while proximity to foreign cores without a trade agreement has a negative impact for peripheral regions. Interestingly, the effect of proximity to foreign cores with a free trade agreement is not significantly different from zero, suggesting that the presence of a free trade agreement appears to mitigate the negative impact of close proximity to foreign core markets, or at least reduce regional disparities. Indeed, the analysis using panel data indicates that the act of signing a trade agreement, leading to an increase in centrality to foreign cores with FTA, is associated with a rise in GDP per capita, although the statistical significance is modest. These results imply that the existence of a free trade agreement shows potential in fostering regional development, including in peripheral areas, as also evidenced in the literature (Brülhart et al., 2004; Adam et al., 2023; Bonadio et al., 2023).

These findings underscore the importance of understanding the unique challenges faced by peripheral regions in achieving economic development and suggest the need for policies that are tailored to the specific circumstances of these regions. One policy suggestion that arises from the study is to enhance connectivity between peripheral regions and core domestic regions in order to reduce transportation costs and increase access to domestic demand. Nonetheless, it's worth noting that this policy might not be the most suitable approach for low-income countries, as there's a potential risk of regional economic activity and population relocating from the periphery to existing core regions. Additionally, policies aiming at improving trade relations and establishing free trade agreements with foreign countries may be beneficial to peripheral regions.

References

A'Hearn, B. and Venables, A. J. (2013) Regional Disparities: Internal Geography and External Trade. In: *The Oxford Handbook of the Italian Economy Since Unification.* Oxford University Press.

Abreu, M. and Öner, Ö. (2020) Disentangling the brexit vote: the role of economic, social and cultural contexts in explaining the uk's eu referendum vote, *Environment and Planning A: Economy and Space* **52**, 1434–1456.

Acemoglu, D., Gallego, F. A., and Robinson, J. A. (2014) Institutions, human capital, and development, *Annual Review of Economics* **6**, 875–912.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2005) Institutions as a fundamental cause of long-run growth, *Handbook of economic growth* **1**, 385–472.

Adam, H. L., Larch, M., and Stadelmann, D. (2023) Trade agreements and subnational income of border regions, *Economic Inquiry.*

Ahmad, S. and Riker, D. (2020) Updated estimates of the trade elasticity of substitution, *US International Trade Commission, Economics Working Paper.*

Anderson, J. E., Larch, M., and Yotov, Y. V. (2020) Transitional growth and trade with frictions: a structural estimation framework, *The Economic Journal* **130**, 1583–1607.

Baldwin, R. E. and Martin, P. (2004) Agglomeration and regional growth (chapter 60). In: *Cities and geography.* Ed. by J. V. Henderson and J.-F. Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, 2671–2711.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., and Brandt, L. (2020) Does investment in national highways help or hurt hinterland city growth?, *Journal of Urban Economics* **115**, 103124.

Blonigen, B. A. and Wilson, W. W. (2008) Port efficiency and trade flows*, *Review of International Economics* **16**, 21–36.

Bonadio, B. (2022) Ports vs. roads: infrastructure, market access and regional outcomes, *Working Paper.*

Bonadio, B., Dhabi, N. A., Brülhart, M., Cadot, O., and Rais, G. (2023) And there was light: trade and the development of border regions.

Boulhol, H. and De Serres, A. (2010) Have developed countries escaped the curse of distance?, *Journal of Economic Geography* **10**, 113–139.

Brakman, S., Garretsen, H., and Marrewijk, C. V. (Oct. 2009) Economic Geography Within And Between European Nations: The Role Of Market Potential And Density Across Space And Time, *Journal of Regional Science* **49**, 777–800.

Brakman, S., Garretsen, H., and Schramm, M. (2004) The spatial distribution of wages: estimating the helpman-hanson model for germany, *Journal of regional science* **44**, 437–466.

Breinlich, H. (2006) The spatial income structure in the european union—what role for economic geography?, *Journal of Economic Geography* **6**, 593–617.

Brülhart, M. (2006) The fading attraction of central regions: an empirical note on core–periphery gradients in western europe, *Spatial Economic Analysis* **1**, 227–235.

Brülhart, M., Crozet, M., and Koenig, P. (2004) Enlargement and the eu periphery: the impact of changing market potential, *World Economy* **27**, 853–875.

Brülhart, M., Desmet, K., and Klinke, G.-P. (2020) The shrinking advantage of market potential, *Journal of Development Economics* **147**, 102529.

Bruna, F., Lopez-Rodriguez, J., and Faíña, A. (2016) Market potential, spatial dependences and spillovers in european regions, *Regional Studies* **50**, 1551–1563.

Chiquiar, D. (2008) Globalization, regional wage differentials and the Stolper-Samuelson Theorem: Evidence from Mexico, *Journal of International Economics* **74**, 70–93.

Combes, P. P., Duranton, G., and Overman, H. G. (Aug. 2005) Agglomeration and the adjustment of the spatial economy§, *Papers in Regional Science* **84**, 311–349.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012) The productivity advantages of large cities: distinguishing agglomeration from firm selection, *Econometrica* **80**, 2543–2594.

Coughlin, C. C. and Novy, D. (2021) Estimating border effects: the impact of spatial aggregation, *International Economic Review* **62**, 1453–1487.

Daniele, V., Malanima, P., and Ostuni, N. (2018) Geography, market potential and industrialization in italy 1871-2001, *Papers in Regional Science* **97**.

de Sousa, J. (2012) The currency union effect on trade is decreasing over time, *Economics Letters* **117**, 917–920.

Dijkstra, L., Poelman, H., and Rodrıéguez-Pose, A. (2020) The geography of eu discontent, *Regional Studies* **54**, 737–753.

Donaldson, D. (Apr. 2018) Railroads of the raj: estimating the impact of transportation infrastructure, *American Economic Review* **108**, 899–934.

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

Ducruet, C. (2020) The geography of maritime networks: a critical review, *Journal of Transport Geography* **88**, 102824.

Fallah, B., Partridge, M., and Olfert, M. (2009) New economic geography and us metropolitan wage inequality, *Journal of Economic Geography* **11**.

Fally, T., Paillacar, R., and Terra, C. (2010) Economic geography and wages in Brazil: Evidence from micro-data, *Journal of Development Economics* **91**, 155–168.

Fingleton, B. (2008) Competing models of global dynamics: evidence from panel models with spatially correlated error components, *Economic Modelling* **25**, 542–558.

Fontagné, L., Guimbard, H., and Orefice, G. (2022) Tariff-based product-level trade elasticities, *Journal of International Economics* **137**, 103593.

Frankel, J. A. and Romer, D. H. (1999) Does trade cause growth?, *American Economic Review* **89**, 379–399.

Frensch, R., Fidrmuc, J., and Rindler, M. (2023) Topography, borders, and trade across europe, *Journal of Comparative Economics*.

Fujita, M., Krugman, P., and Venables, A. J. (1999) *The spatial economy: cities, regions, and international trade*. MIT press.

Gaulier, G. and Zignago, S. (2010) *Baci: international trade database at the product-level. the 1996-2020 version*. Working Papers 2010-23. CEPII.

Gennaioli, N., LaPorta, R., Lopez-de-Silanes, F., and Shleifer, A. (2013) Human capital and regional development, *Quarterly Journal of Economics* **128**, 105–164.

Gennaioli, N., Porta, R. L., Silanes, F. L. D., and Shleifer, A. (2014) Growth in regions, *Journal of Economic Growth* **19**, 259–309.

Guerriero, M. (2019) *The labor share of income around the world: evidence from a panel dataset*. Springer.

Hanson, G. H. (2005) Market potential, increasing returns and geographic concentration, *Journal of International Economics* **67**, 1–24.

Harris, C. D. (1954) The market as a factor in the localization of industry in the united states, *Annals of the Association of American Geographers* **44**, 315–348.

Head, K. and Mayer, T. (2004) Market potential and the location of japanese investment in the european union, *Review of Economics and Statistics* **86**, 959–972.
Head, K. and Mayer, T. (2006) Regional wage and employment responses to market potential in the EU, *Regional Science and Urban Economics* **36**, 573–594.
Head, K. and Mayer, T. (2011) Gravity, market potential and economic development, *Journal of Economic Geography* **11**, 281–294.
Head, K. and Mayer, T. (2014) Gravity Equations: Workhorse,Toolkit, and Cookbook. In: *Handbook of International Economics.* Ed. by G. Gopinath, . Helpman, and K. Rogoff. Vol. 4. Handbook of International Economics. Elsevier. Chap. 0, 131–195.
Head, K. and Ries, J. (2001) Increasing returns versus national product differentiation as an explanation for the pattern of u.s.-canada trade, *American Economic Review* **91**, 858–876.
Hering, L. and Paillacar, R. (2016) Does access to foreign markets shape internal migration? evidence from brazil, *World Bank Economic Review* **30**, 78–103.
Hering, L. and Poncet, S. (2010) Market access and individual wages: evidence from china, *The Review of Economics and Statistics* **92**, 145–159.
Hornbeck, R. and Rotemberg, M. (2021) Railroads, market access, and aggregate productivity growth, *University of Chicago Booth School of Business, mimeo.*
Hummels, D. (2007) Transportation costs and international trade in the second era of globalization, *Journal of Economic perspectives* **21**, 131–154.
Jacks, D. S. and Novy, D. (2018) Market Potential and Global Growth over the Long Twentieth Century, *Journal of International Economics* **114**, 221–237.
Karahasan, B. C., Dogruel, F., and Dogruel, A. S. (2016) Can market potential explain regional disparities in developing countries? evidence from turkey, *The Developing Economies* **54**, 162–197.
Kosfeld, R. and Eckey, H.-F. (2010) Market access, regional price level and wage disparities: the german case, *Jahrbuch für Regionalwissenschaft* **30**, 105–128.
Krugman, P. (1980) Scale economies, product differentiation, and the pattern of trade, *American Economic Review* **70**, 950–59.
Krugman, P. (1991) Increasing returns and economic geography, *Journal of Political Economy* **99**, 483–99.
Lai, H. and Trefler, D. (2002) *The gains from trade with monopolistic competition: specification, estimation, and mis-specification.*
Limão, N. and Venables, A. J. (2001) Infrastructure, geographical disadvantage, transport costs, and trade, *The World Bank Economic Review* **15**, 451–479.
López-Rodrıéguez, J. and Andrés Faıéña, J. (2006) Does distance matter for determining regional income in the european union? an approach through the market potential concept, *Applied Economics Letters* **13**, 385–390.
López-Rodrıéguez, J., Márquez, M. A., and Faiña, A. (2008) Economic geography and spatial wage structure in spain, *REAL Discussion Papers 08-T* **4**.
McCallum, J. (1995) National borders matter: canada-us regional trade patterns, *The American Economic Review* **85**, 615–623.
Melitz, M. J. and Ottaviano, G. I. (2008) Market size, trade, and productivity, *The review of economic studies* **75**, 295–316.
Mion, G. (2004) Spatial externalities and empirical analysis: the case of Italy, *Journal of Urban Economics* **56**, 97–118.
Mion, G. and Naticchioni, P. (2009) The spatial sorting and matching of skills and firms, *Canadian Journal of Economics/Revue canadienne d'économique* **42**, 28–55.

Niebuhr, A. (2006) Market access and regional disparities: new economic geography in europe, *The Annals of Regional Science* **40**, 313–334.

Ottaviano, G. I. and Pinelli, D. (2006) Market potential and productivity: evidence from finnish regions, *Regional Science and Urban Economics* **36**, 636–657.

Özgüzel, C. (2022) Agglomeration effects in a developing economy: evidence from Turkey, *Journal of Economic Geography*. lbac035.

Paredes, D. (2013) The role of human capital, market potential and natural amenities in understanding spatial wage disparities in chile, *Spatial Economic Analysis* **8**, 154–175.

Pires, A. J. G. (2006) Estimating krugman's economic geography model for the spanish regions, *Spanish Economic Review* **8**, 83–112.

Portugal-Perez, A. and Wilson, J. S. (2012) Export Performance and Trade Facilitation Reform: Hard and Soft Infrastructure, *World Development* **40**, 1295–1307.

Redding, S. J. (2022) Chapter 3 - trade and geography. In: *Handbook of international economics: international trade, volume 5*. Ed. by G. Gopinath, E. Helpman, and K. Rogoff. Vol. 5. Handbook of International Economics. Elsevier, 147–217.

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

Reshef, A. and Santoni, G. (2023) Are your labor shares set in beijing? the view through the lens of global value chains, *European Economic Review* **155**, 104459.

Rodríguez-Pose, A., Lee, N., and Lipp, C. (2020) Golfing with trump: social capital, decline, inequality, and the rise of populism in the us.

Romalis, J. (2007) Nafta's and cusfta's impact on international trade, *The Review of Economics and Statistics* **89**, 416–435.

Rosenthal, S. S. and Strange, W. C. (2004) Evidence on the nature and sources of agglomeration economies. In: *Handbook of regional and urban economics*. Vol. 4. Elsevier, 2119–2171.

Wessel, P. and Smith, W. (1996) A global, self-consistent, hierarchical, high-resolution shoreline database, *Journal of Geophysical Research* **101**, 8741–8743.

Zou, W., Chen, L., and Xiong, J. (2021) High-speed railway, market access and economic growth, *International Review of Economics & Finance* **76**, 1282–1304.

Appendix A Regional Trade Gravity Equation and Trade Elasticity

A.1 Regional Trade Gravity Equation

To express and compute regional market potential indexes, I estimate the gravity equation in order to use the coefficients of trade elasticity to trade costs as explained in section 3. The gravity model states that the force of attraction between two entities depends positively on their respective mass and negatively on the distance between them - the distance playing as a resistance force. Applied to trade, it states that the wealthier and the closer two markets are, the more intense their bilateral trade activity. The underlying mechanisms follow a supply and demand analysis. The larger the exporter's production, the more it is going to sell. The larger the importer's demand, the more it is going to expend. If both markets are close, transportation costs from one to another are low. Thus, the two markets are likely to fulfill their need: to trade between each other.

To estimate the gravity equation, I use transnational trade panel data and estimate the expected trade flows between regions of each pair of countries. I assume that the bilateral trade flows between two regions of different countries is proportional to the share of their regional Gross Domestic Product (GDP) in the overall national GDP. Hence, the wealthier the region, the more it participates to the wealth of its country, and the more intense its international activity. Hence, bilateral trade flows between regions are proxied by:

$$\tilde{\mathrm{TF}}_{ij} = \mathrm{TF}_{c_i c_j} \times \frac{\mathrm{GDP}_i}{\mathrm{GDP}_{c_i}} \times \frac{\mathrm{GDP}_j}{\mathrm{GDP}_{c_j}} \tag{19}$$

with $\tilde{\mathrm{TF}}_{ij}$ the expected bilateral trade flows between region i in country c_i and region j in country c_j, with $c_i \neq c_j$, $\mathrm{TF}_{c_i c_j}$ the observed value of bilateral trade flows observed between countries c_i and c_j in the data, GDP_i and GDP_{c_i} the Gross Domestic Product of region i and its country c_i respectively. The closer the ratio $\mathrm{GDP}_i/\mathrm{GDP}_{c_i}$ to unity, the more country c_i's economic activity relies on the one of region i. The same applies for the ratio $\mathrm{GDP}_j/\mathrm{GDP}_{c_j}$.

Then, the gravity equation is expressed as follows:

$$\tilde{\mathrm{TF}}_{ij} = A \times \mathrm{Y}_i^{\alpha_1} \times \mathrm{Y}_j^{\alpha_2} \times \tau_{ij} \tag{20}$$

with A a constant, $\tilde{\mathrm{TF}}_{ij}$ the expected total bilateral trade flows between regions i and j, Y_i and Y_j their respective mass, and τ_{ij} the resistance term -or proxy of trade costs- between them. The latter variable has been expressed in two manners, as showed in equations 3 and 5.

The gravity equation is estimated in cross-section as follows:

$$\begin{aligned}\ln \tilde{\mathrm{TF}}_{ij} = \; & \beta_1 \ln \mathrm{dist}_{ij}^{(\text{haversine})} + \beta_2 \mathbb{1}_{\text{language}_{ij}} + \beta_3 \mathbb{1}_{\text{contiguity}_{ij}} + \beta_4 \mathbb{1}_{\text{colony}_{ij}} \\ & + \beta_5 \mathbb{1}_{\text{rta}_{ij}} + \beta_6 \mathbb{1}_{\text{currency}_{ij}} + \delta_i + \delta_j + \epsilon_{ij}\end{aligned} \tag{21}$$

$$\begin{aligned}\ln \tilde{\mathrm{TF}}_{ij} = \; & \gamma_1 \ln \mathrm{dist}_{io}^{(\text{land, from exporter to origin port})} \mathbb{1}_{\text{maritime route}} \\ & + \gamma_2 \ln \mathrm{dist}_{od}^{(\text{sea, between ports})} \mathbb{1}_{\text{maritime route}} \\ & + \gamma_3 \ln \mathrm{dist}_{dj}^{(\text{land, from destination port to importer})} \mathbb{1}_{\text{maritime route}} \\ & + \gamma_4 \ln \mathrm{dist}_{ij}^{(\text{land})} (1 - \mathbb{1}_{\text{maritime route}}) \\ & + \beta_2 \mathbb{1}_{\text{language}_{ij}} + \beta_3 \mathbb{1}_{\text{contiguity}_{ij}} + \beta_4 \mathbb{1}_{\text{colony}_{ij}} + \beta_5 \mathbb{1}_{\text{rta}_{ij}} + \beta_6 \mathbb{1}_{\text{currency}_{ij}} \\ & + \delta_i + \delta_j + \epsilon_{ij}\end{aligned} \tag{22}$$

with $\text{dist}_{ij}^{(\text{land})} = \kappa_{ij}$, $\text{dist}_{io}^{(\text{land, from exporter to origin port})} = \kappa_{io}$, $\text{dist}_{dj}^{(\text{land, from destination port to importer})} = \kappa_{od}$ and $\text{dist}_{ij}^{(\text{sea, between ports})} = \kappa_{od}$ from equation 5. The dummy $\mathbb{1}_{\text{maritime route}}$ equals 1 if the condition expressed in equation 4 is respected, i.e. if it is more convenient for region i to send its goods overseas to region j, and 0 otherwise. $\tilde{\text{TF}}_{ij}$ is the expected total bilateral trade flows between regions i and j, in countries c_i and c_j respectively, with $c_i \neq c_j$. Since I estimate the effect of distance on trade flows which is time invariant, I cannot use pairs fixed effects to reduce endogeneity problems. However, I include fixed effects for both regions of the pair, δ_i and δ_j, in order to control for the unobservable and unit-specific confounders in 2005. Therefore, it controls for supply, demand and multilateral resistance terms of each region.

The trade literature has widely estimated coefficients α_1, α_2 and β_1 in cross-section and repeated cross-section estimations at the national level. Coefficients α_1 and α_2 are generally found to equal 1, and β_1 to -1. This result is stable across different periods of time and samples of countries. Coefficients β_2, β_3, β_4, β_5 and β_6 estimates are supposed to be positive, as they play as trade facilitators. Next subsection presents the results of the estimated coefficients of trade elasticity to trade resistance and facilitator terms at the regional level using expected bilateral trade flows and compares the results with those found in the literature.

A.2 Regional Trade Costs Parameters Estimations

Table 10 shows the estimated coefficients from equations 21 and 22 in columns 1 and 2 respectively. Without surprise, the estimator of trade elasticity to physical distance is negative regardless of the specification of distance. Trade elasticity to distance, as measured as haversine, is about -1.18 (column 1), meaning that a 1% increase in the great-circle distance in kilometers between two regions of different countries decreases the total bilateral trade flows by 1.18% between them.

Distance as measured by the shortest path in kilometers between regions i and j considering land and ocean areas also give negative coefficients of trade elasticity. When it is considered to be convenient for the shipment to be conducted on maritime routes between regions i and j, trade elasticity to distance from regions to their closest port is estimated to be about $[-0.07; -0.06]$, and between ports about -0.96. Otherwise, for those pairs that are considered to ship goods by land routes only, the trade elasticity coefficient is -1.

As expected, contiguity, common language, colonial ties and being part of a regional trade agreement play as trade facilitator forces. Indeed, the corresponding coefficients are found to be significantly positive. Trade elasticity to contiguity and colonial ties are found to be greater than 1. Unexpectedly, the dummy of common currency gives a negative and significant coefficient. It suggests that two regions belonging to two different countries tend to trade less between each other if their countries share a common currency, whereas it is supposed to be a trade facilitator as well.

Section A.3 in the appendix presents the results obtained from estimations undergone at the country level, compares the coefficients with those found in the literature and presents the statistical tests conducted. Elasticity coefficients estimated here are found to be similar to estimates from the country level analysis, as well as to coefficients usually obtained from gravity equation estimations in the literature (see table 12), except for the effect of sharing a common currency.

For the computation of the market potential indexes, I am going to use estimators of distance, and the estimates of trade facilitator found in column 2 table 10, except for the effect of currency which may be badly estimated. Instead, I use the average coefficient of 0.79 estimated in the literature (see column 1 of table 12). Since I express the market potential of subnational regions, I consider regions belonging to same countries as well. However, the worldwide coverage data I use for the gravity equation estimations do not contain intra-national trade flows. The coefficient β_7,

	(1)	(2)
$\text{dist}_{ij}^{(\text{haversine})}$	-1.18^{***}	
	(0.02)	
language_{ij}	0.71^{***}	0.66^{***}
	(0.04)	(0.04)
contiguity_{ij}	1.13^{***}	1.12^{***}
	(0.05)	(0.05)
colony_{ij}	1.34^{***}	1.37^{***}
	(0.07)	(0.07)
rta_{ij}	0.40^{***}	0.47^{***}
	(0.04)	(0.03)
currency_{ij}	-0.92^{***}	-0.81^{***}
	(0.11)	(0.11)
$\mathbb{1}_{\text{maritime route}} \times \text{dist}_{io}^{(\text{land, from i to origin port})}$		-0.07^{***}
		(0.01)
$\mathbb{1}_{\text{maritime route}} \times \text{dist}_{od}^{(\text{sea, between ports})}$		-0.96^{***}
		(0.02)
$\mathbb{1}_{\text{maritime route}} \times \text{dist}_{dj}^{(\text{land, from destination port to j})}$		-0.06^{***}
		(0.01)
$(1 - \mathbb{1}_{\text{maritime route}}) \times \text{dist}_{ij}^{(\text{land})}$		-1.00^{***}
		(0.02)
Fixed Effects		
region i	Yes	Yes
region j	Yes	Yes
Num. obs.	2,378,212	2,378,212
Num. groups: region i	1,617	1,617
Num. groups: region j	1,617	1,617
Adj. R^2 (proj model)	0.33	0.33

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. In parenthesis, the robust standard errors are two-way clustered at the regional level.

Table 10: Gravity Equation Estimates - regional level

which represents the border effect has been estimated to equal 1.96 on average, such as displayed by the coefficients of the variable home_{ij} from the meta-analysis of Head and Mayer (2014) (see column 1 of table 12).

Therefore, the vectors of estimators is the following: $\left[\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_7, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4\right] = [-1.18, 0.66, 1.12, 1.37, 0.47, 0.79, 1.96, -0.07, -0.96, -0.06, -1.00]$ From these estimates coefficients, market potential indexes are computed following equation 1.

A.3 Robustness - Trade Elasticity Estimates Comparison

For robustness, regression 21 is estimated at the country level with the observed trade flows between countries as dependant variable. Results are displayed in table 11. The different columns display the results for different samples of countries. Column 1 includes countries both present in BACI and Gennaioli datasets for the year 2005. Column 2 includes all countries present in BACI dataset for the year 2005. Column 3 and 4 follow columns 1 and 2 respectively, but including all the years between 1996 and 2010.

Coefficients found with countries as level of geographic unit are very similar. Regarding the effects of having a common currency is still found to be negative, but statistically insignificant for specifications in columns 2 to 4. Hence, the negative coefficient displayed may result from an endogeneity problem and/or a sample selection bias. Further investigation shows that the negative sign becomes positive when adding pair fixed effects. The coefficient takes value around 0.17 with the sample of regression in column (3), significant at the 1% level, and 0.16 with the sample of regression in column (4), significant at the 5% level.

	(1)	(2)	(3)	(4)
$\text{dist}_{c_1c_2}^{(\text{haversine})}$	-1.21^{***}	-1.40^{***}	-1.17^{***}	-1.38^{***}
	(0.09)	(0.07)	(0.08)	(0.06)
$\text{language}_{c_1c_2}$	0.92^{***}	0.82^{***}	0.87^{***}	0.72^{***}
	(0.16)	(0.11)	(0.21)	(0.10)
$\text{contiguity}_{c_1c_2}$	0.86^{***}	0.72^{***}	0.81^{***}	0.70^{***}
	(0.23)	(0.21)	(0.14)	(0.19)
$\text{colony}_{c_1c_2}$	1.02^{***}	0.65^{***}	1.18^{***}	0.66^{***}
	(0.29)	(0.15)	(0.29)	(0.14)
$\text{rta}_{c_1c_2t}$	0.32^{**}	0.51^{***}	0.32^{**}	0.44^{***}
	(0.14)	(0.11)	(0.12)	(0.10)
$\text{currency}_{c_1c_2t}$	-0.83^{**}	-0.04	-0.94^{***}	-0.11
	(0.38)	(0.35)	(0.31)	(0.28)
Fixed Effects				
Country c_1	Yes	Yes	No	No
Country c_2	Yes	Yes	No	No
Country c_1 × year t	No	No	Yes	Yes
Country c_2 × year t	No	No	Yes	Yes
Sample				
Year	2005	2005	[1996; 2010]	[1996; 2010]
Countries	BACI ∩ Gennaioli	All in BACI	BACI ∩ Gennaioli	All in BACI
Num. obs.	7,057	18,618	99,928	254,889
Num. groups: $c_1(\times t)$	77	146	1,155	2,190
Num. groups: $c_2(\times t)$	103	208	1,545	3,116
Adj. R^2 (proj model)	0.30	0.26	0.30	0.25

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. In parenthesis, the robust standard errors are two-way clustered at the national level. Note that the four first variables in the estimation are time invariant in the period [1996; 2010], while the two latter are time varying for some pairs of countries.

Table 11: Gravity Equation Estimates - national level

The significance tests conducted over the coefficients in tables 10 and 11 examined whether coefficients where statistically different from zero in order to assess the effect of the presented variable on trade. In other words, the test conducted can be written as: $H_0^{(1)} : \beta = 0$, with $H_1^{(1)} : \beta \neq 0$ the alternative, with $\beta = [\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \gamma_1, \gamma_2, \gamma_3, \gamma_4]$ the set of coefficients to be estimated.

I conduct other statistical tests here. The second test evaluates whether the estimated coefficients in the expected regional trade flows are equal to the ones found from the estimations con-

	HM 1	HM 2
$\text{dist}_{ij}^{(\text{haversine})}$	-0.93	-1.10
language_{ij}	0.54	0.39
contiguity_{ij}	0.53	0.66
colony_{ij}	0.92	0.75
rta_{ij}	0.59	0.75
currency_{ij}	0.79	0.86
home_{ij}	1.96	1.90

Average trade elasticity to resistance and facilitator variables found in the literature, gathered in the meta-analysis by Head and Mayer (2014). First column displays the average coefficients in the whole sample of gravity models in their analysis, while the second only includes coefficients estimated from structural gravity models.

Table 12: Gravity Equation Estimates - Head and Mayer (2014)

ducted at the national level. The corresponding hypothesis can be written as: $H_0^{(2)} : \beta^{(1)} = \hat{\beta}_{C1}$, and $H_1^{(2)} : \beta^{(1)} \neq \hat{\beta}_{C1}$ the alternative, with $\hat{\beta}_{C1}$ the set of estimated coefficients table 11 column 1. The same test is undertaken for the estimates $\beta^{(2)}$ in table 11 column 2.

The third and fourth tests evaluate whether the estimated coefficients from national estimations are equal to the ones found from the meta-analysis of all gravity papers and of structural gravity papers used by Head and Mayer (2014). The hypothesis can respectively be written as: $H_0^{(3)} : \beta_{C1} = \hat{\beta}_{HM1}$, with $H_1^{(3)} : \beta_{C1} \neq \hat{\beta}_{HM1}$ the alternative, and $H_0^{(4)} : \beta_{C1} = \hat{\beta}_{HM2}$, with $H_1^{(4)} : \beta_{C1} \neq \hat{\beta}_{HM2}$ the alternative. The sets of estimated coefficients $\hat{\beta}_{HM1}$ and $\hat{\beta}_{HM2}$ are transcribed in columns 1 and 2 of table 12 respectively. Finally, the same hypothesis are tested with $\beta^{(1)}$ and $\beta^{(2)}$.

	$\beta^{(1)} = \hat{\beta}_{C1}$	$\beta^{(2)} = \hat{\beta}_{C1}$	$\beta_{C1} = \hat{\beta}_{HM1}$	$\beta_{C1} = \hat{\beta}_{HM2}$	$\beta^{(1)} = \hat{\beta}_{HM1}$	$\beta^{(1)} = \hat{\beta}_{HM2}$
$\text{dist}_{ij}^{(\text{haversine})}$	0.23	-	0.00	0.24	0.00	0.00
language_{ij}	0.00	0.00	0.01	0.00	0.00	0.00
contiguity_{ij}	0.00	0.00	0.15	0.36	0.00	0.00
colony_{ij}	0.00	0.00	0.83	0.44	0.00	0.00
rta_{ij}	0.00	0.00	0.03	0.62	0.00	0.32
currency_{ij}	0.28	0.95	0.00	0.00	0.00	0.00

The table presents p-values of the tests related to the null hypothesis displayed in the top of each column. A p-value of $p < 0.05$ rejects the null hypothesis at the 95% confidence level. If $p \geq 0.05$, we can consider that the null hypothesis is not rejected.

Table 13: Statistical tests - p-values

Table 13 presents the p-values of the different statistical tests. Columns 1 and 2 show that all variables give elasticity coefficients significantly different from the regional to national gravity equations, the former using expected inter-regional trade flows and the latter using observed international trade flows. This is true except for the haversine distance and the dummy of common currency. Despite the significant difference between estimates, they are similar.

Then, coefficients from the country-level-estimations in table 11 are compared to those from the literature displayed in table 12. The haversine distance is found to be statistically different from -0.93 but not from -1.10. Among the trade facilitator variables, contiguity, colonial ties and regional trade agreement are found to not be statistically different from the average coefficients estimated in the literature. However, the effect of sharing a currency is unexpectedly found to be negative, and especially to be equal to the opposite of the ones found in the literature.

Appendix B Figures

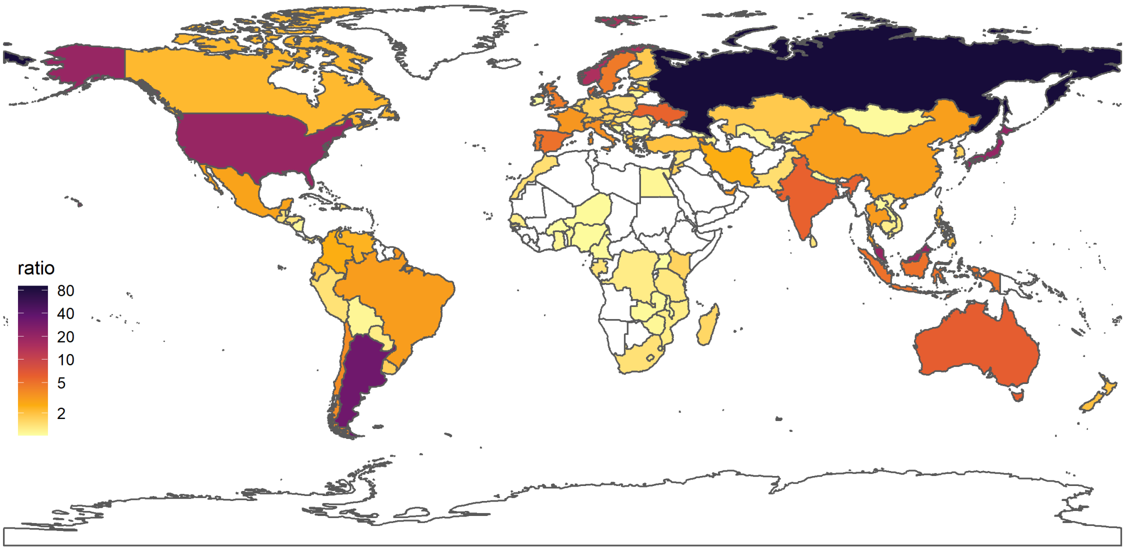

Figure 4: Ratio highest/lowest regional market potential $\mathrm{MP}^{(s)}$ in 2005

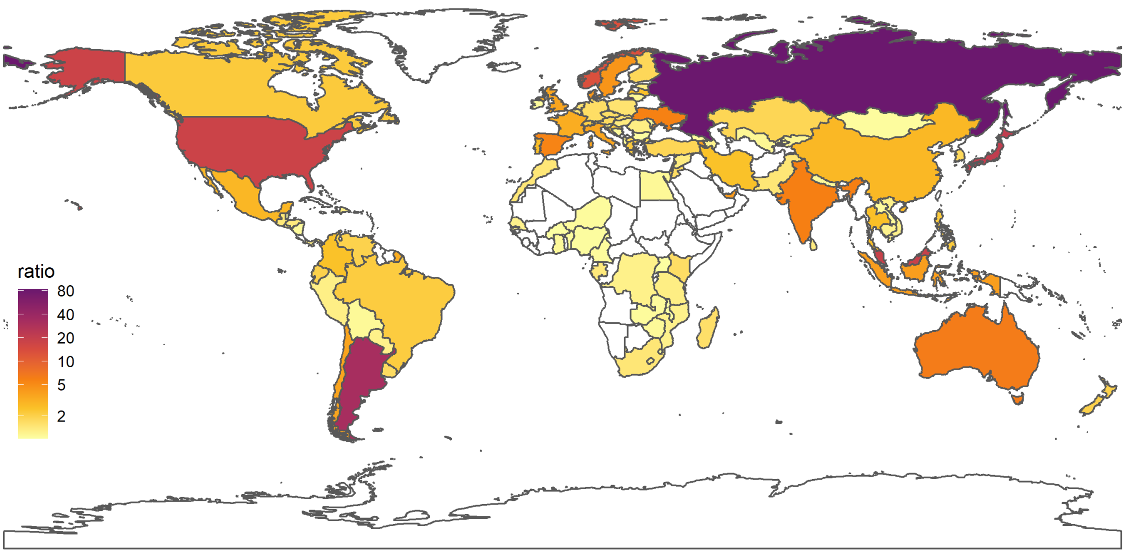

Figure 5: Ratio highest/lowest regional non-local market potential $\mathrm{NLMP}^{(s)}$ in 2005

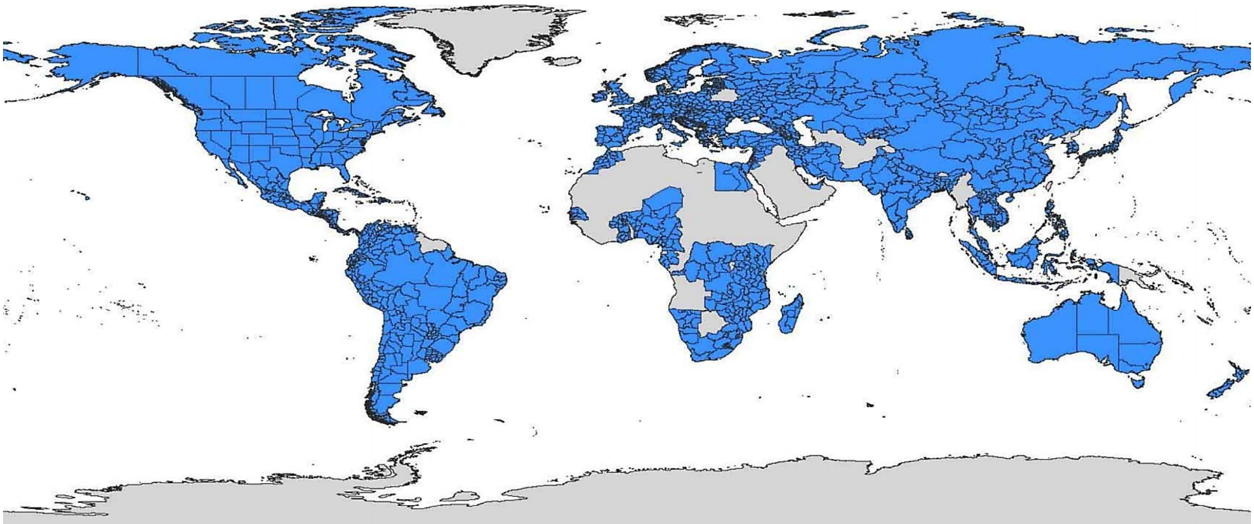

Figure 6: Gennaioli et al. (2013) regional dataset

Figure 7: World ports selected as the closest ports to each region in the sample

Appendix C Tables

	All countries					*High income countries*				
	mean	**min**	**max**	**sd**	**n**	**mean**	**min**	**max**	**sd**	**n**
log(GDP pc)	8.77	1.76	11.87	1.32	1502	10.02	8.9	11.24	0.45	289
temperature	15	-12.73	29.15	8.45	1501	10	-12.2	27.27	4.94	289
inv. dist. port	0.04	0	1	0.18	1502	0.07	0	1	0.23	289
log(oil production pc)	0.1	0	4.16	0.4	1502	0.05	0	3.13	0.27	289
log(density)	4.17	-2.93	12.01	1.81	1502	4.79	-2.93	10.23	1.55	289
years education	7.13	0.22	13.21	3.16	1465	10.09	5.3	12.94	1.71	289
$\log(MP^{(h)})$	21.21	19.07	25.49	1.32	1502	22.74	20.6	25.49	0.93	289
$\log(MP^{(s)})$	22.23	20.06	25.92	1.22	1502	23.62	21.81	25.92	0.89	289
$\log(NLMP^{(h)})$	21.18	19.06	25.46	1.31	1502	22.69	20.55	25.46	0.94	289
$\log(NLMP^{(s)})$	22.21	20.04	25.9	1.21	1502	23.59	21.74	25.9	0.89	289
$\log(FMP^{(h)})$	20.53	19.06	24.33	1.05	1502	21.61	19.32	24.33	1.29	289
$\log(FMP^{(s)})$	21.62	20.03	25.2	0.98	1502	22.59	20.19	25.2	1.29	289
	Upper middle income countries					*Lower middle income countries*				
	mean	**min**	**max**	**sd**	**n**	**mean**	**min**	**max**	**sd**	**n**
log(GDP pc)	8.9	7.35	11.12	0.55	485	7.84	06.09	9.75	0.57	393
temperature	14.69	-12.73	27.99	8.66	485	19	-5.79	28.26	8.18	393
inv. dist. port	0.05	0	1	0.2	485	0.02	0	1	0.11	393
log(oil production pc)	0.23	0	3.48	0.59	485	0.02	0	1.98	0.12	393
log(density)	3.74	-2.67	10.43	1.82	485	4.24	-1.91	9.56	1.78	393
years education	6.89	2.37	11.45	1.83	485	5.05	0.44	12.5	2.42	393
$\log(MP^{(h)})$	20.95	19.54	25.04	0.76	485	20.31	19.24	23.18	0.53	393
$\log(MP^{(s)})$	21.98	20.75	25.45	0.69	485	21.4	20.31	23.68	0.47	393
$\log(NLMP^{(h)})$	20.91	19.52	24.91	0.75	485	20.3	19.24	23.18	0.53	393
$\log(NLMP^{(s)})$	21.96	20.75	25.34	0.68	485	21.39	20.31	23.68	0.47	393
$\log(FMP^{(h)})$	20.3	19.21	22.03	0.66	485	20.13	19.19	21.86	0.46	393
$\log(FMP^{(s)})$	21.42	20.59	23.05	0.6	485	21.26	20.17	22.78	0.44	393
	Low income countries									
	mean	**min**	**max**	**sd**	**n**					
log(GDP pc)	6.38	1.76	8.56	1.37	134					
temperature	22.8	-1.58	29.15	6.19	134					
inv. dist. port	0.03	0	1	0.17	134					
log(oil production pc)	0	0	0.08	0.01	134					
log(density)	4.29	-0.55	12.01	1.73	134					
years education	2.98	0.22	11.7	2.5	134					
$\log(MP^{(h)})$	19.7	19.07	20.58	0.25	134					
$\log(MP^{(s)})$	20.85	20.06	21.34	0.23	134					
$\log(NLMP^{(h)})$	19.7	19.06	20.58	0.25	134					
$\log(NLMP^{(s)})$	20.85	20.04	21.34	0.23	134					
$\log(FMP^{(h)})$	19.67	19.06	20.22	0.24	134					
$\log(FMP^{(s)})$	20.83	20.03	21.27	0.22	134					

Country income groups are defined following the World Bank classification: *High income* (21 countries in the sample), *Upper middle income* (28 countries), *Lower middle income* (31 countries) and *Low income* (14 countries).

Table 14: Descriptive statistics - 2005

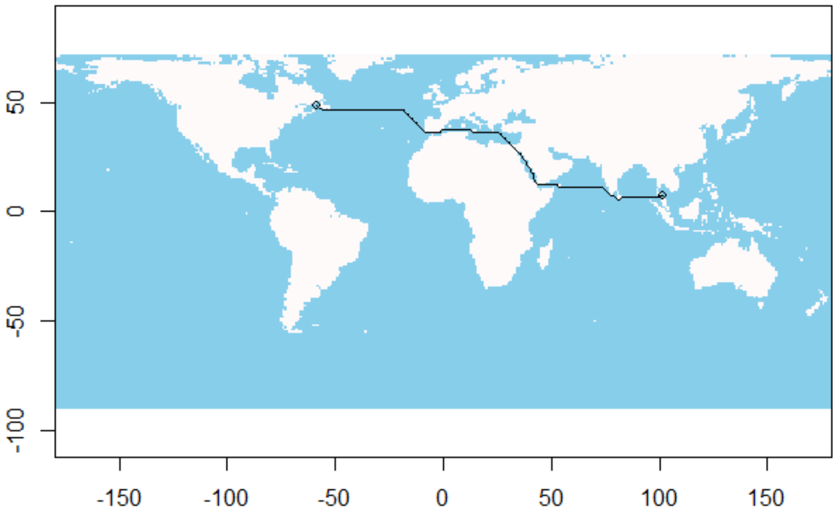

Figure 8: Example of shortest path between ports in Canada and Thailand

	core					*semi-periphery*				
	mean	**min**	**max**	**sd**	**n**	**mean**	**min**	**max**	**sd**	**n**
log(GDP pc)	9.11	2.63	11.42	1.31	142	8.76	1.98	11.24	1.37	379
temperature	16.69	-4.66	29.1	7.85	142	15.69	-9.1	29.15	7.95	379
inv. dist. port	0.08	0	1	0.24	142	0.06	0	1	0.22	379
log(oil per cap.)	0.09	0	3.48	0.45	142	0.05	0	2.42	0.24	379
log(density)	6.17	1.42	11.1	1.86	142	4.77	-0.23	12.01	1.46	379
years education	7.84	0.25	13.21	2.98	142	7.1	0.27	12.85	3.29	379
log($MP^{(h)}$)	21.35	19.09	25.04	1.47	142	21.25	19.07	25.49	1.4	379
log($MP^{(s)}$)	22.29	20.13	25.45	1.31	142	22.26	20.06	25.92	1.28	379
log($NLMP^{(h)}$)	21.17	19.09	24.91	1.42	142	21.21	19.06	25.46	1.39	379
log($NLMP^{(s)}$)	22.19	20.13	25.34	1.28	142	22.24	20.04	25.9	1.28	379
log($FMP^{(h)}$)	20.57	19.08	24.27	1.14	142	20.6	19.06	24.33	1.12	379
log($FMP^{(s)}$)	21.69	20.13	25.17	1.05	142	21.69	20.03	25.2	1.05	379
	periphery									
	mean	**min**	**max**	**sd**	**n**					
log(GDP pc)	8.71	1.76	11.87	1.3	977					
temperature	14.48	-12.73	28.8	8.69	977					
inv. dist. port	0.03	0	1	0.14	977					
log(oil per cap.)	0.12	0	4.16	0.44	977					
log(density)	3.65	-2.93	9.17	1.64	977					
years education	07.03	0.22	12.83	3.12	977					
log($MP^{(h)}$)	21.18	19.13	24.88	1.27	977					
log($MP^{(s)}$)	22.21	20.09	25.54	1.19	977					
log($NLMP^{(h)}$)	21.16	19.12	24.62	1.26	977					
log($NLMP^{(s)}$)	22.2	20.09	25.44	1.18	977					
log($FMP^{(h)}$)	20.49	19.12	24.2	1	977					
log($FMP^{(s)}$)	21.58	20.08	25.11	0.95	977					

Table 15: Descriptive statistics by cluster characteristic

	(1)	(2)	(3)	(4)
	ln education		ln density	
$\mathbb{1}_{\text{semi-periphery}}$	−0.18***	−0.04***	−1.42***	−1.13***
	(0.03)	(0.01)	(0.16)	(0.28)
$\mathbb{1}_{\text{periphery}}$	−0.27***	−0.07***	−2.42***	−1.99***
	(0.04)	(0.02)	(0.16)	(0.29)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{upper-middle income group}}$		−0.09***		−0.51
		(0.03)		(0.41)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{upper-middle income country}}$		−0.16***		−0.88**
		(0.03)		(0.39)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{lower-middle income country}}$		−0.17***		−0.14
		(0.06)		(0.40)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{lower-middle income country}}$		−0.28***		−0.24
		(0.09)		(0.40)
$\mathbb{1}_{\text{semi-periphery}} \times \mathbb{1}_{\text{low income country}}$		−0.55***		−0.89
		(0.17)		(0.68)
$\mathbb{1}_{\text{periphery}} \times \mathbb{1}_{\text{low income country}}$		−0.67***		−0.82
		(0.19)		(0.55)
Num. obs.	1479	1479	1516	1516
Num. groups: code	102	102	104	104
Adj. R^2 (full model)	0.90	0.91	0.54	0.55
Adj. R^2 (proj model)	0.14	0.20	0.27	0.28

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The dependant variable is the GDP per capita (in log). Robust standard errors adjusted for clustering on each country are in parentheses. Estimations include country fixed-effects.

Table 16: Core and Periphery Divide, Education and Density

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln \text{MP}^{(h)}$	$\ln \text{MP}^{(s)}$	$\ln \text{NLMP}^{(h)}$	$\ln \text{NLMP}^{(s)}$	$\ln \text{FMP}^{(h)}$	$\ln \text{FMP}^{(s)}$
$\mathbb{1}_{\text{semi-periphery}}$	−0.20***	−0.13***	−0.05	−0.05	−0.00	−0.02*
	(0.06)	(0.05)	(0.06)	(0.05)	(0.01)	(0.01)
$\mathbb{1}_{\text{periphery}}$	−0.36***	−0.25***	−0.18***	−0.15***	−0.02**	−0.05***
	(0.07)	(0.06)	(0.07)	(0.05)	(0.01)	(0.01)
Num. obs.	1498	1498	1498	1498	1498	1498
Num. groups: code	103	103	103	103	103	103
Adj. R^2 (full model)	0.91	0.93	0.91	0.93	0.98	0.98
Adj. R^2 (proj model)	0.07	0.05	0.03	0.03	0.00	0.01

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The dependant variable is the GDP per capita (in log). Robust standard errors adjusted for clustering on each country are in parentheses. Estimations include country fixed-effects.

Table 17: Core and Periphery Divide, Market Potential

	(1)	(2)	(3)	(4)	(5)	(6)
market potential	0.22***	0.22***	0.14***	0.15***	0.13	0.12
	(0.05)	(0.06)	(0.04)	(0.05)	(0.11)	(0.13)
Num. obs.	1,502	1,502	1,502	1,502	1,502	1,502
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Num. groups: code	105	105	105	105	105	105
Adj. R^2 (proj model)	0.06	0.04	0.02	0.01	0.00	0.00
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The table displays the estimated coefficients from equation 6. Robust standard errors adjusted for clustering on each country are in parentheses. $MP^{(h)}$ considers physical distance between regions measured as the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 18: Regional Development and Market potential - Univariate regressions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
market potential	0.16**	0.16**	0.11**	0.11**	0.03	−0.02
	(0.06)	(0.07)	(0.05)	(0.06)	(0.12)	(0.10)
education +65 years old	0.22***	0.22***	0.22***	0.22***	0.22***	0.22***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
Num. obs.	607	607	607	607	607	607
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Num. groups: code	39	39	39	39	39	39
Adj. R^2 (proj model)	0.38	0.37	0.37	0.37	0.36	0.36
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses.

Table 19: Regional development and Market Potential - Education of old

	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$	N regions
AUS	-0.12	-0.23	-0.24	-0.23	0.09	0.39	8
AUT	-3.07	-2.43	-2.11	-2.12	0.31	0.20	9
BEL	0.03	-0.21	-0.25	-0.39	-2.24	-2.83	11
CAN	-0.22	-0.23	-0.26	-0.25	-0.16	-0.16	12
CHE	0.41	0.50	0.31	0.42	-0.70	0.74	25
CZE	-0.82	-1.04	-0.80*	-0.97*	-3.08	0	8
DEU	0.28	0.17	0.01	-0.00	-0.09	-0.10	16
DNK	0.57*	0.79**	0.36	0.63**	0.09	0.50	15
ESP	0.07	-0.00	0.05	-0.00	0.12	-0.02	19
EST	0.36	-0.25	0.19	-0.25	0.28	-0.40	15
FIN	-	-	-	-	-	-	5
FRA	0.02	-0.10	-0.06	-0.14	-0.00	-0.09	22
GBR	-0.02	-0.09	-0.09	-0.12	-0.25	-0.27	12
GRC	0.57	0.15	0.38	0.03	-0.83	-0.25	13
HRV	0.52**	0.72**	0.52**	0.72**	0.98**	1.16**	20
HUN	-	-	-	-	-	-	7
IRL	-	-	-	-	-	-	2
ISR	-	-	-	-	-	-	6
ITA	0.43**	0.29	0.37**	0.26	0.57***	0.54**	20
JPN	-0.01	-0.02	-0.03	-0.03	0.19	0.09	47
NLD	-1.21	-1.01	-1.13	-0.97	0.45	0.62	12
NOR	-0.06	-0.09	-0.06	-0.09	-0.00	-0.25	19
NZL	-0.04	-0.09	-0.09	-0.09	-8.57	4.69	14
POL	-0.04	-0.12	-0.19	-0.21	-0.12	-0.18	16
PRT	-0.21**	-0.26**	-0.33**	-0.37**	-0.48*	0	7
SVK	-0.27	-0.40	-0.23	-0.35	0.36	0.63	8
SVN	0.30	0.37	0.02	0.17	0.82	0.79	12
SWE	-0.11	-0.15	-0.11	-0.14	0.20	0.31	8
USA	-0.00	-0.03	-0.02	-0.04	-0.30	-0.27	51

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. These coefficients are elasticity coefficients of regional GDP per capita to market potential indexes for each high income country. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c. List of countries: Australia, Austria, Belgium, Canada, Switzerland, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, Great Britain, Greece, Croatia, Hungary, Ireland, Israel, Italy, Japan, Netherlands, Norvege, New Zeland, Poland, Portugal, Slovakia, Slovenia, Sweden, United States of America.

Table 20: Market Potential Elasticity - High income countries (2005)

	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$	N regions
AZE	-0.27	-0.56	-0.27	-0.55	10.89	15.10	11
BGR	-	-	-	-	-	-	6
BIH	-	-	-	-	-	-	3
BRA	0.19	0.21	-0.01	0.04	-0.55	-0.42	27
CHL	-0.19	-0.31	-0.22	-0.34	-1.28	-0.33	13
CHN	0.70***	0.69**	0.42	0.48	1.44*	-0.61	31
COL	0.64**	0.88*	0.56*	0.78*	2.47**	4.68**	33
DOM	0.04	0.20	0.57	0.84	-1.60	0.44	9
ECU	-1.44	-2.84**	-1.03	-2.82**	-2.71*	-2.84***	21
GAB	-	-	-	-	-	-	4
IRN	0.54	0.41	0.11	0.10	0.13	0.67	30
JOR	-0.18	-0.52	-0.20	-0.53	0.06	-0.18	12
KAZ	-	-	-	-	-	-	6
LBN	-	-	-	-	-	-	6
LTU	-0.63	-0.71	-0.46	-0.60	-1.10	-1.55	10
LVA	0.21	0.21	0.17	0.16	2.58	-2.60	33
MEX	-0.04	-0.35	-0.26	-0.41	0.34	-0.55	32
MKD	-0.95	-1.14	-0.96	-1.15	-0.99	-1.17	8
MYS	-0.07	-0.15	-0.11	-0.17	-2.78	-3.10	14
PAN	-1.46*	-2.32**	-1.56**	-2.34***	5.11	2.07	9
PER	0.50	0.67	0.13	0.40	0.88	1.70	24
ROU	1.06	1.60	1.08	1.64	2.00*	2.33	8
RUS	0.27***	0.25**	0.24***	0.23**	0.73***	0.63**	80
THA	-	-	-	-	-	-	5
TUR	0.32	0.43	0.34	0.46	0.20	0.42	12
URY	0.21	0.26	0.20	0.22	0.22	0.42	19
VEN	0.03	0.03	-0.01	-0.01	0.43	0.40	23
ZAF	0.07	0.03	-0.03	-0.0	5.79	10.70	9

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. These coefficients are elasticity coefficients of regional GDP per capita to market potential indexes for each upper middle income country. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c. List of countries: Azerbaijan, Bulgaria, Bosnia and Herzegovina, Brazil, Chile, Schina, Columbia, Dominican Republic, Ecuador, Gabon, Iran, Jordania, Kazakhstan, Liban, Lituania, Latva, Mexico, Macedonia, Malaysia, Panama, Peru, Romania, Russia, Thailand, Turkey, Uruguay, Venezuela, South Africa.

Table 21: Market Potential Elasticity - Upper middle income countries (2005)

	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$	N regions
ARM	0.22	0.44	0.21	0.44	0.36	0.43	11
BLZ	-	-	-	-	-	-	6
BOL	-11.12	-11.33	-28.04	-16.17	20.23	8.39	9
CMR	-3.29	-3.31	-3.20	-3.28	-3.89	-4.12	10
EGY	-	-	-	-	-	-	2
GEO	-3.16	-5.79	-3.10	-5.56	-3.96	-5.32	11
GHA	-8.39	0.10	-8.49*	0.09	-1.67**	3.22	10
GTM	-0.24	-0.40	-0.24	-0.40	1.32	3.23	8
HND	-3.66**	1.22	-3.49	1.11	-3.19**	1.13	18
IDN	0.94**	0.87*	0.68	0.63	0.69	0.70	33
IND	0.19*	0.26*	0.17*	0.26*	-0.01	0.08	32
LAO	0.10	0.77	0.06	0.73	-0.18	0.66	18
LKA	-0.37	-0.58	-0.43	-0.60	0.17	-2.12	9
LSO	-0.84	-1.30	-0.86	-1.32	-0.89	-1.30	10
MAR	-0.57	-0.64	-0.57	-0.64	-0.48	-0.56	14
MDA	-	-	-	-	-	-	5
MNG	0.41	-1.07	0.05	-1.20	-0.21	-1.29	22
NGA	-	-	-	-	-	-	6
NIC	-2.50	-1.08	-2.49**	-1.15	-1.55	-0.81	17
PAK	-	-	-	-	-	-	5
PHL	-0.02	-0.05	-0.03	-0.05	0.30	-0.14	17
PRY	-4.11	-6.72	-6.45**	-7.97**	-4.24	-4.95	18
SEN	5.43	4.90	5.48	4.92	7.81*	5.87	10
SLV	-0.21	-0.10	-0.20	-0.12	-0.53	-0.03	14
SWZ	-	-	-	-	-	-	4
SYR	-0.52	-0.74	-0.51	-0.72	-0.61*	-0.98	13
UKR	-0.22	0.30*	0.21*	0.30*	-1.85**	-1.61	27
UZB	-	-	-	-	-	-	5
VNM	1.20	-3.01	-2.89	-6.28	-2.95	-6.95	8

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. These coefficients are elasticity coefficients of regional GDP per capita to market potential indexes for each lower middle income country. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c. List of countries: Armenia, Belize, Bolivia, Cameroon, Egypt, Georgia, Ghana, Guatemala, Honduras, Indonesia, India, Laos, Morocco, Moldova, Mongolia, Nigeria, Nicaragua, Pakistan, Philippines, Paraguay, Senegal, El Salvador, Swaziland, Syria, Ukrania, Uzbekistan, Vietnam.

Table 22: Market Potential Elasticity - Lower middle income countries (2005)

	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$	N regions
BFA	0.34	-5.20	0.05	-6.69	1.91	6.27	13
COD	-3.08	-5.01	-3.34	-5.00	-3.05	-4.99	11
KEN	-0.12	-0.20	-0.15	-0.23	0.98	-.78	8
KGZ	3.18	5.88*	3.16	5.87	3.05*	5.57	8
KHM	1.24	1.54	1.13	1.48	1.15	1.46	15
MDG	-	-	-	-	-	-	6
MOZ	0.10	0.08	0.11	0.08	0.05	0.02	10
MWI	-	-	-	-	-	-	3
NER	-0.94	-0.54	-0.97	-0.56	-0.98	-0.56	8
NPL	-	-	-	-	-	-	5
TZA	-1.05	-0.95	-1.14	-1.00	-1.46	-1.36	21
UGA	-	-	-	-	-	-	4
ZWE	5.68***	6.99**	5.68***	6.99**	5.68**	6.99**	10

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. These coefficients are elasticity coefficients of regional GDP per capita to market potential indexes for each low income country. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c. List of countries: Benin, Burkina Faso, Kenya, Kyrgyzstan, Cambodia, Madagascar, Mozambic, Malawi, Niger, Nepal, Tanzania, Uganda, Democratic Republic of the Congo, Zimbabwe.

Table 23: Market Potential Elasticity - Low income countries (2005)

	(1)	(2)	(3)	(4)	(5)	(6)
market potential	0.12^{***}	0.08^{**}	0.05^{**}	0.04	0.05	-0.02
	(0.04)	(0.03)	(0.02)	(0.03)	(0.07)	(0.09)
centrality	0.05^{***}	0.05^{***}	0.06^{***}	0.06^{***}	0.06^{***}	0.06^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Num. obs.	1464	1464	1464	1464	1464	1464
Num. groups: code	103	103	103	103	103	103
Adj. R^2 (full model)	0.95	0.95	0.95	0.95	0.95	0.95
Adj. R^2 (proj model)	0.44	0.43	0.43	0.43	0.43	0.42
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Market potential and centrality indexes are introduced in the logarithm form. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 24: Market Potential and Centrality (2005)

	(1)	(2)	(3)	(4)	(5)	(6)
market potential	0.17^{***}	0.17^{***}	0.12^{***}	0.13^{***}	0.11^{*}	0.05
	(0.04)	(0.04)	(0.03)	(0.04)	(0.06)	(0.06)
market potential $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.04^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
market potential $\times$ $\mathbb{1}_{\text{periphery}}$	-0.03^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}	-0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
centrality	-0.00	-0.00	0.00	0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
centrality $\times$ $\mathbb{1}_{\text{semi-periphery}}$	-0.17^{***}	-0.16^{***}	-0.15^{***}	-0.15^{***}	-0.11^{***}	-0.11^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
centrality $\times$ $\mathbb{1}_{\text{periphery}}$	-0.12^{**}	-0.13^{**}	-0.10^{*}	-0.11^{**}	-0.07	-0.07
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)
Num. obs.	1460	1460	1460	1460	1460	1460
Num. groups: code	101	101	101	101	101	101
Adj. R^2 (full model)	0.95	0.95	0.95	0.95	0.95	0.95
Adj. R^2 (proj model)	0.48	0.48	0.47	0.4708	0.46	0.46
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Market potential and centrality indexes are introduced in the logarithm form. In each regression, I control for the temperature, the inverse distance to the closest port, oil production per capita, density and average level of education. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 25: Market Potential and Centrality - Core and Periphery (2005)

	(1)	(2)	(3)	(4)	(5)	(6)
1995 sample						
market potential	0.10**	0.08*	0.03	0.02	0.13*	0.08
	(0.05)	(0.05)	(0.04)	(0.04)	(0.07)	(0.09)
2000 sample						
market potential	0.10**	0.09*	0.04	0.04	0.22***	0.18*
	(0.05)	(0.05)	(0.04)	(0.05)	(0.07)	(0.09)
2005 sample						
market potential	0.11***	0.09**	0.06**	0.05*	0.08	−0.00
	(0.03)	(0.04)	(0.03)	(0.03)	(0.08)	(0.10)
2010 sample						
market potential	0.01	−0.07*	−0.06	−0.11**	−0.05	−0.23**
	(0.03)	(0.04)	(0.03)	(0.03)	(0.08)	(0.10)
1995-2005 sample						
market potential	0.10***	0.08***	0.04**	0.03	0.13***	0.06
	(0.02)	(0.03)	(0.02)	(0.02)	(0.05)	(0.06)
1995-2010 sample						
market potential	0.08**	0.05*	0.02	0.00	0.09**	−0.01
	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.05)
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country-time group are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. For the cross-sectional estimations of each year, country fixed effects are included. For the panel estimations, country-year fixed effects are included. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.
Countries in the 1995 sample: ARG, SVN, NOR, PAK, PAN, PER, POL, PRT, SLV, SWE, THA, TUR, TZA, URY, AUS, AUT, BEL, BGR, BOL, BRA, CAN, CHE, CHL, CHN, COL, DEU, ECU, ESP, FIN, FRA, GRC, GTM, HND, IDN, IND, IRL, ITA, JOR, JPN, KAZ, LKA, LSO, LTU, LVA, MEX, MNG, MYS, NLD. The sample gathers 738 regions in 48 countries.
Countries in the 2000 sample: ARG, SVN, NOR, PAN, PER, POL, PRT, PRY, SLV, SRB, SWE, THA, TUR, TZA, URY, ALB, ARE, AUS, AUT, BEL, BEN, BGR, BOL, BRA, CANn CHE, CHL, CHN, COL, DEU, ECU, ESP, EST, FIN, FRA, GRC, HND, HRV, HUN, IND, IRL, ITA, JOR, JPN, KAZ, LKA, LSO, LTU, LVA, MEX, MKD, MNG, MOZ, MYS, NLD. The sample gathers 802 regions in 55 countries.
Countries in the 2005 sample: ARG, SVN, NOR, NPL, PAK, PAN, PER, POL, PRT, PRY, RUS, SLV, SVK, SWE, THA, TUR, TZA, UKR, URY, USA, UZB, VNM, ARE, AUS, AUT, BEL, BEN, BGR, BIH, BOL, BRA, CAN, CHE, CHL, CHN, COL, CZE, DEU, ECU, ESP, EST, FIN, FRA, GBR, GRC, GTM, HND, HRV, HUN, IDN, IND, IRL, IRN, ITA, JOR, JPN, KAZ, KGZ, LKA, LSO, LTU, LVA, MEX, MKD, MNG, MOZ, MYS, NGA, NLD. The sample gathers 1129 regions in 69 countries.
Countries in the sample: SVN, NOR, PAN, PRT, RUS, SVK, SWE, USA, VNM, BGR, BRA, CHE, CHN, CZE, FIN, IDN, JPN, KOR, MEX, NLD. The sample gathers 421 regions in 20 countries.

Table 26: Market potential elasticity coefficients - panel

	(1)	(2)	(3)	(4)	(5)	(6)
1995-2005 sample						
market potential	0.09^{**}	0.08^{**}	0.05	0.05	0.11^{*}	0.06
	(0.04)	(0.04)	(0.03)	(0.03)	(0.06)	(0.07)
market potential × $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential × $\mathbb{1}_{\text{periphery}}$	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.03^{***}	-0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
1995-2010 sample						
market potential	0.08^{**}	0.05	0.03	0.01	0.08	0.00
	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
market potential × $\mathbb{1}_{\text{semi-periphery}}$	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
market potential × $\mathbb{1}_{\text{periphery}}$	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.03^{***}	-0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Regressor	$MP^{(h)}$	$MP^{(s)}$	$NLMP^{(h)}$	$NLMP^{(s)}$	$FMP^{(h)}$	$FMP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country-time group are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. For the cross-sectional estimations of each year, country fixed effects are included. For the panel estimations, country-year fixed effects are included. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures. $NLMP^{(h)}$ and $NLMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the local market, i.e. $y_i\tau_{ii}$. $FMP^{(h)}$ and $FMP^{(s)}$ are proxies for $MP^{(h)}$ and $MP^{(s)}$ respectively, excluding the domestic markets, i.e. $\sum_{j\neq i} y_j\tau_{ij}$ with i and j both in country c.

Table 27: Market potential elasticity coefficients - panel - core-periphery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
market potential	0.14^{***}	0.12^{***}	0.15^{***}	0.12^{***}	0.15^{***}	0.13^{**}	0.16^{***}	0.13^{**}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
$\text{centrality}^{\text{cores}}$	−0.11	−0.09						
	(0.07)	(0.06)						
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{core})$			−0.37	−0.12				
			(0.26)	(0.22)				
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$			-0.13^{**}	-0.10^{*}				
			(0.06)	(0.06)				
$\text{centrality}^{\text{cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$			−0.10	−0.08				
			(0.08)	(0.08)				
$\text{centrality}^{\text{domestic cores}}$					−0.00	0.01		
					(0.02)	(0.02)		
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{core})$							−0.37	−0.12
							(0.28)	(0.24)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$							−0.02	−0.01
							(0.02)	(0.02)
$\text{centrality}^{\text{domestic cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$							0.01	0.03
							(0.02)	(0.02)
$\text{centrality}^{\text{foreign cores}}$					-0.42^{**}	-0.39^{**}		
					(0.18)	(0.18)		
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{core})$							-0.32^{*}	−0.27
							(0.18)	(0.17)
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{semi-periphery})$							-0.43^{**}	-0.40^{**}
							(0.17)	(0.17)
$\text{centrality}^{\text{foreign cores}} \times \mathbb{1}(\gamma_g = \text{periphery})$							-0.44^{**}	-0.41^{**}
							(0.19)	(0.19)
Num. obs.	1460	1460	1460	1460	1460	1460	1460	1460
Country and Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. groups: code	101	101	101	101	101	101	101	101
Adj. R^2 (proj model)	0.32	0.31	0.32	0.31	0.33	0.32	0.34	0.33
Regressor	$MP^{(h)}$	$MP^{(s)}$	$MP^{(h)}$	$MP^{(s)}$	$MP^{(h)}$	$MP^{(s)}$	$MP^{(h)}$	$MP^{(s)}$

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust standard errors adjusted for clustering on each country are in parentheses. The following covariates are included: temperature, inverse distance to the closest port, oil production per capita and average educational level. The market potential and centrality variables are entered with the logarithm in the regressions. $MP^{(h)}$ considers physical distance between regions measured by the haversine distance, cultural proximity measures depending on common language, national common border, colonial ties, and trade facilities implied by a common currency and regional trade agreements. $MP^{(s)}$ considers the shortest path by land and sea between regions, passing through their closest ports if needed, in addition to the cultural proximity and trade facilities measures.

Table 28: Regional Development, the Core and Periphery, and Centrality to cores (2)