Politecnico di Torino

Collegio di Ingegneria Gestionale e della Produzione

Class LM31

Master's Degree in Engineering and Management

'MADE NEAR ITALY': HOW NEARSHORING IS REDEFINING MANUFACTURING TRADE

Exploring how proximity and trade dynamics shape Italian manufacturing strategies, 2008-2019

Thesis supervisor:
Anna D'Ambrosio

Candidate:
Ilaria Auda Gioanet

Anno Accademico 2021-2022

Table of contents

Introduction

In a world where global supply chains stretch across continents and geopolitical shifts redefine market dynamics overnight, companies are increasingly rethinking their strategies for production and trade. Nearshoring has emerged as a powerful strategy—not just as a response to the pressures of cost and efficiency, but as a way to navigate the volatile pulse of modern economics.

Italy, a nation historically at the crossroads of trade routes, finds itself in this strategic shift, navigating the complexities of bringing production closer to home.

This Thesis dives into the heart of Nearshoring, focusing on how Italy, with its unique economic and geopolitical landscape, has adapted to this trend between 2008 and 2019. The goal is dual: to define the driving forces behind Italy’s nearshoring decisions and to build a robust econometric model that quantifies the impact of these factors on international trade, with a spotlight on manufacturing imports.

The present work begins by untangling the concepts of offshoring, reshoring, and nearshoring, setting the stage with a comprehensive look at the literature. Theoretical frameworks, including the gravity model of trade and Ghemawat's CAGE framework, are utilized to examine the non-economic factors, such as cultural, administrative, and geographic distance, that influence trade and production decisions.

In the second part, attention shifts to the construction of the econometric model, meticulously designed to capture the differences of Italy's trade strategies and the complex interplay of factors influencing nearshoring. The model leverages an extensive dataset of Italian manufacturing firms engaged in international trade, spanning more than a decade of economic activity. Key independent variables—such as GDP per capita, geographic distance, time zones, shared currency, and regional trade agreements—are incorporated into the model to assess their influence on the nearshoring and offshoring behavior of Italian firms. The model's goal is to measure how much different factors influence the volume and direction of offshoring and nearshoring decisions. This approach helps clarify which factors are most important in driving Italy's trade and production relocation strategies.

Finally, the third part of the thesis presents an analysis of the results, uncovering critical insights into the factors that have influenced Italy’s nearshoring and offshoring activities over the studied period.

Monitoring shifts in nearshoring and offshoring across specific time frames and countries is essential for understanding broader implications on trade patterns and economic resilience. This thesis aims to demonstrate that by focusing on particular contexts, valuable insights can be gained to guide more informed decision-making and ensure competitiveness in an increasingly interconnected global economy.

1. Near to the origin: Nearshoring and its framework

1.1. Definitions Offshoring, Reshoring and Nearshoring

Since the late 1980s, the Italian business landscape has experienced an important evolution, driven by the opportunities offered by an increasingly interconnected world. In this transformative journey, businesses have been forced to review traditional approaches in light of new strategies aimed at reinforcing the operational efficiency, cost reduction, and sustainable growth trajectories of operations. Central in that narrative is the rise of offshoring as one of the keys that reshapes the profile of international commerce.

Offshoring, which rose to prominence in the early 2000s, entailed the strategic relocation of business processes and operations from high-cost regions to more economically advantageous destinations across international borders, notably in regions like Asia. This strategic plan, exemplified by the widespread availability of "Made In China" products or the presence of "Made in Bangladesh" merchandise on store shelves, became emblematic of the global business panorama, representing the search for cost-effectiveness and access to diverse markets and resources.

This practice is sometimes confused with Outsourcing, which involves purchasing products from external suppliers that were previously produced internally within the company. If these external suppliers are located in foreign countries, then it also constitutes offshoring.

After many years of trading across continents, businesses started to move to a counter-trend, inducing the creation of new conceptual frameworks like reshoring and nearshoring. Recently, these phenomena have gained relevance not only among governments and firms but also in public opinion, popularizing neologisms coined to describe such tendencies.

The most known is Reshoring, also called Back-shoring, that is the practice of bringing back business operations, manufacturing processes, or services to the domestic country from which they were previously offshored.

Unlike its predecessors, Nearshoring involves the relocation of business activities to nearby, often neighboring, countries. This strategic pivot is promoted by a multiplicity of factors, which vary across industries and sectors. According to The Economist (2005),

nearshoring refers to the practice of relocating business activities “to countries that are quite cheap and very close rather than very cheap and far away”.

The shift is noticeable across Western Europe. Rather than relying on East Asia for production, Italian manufacturers are increasingly establishing operations in Turkey, Romania, and the Balkans. While these regions may present some uncertainties, collaborating with them is considered a more secure option compared to dealing with factories located on the other side of the globe.

Alongside nearshoring, other similar frameworks have emerged, such as Allyshoring and Friendshoring. These phenomena have gathered increased attention from governments, particularly in light of recent global events. For instance, tensions between China and the USA, or Russia's suspension of gas supplies in attempts to influence the EU's stance on Ukraine, have underscored the importance of strategic partnerships and alliances in international trade and economic relations.

Importing goods from foreign countries, often a target of government scrutiny, remains a crucial pillar of a nation's production ecosystem. This is especially evident in the need to import raw materials when certain resources are either unavailable domestically or insufficient to meet local demand. For example, the EU's production only accounts for 4% of the global supply chain of critical materials used in digital equipment production (Strategic Foresight Report EU, 8). However, the recognition of the crucial aspect of securing access to critical raw materials is paramount, with the EU striving for open strategic autonomy throughout the entirety of the value chain. Recent military aggression against Ukraine has further underscored the significance of EU autonomy, particularly in the energy sector, prompting member states to embrace the nearshoring movement.

Overall, all shoring phenomena represent strategic adjustments aimed at optimizing business operations in response to evolving global trends and challenges.

These strategies are often mixed within the same company and are decided based on the specific components or products considered and their quality, cost, and timing.

1.1.1. The role of Globalization in shaping strategic relocation

The rise of offshoring was deeply interlaced with the broader process of globalization, which facilitated the seamless flow of goods, capital, information, and people across

borders. This process has been propelled by advancements in technology, reductions in trade barriers, and the liberalization of markets, which collectively have created a more integrated global economy.

According to the DHL Global Connectedness Report 2024, globalization, despite recent challenges, has not reversed but evolved, reaching new heights in terms of international connectivity. This persistent global interconnectedness has been the driving force behind companies' decisions to offshore production, taking advantage of cost efficiencies and the availability of resources in distant regions.

Despite numerous global crises, the report reveals that globalization has not only persisted but reached record levels. International flows remained robust in 2022 and 2023, highlighting the resilience of global connections in the face of significant challenges.

One of the key findings of the report is that, contrary to what one might think, globalization has not given way to regionalization on a broad scale. While there is evidence of increased regionalization in specific areas, such as North America, where nearshoring trends are particularly prominent, the overall global flows continue to span considerable distances. This suggests that while companies and nations may be reconfiguring their strategies to address new risks, they are not abandoning global markets.

In particular, the report identifies that North America is the only region where nearshoring has clearly taken root. Companies in the United States, for example, have increasingly shifted their production activities to nearby countries such as Mexico, leveraging geographic proximity to enhance supply chain resilience and reduce costs.

For Italy and other European countries, the report underscores the potential benefits of embracing similar strategies. Although the European region has not seen the same level of nearshoring as North America, the increasing importance of regional trade partnerships and the need for greater supply chain security suggest that nearshoring could become a more significant trend in Europe. By relocating production to neighboring countries within the Euro-Mediterranean region, Italian businesses could enhance their competitiveness, reduce their exposure to global disruptions, and respond more effectively to market changes.

The "DHL Global Connectedness Report 2024" also highlights that while international flows remain strong, they are still far from reaching their full potential. The global connectedness index, a DHL measure of the depth of global integration, indicates that international flows account for only 25% of all economic activity, with domestic flows still

dominating the global economy. This suggests that there is significant room for growth in international trade and investment, provided that companies and governments can navigate the challenges posed by the current geopolitical landscape.

1.1.2. An example: Benetton's case

An explanatory case of these varied approaches can be presented through the production and supply chain of Benetton. After various experiences with offshoring, in 2016 Benetton decided to adopt the reshoring and nearshoring strategies, bringing back to Italy a portion of its production. In recent times, the Italian fashion firm has shifted over 10% of its production output from countries like Bangladesh, Vietnam, China, and India to European production locations. Looking ahead, Benetton plans to further reduce its production in Asia, aiming to increase manufacturing in countries such as Croatia, Serbia, Turkey, Egypt, and Tunisia. The different strategies adopted by the Benetton company can be observed in the explicative Image 1.1.

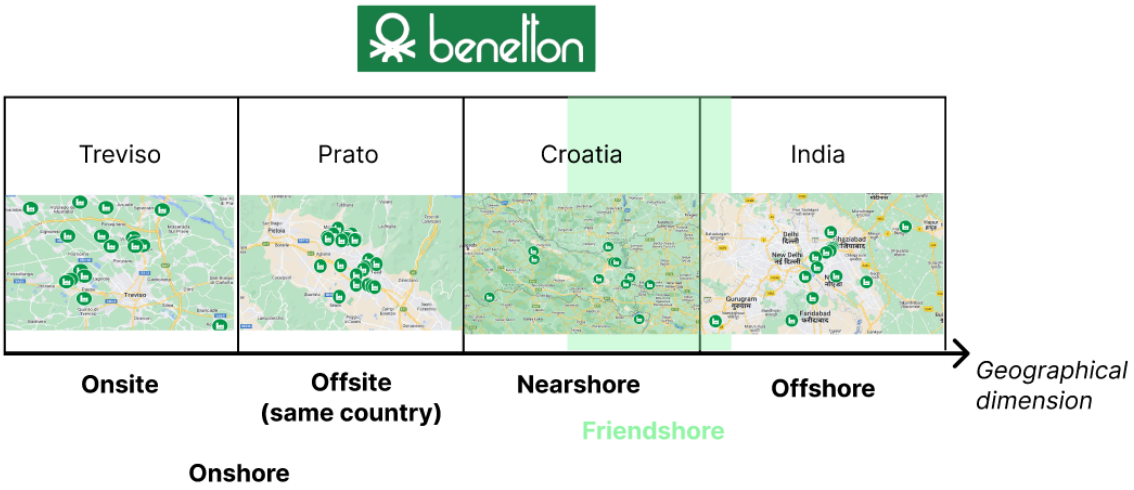

Image 1.1.: Points from the Maps of Benetton's suppliers around the world and the relationship with Onshoring, Nearshoring, Offshoring

1.2. Literature analysis

The phenomenon of nearshoring is still relatively underrepresented in literature, often analyzed not individually but alongside its counterpart, reshoring. The definition of nearshoring is present in a rather limited body of literature, and the limits in terms of distances between near and offshoring is not defined.

Literature provides the analysis of the factors implicit in distance: the distance between locations encompasses not only geographic proximity, but psychic distance as well. Literature also provides a comprehensive model that relates trades and distances within two countries and their measure of the economies, this is known as the Gravity model of trade.

1.2.1.Psychic distance

The concept of psychic distance was first introduced to the international economics literature by Beckerman (1956) who explored the effects of economic distance (such as transportation costs and physical distance) on intra-European trade. Since then, it has become a prominent concept in international business literature. As defined by Johanson and Vahlne (2009), psychic distance is the sum of factors preventing the flow of information from and to the market. Examples include disparities in language, education, business practices, culture, and industrial development. Psychic and physical distances often align: the farther a host country is from a home country, the greater the psychic gap. However, there might be exceptions to this correlation between psychic and geographic distances. For instance, despite similar geographic distances between Italy and Spain and between Italy and Tunisia, cultural factors such as language and religion may make Spain feel closer to Italy than Tunisia due to a smaller psychic distance.

In Table 1.1. are reported some variables studied by the literature and affecting the psychic distance.

Factor	Literature studies	Results
Time Zone	*Dow and Karunaratna, 2006*	Time zone differences complicate communication and problem-solving between regions with divergent working hours, posing a challenge for international managers despite advancements in telecommunications.
Common language	*Hutchinson, 2002*	Highlights that the level of English proficiency significantly affects services trade, whereas the impact of other languages like French and German appears to be less pronounced and varies.
	Hejazi and Ma, 2011	Highlighted that countries sharing a common official language communicate more easily, facilitating stronger business relations between them, as measured by Foreign Direct Investments.
Colonial ties	*Linders et al., 2005;*	Colonial history between countries often causes a reduction in administrative distance by establishment of harmonized institutions, laws, property rights, etc., thus decreasing the associated costs.
	Head et al., 2010	This study examines post-colonial trade dynamics. It reveals that after a few decades of independence, trade with former colonial powers diminishes significantly, while hostile separations lead to immediate trade reductions. Trade among former colonies also declines gradually, suggesting a loss of

		trading capital over time.
Preferential Trade Agreements	*Kenichi, 2003*	Present a quantitative simulation analysis on the impact of Japan's Free Trade Agreements in Asia using a Computable General Equilibrium model of global trade. It is proposed that promoting liberalization and facilitation is crucial for fostering economic partnerships in Asia.
	Baier and Bergstrand, 2007	It highlights the limitations of the gravity equation by considering trade policy as an exogenous variable. It addresses the endogeneity of Free Trade Agreements (FTAs) econometrically, demonstrating that the effect of FTAs on trade flows is quintupled. On average, an FTA approximately doubles the bilateral trade between two member countries after 10 years.
Pioneering advantage	*Evans and Mavondo, 2002*	Firms based in high-income countries may be more inclined to engage with and achieve success in middle-income and low-income countries.
Income Parity	*Mitra and Golder, 2002*	Similarity in levels of per capita income and economic development facilitates increased engagement between countries.

Tab 1.1.: Literature sources analyzing components of psychic distance

1.2.2. CAGE framework by Ghemawat

The CAGE framework, introduced by Ghemawat in 2001, offers a similar perspective on distance and its components. It delineates four dimensions through which the distance between two countries can manifest:

Cultural Distance: This dimension significantly influences consumer product preferences and interactions. Factors such as religious beliefs, race, social norms, and language create variations in consumer behavior and can generate distance between nations. For instance, trade tends to be more robust between countries sharing a common language.

Administrative or Political Distance: Historical and political ties between countries have a profound impact on trade. Relationships like those between colonizers and colonies can substantially boost trade. Additionally, preferential trading agreements, shared currencies, and political unions foster greater trade collaboration. The integration within entities like the European Union exemplifies deliberate efforts to reduce administrative and political distance among member states.

Geographic Distance: This dimension directly affects transportation and communication costs, making it particularly relevant for businesses dealing with bulky products or requiring extensive coordination across dispersed locations.

Economic Distance: Disparities in consumer wealth or income create significant economic distance between nations, influencing trade levels and partner preferences. Research indicates that wealthier countries engage in more cross-border economic activities, particularly with other affluent nations. However, economic disparities also impact the cost and quality of resources, affecting businesses differently based on their reliance on economies of scale and standardization or economic arbitrage strategies.

These dimensions play a crucial role in the analysis of nearshoring phenomenon.

1.3. Gravity Model of trade

The nearshoring phenomenon is mainly driven by geographic distance, a key factor in the Gravity Model of trade (Gravity Equations: Workhorse, Toolkit, and Cookbook' by K. Head and T. Mayer). Inspired by the fundamental principles of the physical law of gravity, these equations serve as models to depict the intricate flows of trade between nations, leveraging economic size and distance as key determinants. At the core of gravity

equations lies the principle that exports rise proportionately with the economic size of the destination country, while imports increase in proportion to the size of the origin economy. Moreover, these equations highlight a robust negative relationship between physical distance and trade intensity, emphasizing the significance of geographical proximity in shaping trade flows.

Gravity equations encompass several definitions, each offering unique insights into bilateral trade dynamics. The Gravity Model is often introduced in an intuitive manner to help grasp its basic concept. A simplified and commonly used version of the model is presented in the Equation 1.1.

$$X_{ij} = G \times \frac{GDP_i \times GDP_j}{D_{ij}^{\beta}}$$

Equation 1.1.: Intuitive Formulation of the Gravity Model

Where:

- X_{ij} represents the bilateral trade flow between country i and country j.
- GDP_i and GDP_j are the gross domestic products of countries i and j, respectively.
- D_{ij}^{β} is the geographic distance between the two countries.
- G is a constant term often referred to as the "gravitational constant." It captures overall factors influencing trade flows but remains constant across different trade relationships within the same cross-section of data.
- β is the elasticity of trade with respect to distance.

This formulation is useful for introducing the basic concept of the Gravity Model, highlighting how trade volumes between two countries are directly proportional to their economic sizes (measured by GDP) and inversely proportional to the distance between them. However, it's important to note that this specific expression is a simplified and intuitive version often used in educational contexts to explain the model, and it does not appear in the scholarly literature in this exact form.

In the academic literature, the Gravity Model is presented in a more complex and theoretically grounded form. According to the paper to Head and Mayer, the general formulation of the Gravity Model is expressed as in the Equation 1.2.

$$X_{ni} = G \times S_i \times M_n \times \phi_{ni}$$

Equation 1.2.: General formulation of the Gravity Model by K. Head and T. Mayer

Where:

- X_{ni} represents the value of trade from exporter i to importer n.
- S_i captures the "capabilities" of exporter i as a supplier to all destinations.
- M_n represents the characteristics of the destination market n that promotes imports from all sources.
- ϕ_{ni} is a term that combines trade costs and their elasticity, capturing the bilateral accessibility of a country n to exporter i.

This general formulation, also referred to as Structural gravity, is derived from more rigorous economic foundations. It takes into account not just the size and distance between countries but also multilateral resistance terms that account for third-country effects. These terms reflect the broader context in which trade occurs, including the influence of other trading partners and global economic conditions.

The Structural gravity model is more comprehensive and accurate for empirical analysis and policy evaluation because it integrates these additional complexities. This formulation allows economists to assess the impacts of changes in trade costs, trade policies, and other factors on bilateral trade flows more precisely.

K. Head and T. Mayer, in their work, present also an alternative formula that can be considered as a middle ground between the simplified intuitive formula and the more complex general structural gravity model. The naive gravity equation includes some of the basic elements of the gravity model while omitting the full complexity of multilateral resistance terms and other detailed economic factors. This model simplifies bilateral trade relations to the product of country sizes.

$$X_{ni} = G \times Y_i^a \times Y_n^b \times \phi_{ni}$$

Equation 1.3.: Formula of Naive Gravity Model by K. Head and T. Mayer.

Where:

- X_{ni} represents the bilateral trade flow between country i (the exporter) and country n (the importer).
- G is the constant that scales the equation.
- Y_i^a and Y_n^b represent the economic sizes of the exporter and importer, respectively. These are typically measured by variables such as gross domestic product (GDP) or gross national income (GNI).
- a and b are parameters that determine the elasticity of trade flows with respect to the economic sizes of the exporter and importer, respectively. They indicate how changes in the economic size of either country affect bilateral trade.
- ϕ_{ni} represents the bilateral accessibility between the exporter i and importer n. It ranges from 0 to 1 and combines factors such as trade costs and other barriers to trade. A higher value of ϕ_{ni} indicates greater accessibility or lower trade costs between the two countries.

This model is termed naive because it simplifies trade relationships to be solely determined by the economic sizes of the trading partners and their bilateral accessibility, without considering other potential determinants of trade.

The Naive gravity equation maintains the core idea of the gravity model—that trade between two countries is proportional to their economic sizes (GDPs) and affected by the costs of trading between them (captured by ϕ_{ni}). However, it simplifies the relationship by assuming that the influence of GDPs on trade is determined by elasticities a and b, which

may differ from 1, and by not explicitly incorporating multilateral resistance terms that account for the trade relationships with other countries.

The basic Gravity Model, which originally considers the GDPs of two countries and the distance between them, can be extended to include additional factors that can influence trade between countries. These extensions are crucial because they allow the model to more accurately reflect the real-world complexities of international trade.

To estimate the impact of these variables on trade, economists often transform the Gravity Model into a log-linear form, which is easier to work with statistically.

The relationship expressed by the gravity model and its log-linearization will serve as the foundation for the work carried out in the subsequent paragraphs.

1.4. Differences within sectors.

This analysis is focused solely on the manufacturing sector, without delving into specific industries or including service providers. However, it's important to acknowledge that the sector a company operates in heavily influences its strategic decisions regarding offshoring and nearshoring. Historically, offshoring has been closely linked to manufacturing, particularly in labor-heavy fields like apparel production, where moving operations to low-wage regions offers obvious cost benefits. Nevertheless, offshoring isn't limited to manufacturing—its impact stretches into service-based and tech-driven sectors as well, which warrants a deeper look into how these strategies unfold in different areas.

One significant distinction arises between capital-intensive and labor-intensive industries, which strongly affects whether a company chooses to nearshore or offshore its operations. For example, apparel manufacturing, a labor-intensive industry, frequently turns to offshoring to capitalize on lower labor costs. Yet, there are exceptions—such as Italy's high-end fashion industry—where companies deliberately keep production local to maintain quality and uphold brand prestige, going against the grain of typical economic reasoning.

In the services sector, offshoring has been driven by advancements in information technology, making it easier to transmit data across borders and shift tasks like customer service and back-office operations to remote locations. Even with these advances, unique challenges remain. Regulations such as the European Union's Digital Operational Resilience Act (DORA) place strict limits on how and where sensitive services, particularly in finance, can be offshored, especially to areas considered high-risk. This underscores the growing importance of regulatory compliance, data security, and geopolitical stability in shaping offshoring decisions for service-based companies.

Moreover, the pattern of offshoring in services differs from that in manufacturing due to the relative difficulty in trading many services compared to goods. Although services offshoring is less common at present, it's expected to rise as technology continues to break down barriers. Smaller economies, like Luxembourg and Ireland, have embraced this trend more quickly due to their agility, whereas larger economies like the U.S. and China are only beginning to see noticeable growth in this area.

The expansion of offshoring into the service sector highlights a broader economic shift. Whereas offshoring used to revolve around the physical production of goods, today, it increasingly centers on intellectual labor and information-driven tasks. This shift is forcing companies to rethink how they allocate resources and manage their workforce. Although still developing, services offshoring holds the potential to significantly reshape global trade, particularly as more firms adopt digital tools to streamline their international operations.

Ultimately, the choice between nearshoring and offshoring is now influenced by far more than just cost savings. Factors such as proximity to key markets, compliance with regulations, and advancements in technology have become critical to strategic planning. As the lines between goods and services continue to blur and technology makes global operations more seamless, the future of these practices will depend on how companies navigate the complexities of international trade, regulation, and innovation.

While this study focuses on manufacturing, future research should explore the unique trends in nearshoring and offshoring within service and technology sectors. Investigating how shared factors—such as geographical proximity, market size, or language—shape these trends could lead to a more nuanced understanding of global trade patterns and the specific challenges faced by companies across diverse industries.

2. Study framework and model presentation

To delve deeper into the intricate interplay between distance and international trade within the Italian context, it's necessary to scrutinize real-world data and track their evolution over time.

The main source of data analysis for this work is the database of imports and exports of manufacturing firms in Italy from 2008 to 2019. It provides, for all Italian firms engaged in international trade, the value of all different imported and exported goods in a given year. The products are identified by the NC8 code, 8-digit Combined Nomenclature level, which can provide a detailed description of the product type.

For confidentiality reasons, firms are identified by a unique code not traceable to the company, to which the Ateco code related to the sector in which the company operates is associated. The code ateco2007impr can provide a high level of detail in the description up to 6 digits. As fewer digits are considered, the sector becomes more general, encompassing increasingly diverse types of manufacturing sectors.

As the aim of this thesis is to examine the influence of both geographical and psychological distance on international trade over the years, the following paragraphs present a gravity model and outline the steps for its construction. This model will serve to measure offshoring as a dependent variable, incorporating various independent variables integrating all distance factors into the multivariable regression equation. This approach will provide a foundation for observing the phenomenon of nearshoring across the study period.

A key resource for the present research for working with the Gravity model is the CEPII Gravity Database, which contains a wide array of bilateral and unilateral variables relevant to estimating trade flows. The database spans from 1948 to 2019, providing comprehensive data on factors like GDP, geographic distances, shared languages, legal origins, and trade agreements. This wealth of information is organized by origin-destination-year triplets, offering a detailed look at both historical and contemporary trade relationships. Each variable in the database plays a critical role in shaping trade flows, and the use of this resource allows for a deep analysis of global trade patterns over time.

2.1. Empirical definition of offshoring

In the previous chapter, offshoring has been generally defined. To proceed with the analysis of real data, it's essential to establish metrics that determine whether an international transaction qualifies as offshoring. Before proposing these metrics, it's important to emphasize another time that the present analysis exclusively concerns manufacturing products. Therefore, the definition will be tailored to exclude any implications related to services trade, reselling, and procurement of raw materials.

According to the severity level of selection, the possible definitions can vary depending on the type of code or the number of digits considered. For this analysis, it will be adopted the definition provided by "*A Portrait of Backshorers: Evidence from Italian Administrative Data"* (A. D'Ambrosio, K. Lavoratori, L. Benfratello, D. Castellani, A. Manello). Due to the absence of product-level data on domestic sales in the database under study, the Authors rely on the primary industry of activity for each firm. Therefore, they attribute to the 5-digit Ateco 2007 industry activity code the entire production of the firm.

The authors identify the presence of offshoring when:

$$ateco2007\_4d = ateco2007impr\_4d$$

Where $ateco2007\_4d$ represents the 4-digit ATECO code for the transaction, and $ateco2007impr\_4d$ represents the 4-digit ATECO code for the enterprise.

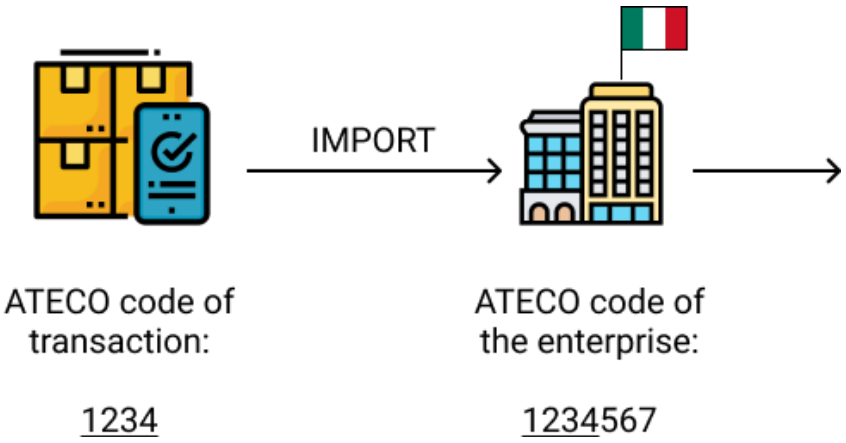

Image 2.1.: Graphical representation and example of offshoring definition according to ATECO codes

This definition considers only imports, this implies that every export entry in the database is not used in the study.

Based on the available data, another possible analysis approach, suggested by A. D'Ambrosio, can be considered. The researcher proposes considering export data and relating them to imports. The phenomenon of offshoring can also be observed when the same type of product is first imported and then exported by the same company. To identify this scenario, it's necessary to examine cases where the same CN8 code from the same firm appears both in export and import records.

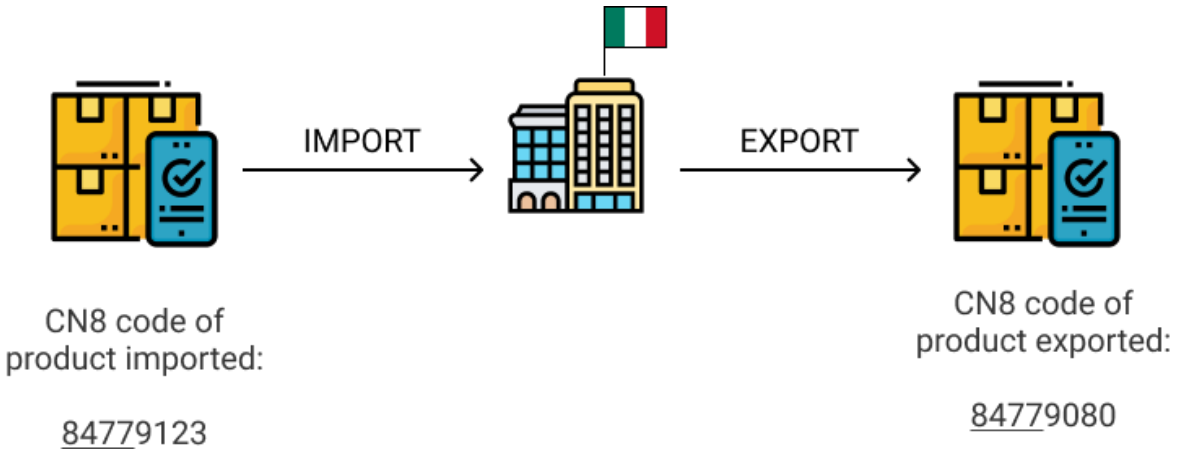

Image 2.2.: Graphical representation of offshoring definition according to CN8 codes of import and export

The forthcoming analysis will utilize the initial definition presented, which will be referred to as "offshoring by ATECO code" from now on.

2.1.1. Strengths and limitations of the empirical definitions

The definition of offshoring by the ATECO code (transactions $ateco2007impr\_4d$ and enterprises $ateco2007impr\_4d$), presents different methodological advantages and challenges. This approach and the composition of the database ensures that the analysis is narrowly tailored to observe offshoring within the manufacturing sector, effectively excluding confounding elements such as services trade, reselling, and raw material procurement. Additionally, utilizing 4-digit ATECO codes allows for a quite good level of detailed classification that can capture specific industry dynamics and variations in offshoring practices across different manufacturing sub-sectors.

However, the reliance on the primary industry activity code for each firm, derived from the 5-digit Ateco 2007 code, assumes that all production activities of a firm fall within a single industry classification. This assumption may not accurately reflect the diversity of activities

within multifaceted enterprises operating across different industry segments, potentially leading to an incomplete understanding of a firm's offshoring practices. Moreover, by excluding exports from the analysis, the approach ignores the broader context of a firm's global supply chain operations. Offshoring strategies may involve intricate combinations of imports and exports, and ignoring export data could result in a significant gap in understanding how firms integrate their international trade activities. This limitation is particularly relevant in industries where global value chains are critical, potentially missing key aspects of how firms optimize their production processes through both imports and exports. Additionally, the definition may not adequately capture indirect forms of offshoring, such as subcontracting or intermediary trade, where the connection between the firm and the imported goods is less direct.

The other analysis approach proposed involves considering export data and relating them to imports. Offshoring can also be observed when the same type of product is first imported and then exported by the same company. To identify this scenario, it is necessary to examine cases where the same CN8 code from the same firm appears in both export and import records. This approach addresses the limitation of excluding export data by integrating it into the analysis, offering a more comprehensive view of the firm's global trade activities. It captures the indirect forms of offshoring by tracking the movement of goods through different stages of the supply chain, offering a more complete picture of offshoring practices. However it completely omits any imported goods that are sold directly in Italy without being exported, which still respects the conceptual definition of offshoring. This limitation arises from the shortage of detailed data about national sales compared to international ones, and thus, cannot be resolved with the available data.

In comparison, while the initial definition based on 4-digit ATECO codes offers precision and detailed industry classification, it falls short in capturing the full scope of offshoring activities, especially concerning multi-industry operations and the exclusion of exports. The suggested approach, by incorporating export data and examining the relationship between imports and exports through CN8 codes, provides a more balanced and refined analysis. However, given the data limitations, the 4-digit ATECO-based definition remains preferable for its clarity and focus within the constraints of the available information.

2.2. Independent variables of the model

A critical phase of the construction of the model implies the definition of the independent variables, this choice will significantly influence the results of the analysis.

The dependent variable, $Offshoring_{ijt}$, measures the total value of imports of manufacturing products from the partener nation i to italian province j in the year t, which respect the definition of offshoring by ATECO code.

The identification of relevant independent variables begins with the components that influence psychic distance, as discussed in Paragraph 1.2.1. These components are essential for understanding the non-economic barriers that impact trade flows, such as cultural and administrative differences. Additionally, it is necessary to incorporate two variables representing the attributes of the Italian provinces and the foreign countries considered. These variables correspond to the monadic effects, or the intrinsic characteristics of each region, which are crucial in the context of the Gravity Model. By including these monadic effects, the model ensures a more comprehensive analysis, accounting for both bilateral and unilateral factors influencing offshoring and nearshoring decisions.

The Tab 2.1 presents the list of variables and the considerations relative to their contribution in the model to be achieved.

Name	Variable	Time dependent	Type
D_{ij}	Distance	No	Continuous
TZ_i	Time Zone	No	Categorical
C_{it}	Currency	Yes	Categorical
CR_i	Colonial Relationship	No	Categorical
EEA_i	European Economic Area	No	Categorical
RTA_i	Regional Trade Agreement	Yes	Categorical
CL_{ij}	Common Language	No	Categorical

Tab 2.1. List of independent variables and their characteristics

Additional factors such as political affinity, common cultural history or common standards in measurement could have been included. However, it is very challenging to find data that are both complete and reliable.

2.2.1. Monadic Effects

To accurately model trade dynamics, it's essential to incorporate variables that capture the unique characteristics of each trading partner independently, beyond just the bilateral relationship. These variables, known as monadic effects, represent factors intrinsic to the Italian provinces and the foreign countries involved in trade.

Monadic effects in the Gravity Model are often implemented through variables that reflect the economic size or "mass" of a region, such as GDP, population, or income levels. According to the theoretical foundations discussed in the literature, including the work by Anderson and van Wincoop, these variables are critical in determining the capacity and propensity of a region to engage in trade. The inclusion of monadic effects addresses unobserved heterogeneity across countries, which is crucial for generating accurate and reliable results. By accounting for country-specific attributes, such as infrastructure quality, political stability, and economic policies, the model controls for these factors' broad impact on trade, ensuring a more precise and robust analysis.

In the empirical application of this model to Italy's offshoring and nearshoring trends, monadic effects are represented by GDP per capita for both the exporter and importer. This choice is grounded in the understanding that GDP per capita serves as a reliable proxy for economic development and market potential, which are critical determinants of trade flows. By incorporating GDP per capita as a monadic variable, the model accounts for the economic strength and consumer purchasing power of each region, thereby influencing the likelihood and volume of trade.

2.2.2. Physical geographic distance: D_{ij}

It is possible to propose different approaches to calculating the geographical distance between one company and another. For a fully accurate picture, one would need to calculate the distance between the manufacturing company in the foreign country and the importing Italian company. However, while this method is highly accurate, it can be computationally intensive and not always feasible due to the format of the input data. In this case, the data of the exact locations of the enterprises involved in trading are omitted for confidentiality reasons.

The general solution is to compute the great circle distance in kilometers between the two countries rather than the relative positions of the two companies. To achieve this, one can

employ the methodology commonly used by mapping software: simplify the problem to a point-to-point distance. In the present analysis, the points considered will be the Italian province and the capital of the foreign country.

The present work uses as the main reference for these measures the CEPII database.

2.2.3. Time zone: TZ_i

Neighboring countries will likely share the same time zone, so this variable will present some correlation with geographical distance. This variable, as categorical, will be boolean.

$TZ_i = 1$, if the country i belongs to the same time zone area of Italy;

$TZ_i = 0$, otherwise.

TZ_i is the same for every Italian province j and is not time dependent.

Also for this variable the CEPII's database will be used. Due to the form of input data and for the sake of simplicity, certain assumptions are made:

- Standard time zones are utilized, without consideration for daylight saving time adjustments;

- In the case of countries having multiple time zones, if the country has in part common time zone is $TZ_i = 1$;

- Time zone is fixed during time and corresponds to the one provided by TimeZoneDB corresponding to 2020.

The significance of this variable lies in the fact that sharing the same time zone simplifies the coordination of international events, such as meetings and conferences, by reducing the need to calculate time differences and thus avoiding confusion. This facilitates communication and the planning of activities among individuals and businesses located in those states. Additionally, having the same time zone makes travel and transportation between states more convenient since there are no significant differences in arrival and departure times.

2.2.4. Currency: C_{it}

The use of the same currency is an extremely important factor for international trade and is indicative of the economic proximity between two countries. The Italian state belongs to the Eurozone, which currently includes 20 states. In the present analysis, C_{it} is a boolean variable, assuming the value 1 in case the foreign country uses the euro as a national currency.

The Eurozone experiences ongoing changes regarding its member countries. Therefore, the currency variable is time-dependent. During the period under consideration, four countries joined the Eurozone: Slovakia in 2009, Estonia in 2011, Latvia in 2014, and Lithuania in 2015.

2.2.5. Colonial relationship: CR_i

Colonial relationships have historically played a significant role in shaping preferential trade paths between colonizers and their former colonies. These relationships can reduce barriers to trade by creating long-standing economic, political, and administrative ties, often resulting in harmonized institutions, shared legal frameworks, and familiarity with business practices. For many nations, the remnants of colonial ties still influence modern trade flows, where former colonies continue to maintain stronger trade relationships with their former colonizers.

Countries such as France and Great Britain, with extensive colonial empires, continue to exhibit strong economic ties with many of their former colonies, particularly in Africa and Asia. These relationships are reflected in preferential trade agreements, shared language, and even common legal systems that persist long after political independence. In many cases, these colonial legacies reduce transaction costs and facilitate smoother trade relations.

In the case of Italy, however, the colonial relationships are far more limited both in scope and duration. Italy's colonial history is primarily concentrated in Africa, specifically with Eritrea (1882-1947), Italian Somalia (1890-1960), Libya (1911-1943), and Ethiopia (1936-1941). Although these regions were once under Italian control, the relatively short

period of colonization and the geopolitical shifts after World War II significantly reduced Italy's influence over its former colonies.

Due to the relatively brief and historically distant nature of these colonial ties, it seems unlikely that these relationships continue to exert a significant influence on Italy's modern trade patterns, especially in the context of offshoring and nearshoring.

While colonial relationships are not particularly relevant to Italy's case, they remain an important factor in studies focusing on countries with more extensive colonial legacies, such as France and Great Britain. In these cases, colonial ties still influence trade flows, particularly through mechanisms like preferential tariffs, development aid, and ongoing political partnerships. Former colonies often have stronger administrative and legal connections with their former colonizers, making trade relationships more robust and smoother. For instance, the Commonwealth of Nations for Britain or La Francophonie for France facilitates easier trade and economic integration with former colonies, often resulting in stronger trade networks.

2.2.6. Trade Agreements: EEA_{it}

Trade agreements can differ significantly in terms of structure, scope, and the level of benefits they provide to participating countries. These agreements are essential in shaping global trade flows by reducing tariffs, simplifying regulations, and promoting cross-border investments. The focus in this analysis is on two key types of trade agreements: the European Economic Area (EEA) and Regional Trade Agreements (RTAs), both of which are central to understanding Italy's offshoring and nearshoring dynamics.

The European Economic Area (EEA) is a significant economic zone that extends the benefits of the European Union's single market to non-EU countries like Iceland, Liechtenstein, and Norway. During the studied period from 2008 to 2019, the only significant change in EEA membership was Croatia's accession in 2014, following its EU accession in 2013. Understanding the trade implications of this accession, particularly in the years leading up to and following Croatia's membership, provides valuable insights into the broader economic impacts of joining such a major economic area.

The process of integrating into the EEA involves extensive preparations, including aligning national regulations with EU standards and gradually opening markets to the competitive

pressures of the single market. For Croatia, this process began well before its official EEA accession in 2014. From as early as 2011, Croatia started implementing measures that would bring its economy closer to EEA norms. This early alignment can have measurable impacts on trade, as businesses and investors anticipate the benefits of future market access.

L. Benfratello e et al, in Inequality, Geography and Global Value Chains cap 6, indicate that these preparatory actions often lead to increased trade flows and investment even before formal membership is granted. This anticipatory effect is critical to consider in any analysis of EEA membership impacts during the period from 2008 to 2019. Croatia's gradual integration into the EU and EEA likely began to influence its trade relations and economic performance as early as 2011, a few years before its official EEA membership.

To accurately capture the dynamics of EEA membership during the 2008-2019 period, the database should incorporate dummy variables that reflect the status of EEA membership for each country. For countries like Norway, Iceland, and Liechtenstein, which were already part of the EEA before 2008, the dummy variable should consistently be set to 1 throughout the entire period from 2008 to 2019. This reflects their ongoing participation in the EEA's single market.

In the case of Croatia, which officially joined the EEA in 2014, the dummy variable should be set to 1 starting from that year to mark its formal accession. However, to account for the effects of Croatia’s integration process prior to official membership, it is beneficial to introduce an additional dummy variable. This pre-accession variable would be set to 1 during the years 2011 to 2013, allowing to assess whether the preparatory steps and gradual alignment with EEA standards had a measurable impact on trade flows even before full membership was achieved.

By including these dummy variables in your regression model, it’s possible to effectively differentiate between the effects of full EEA membership and the earlier integration phase. This approach is crucial for capturing the incremental nature of economic integration and recognizing how market participants often adjust their behavior in anticipation of upcoming changes.

	Condition
EEA_{it} = 1	This variable is set to 1 for all countries that are members of the EEA for all years t from 2008 to 2019, except for Croatia. For Croatia, the dummy variable is set to 1 starting from 2014 onward.
$PEEA_{it} = 1$	This variable is set to 1 if the country i is Croatia and the year t is between 2011 and 2013. For all other countries and for Croatia outside this period, the variable is set to 0.

Tab 2.2. Explanation on how values are assigned to the variables EEA_{it}

2.2.7. Trade Agreements: RTA_{it}

Regional Trade Agreements (RTAs) are treaties between two or more countries that define the rules of trade among the signatories. These agreements typically provide preferential market access, such as reduced or eliminated tariffs on goods traded between the member countries. In economic models, RTAs are crucial variables because they significantly shape trade flows and economic integration among member countries. Their inclusion allows for a more precise analysis of how these agreements influence trade dynamics, investment patterns, and overall economic outcomes within the participating regions.

To conduct this analysis on RTAs, the valuable database of the World Trade Organization (WTO) has been utilized. This database includes all Regional Trade Agreements between the European Union and other countries. Specifically, agreements covering goods have been selected for analysis.

	Condition	Example
$RTA_{it} = 1$	If the foreign country i has a regional trade agreement with the EU in the year t, 0 otherwise.	Venezia (VE) - Montenegro (MNE)

Tab 2.3. Explanation and examples on how values are assigned to the variables RTA_{it}

2.2.8. Common language: CL_{ij}

Language plays a significant role in facilitating international trade, as shared linguistic ties can reduce communication barriers, enhance mutual understanding, and streamline business transactions. In this analysis, the influence of language is captured through the variable Common Official Language (CL_{ij}), which reflects whether two trading partners share a common official language. This factor can potentially foster closer economic relationships by making communication and business operations smoother between countries.

In the case of Italy, the only countries that share Italian as an official language are San Marino and Switzerland. However, several regions within Italy exhibit bilingualism or the use of multiple official languages, extending the relevance of the common language variable to other countries. Specifically:

- In Valle d'Aosta, both Italian and French are recognized as official languages.
- In Alto Adige (South Tyrol), Italian and German are official languages.
- In Friuli Venezia Giulia, Italian and Slovenian are official in certain municipalities.

For instance, provinces in Valle d'Aosta have linguistic ties with Francophone countries such as France, while those in Alto Adige share a common language with

German-speaking countries such as Germany, Austria, and Switzerland. Similarly, in Friuli Venezia Giulia, the presence of Slovenian establishes linguistic connections with Slovenia.

It is particularly important in this analysis to consider individual provinces, rather than treating Italy as a single linguistic unit. By focusing on the specific linguistic characteristics of provinces, the analysis more accurately captures the impact of common official language on trade. This provincial-level approach allows for the inclusion of regional bilingualism, which is not reflected when considering Italy as a whole.

Table 2.4 provides an example of how CL_{ij} values are assigned according to the regions and the corresponding countries with which they share a common official language.

	Condition	Example
CL_{ij} = 1	If the italian region of the province j has the same official language as th foreign country i	Aosta (AO) - France (FRA) Bolzano (BZ) - Germania (DEU) Catania (CA) - San Marino (SMR)
CL_{ij} = 0	Otherwise	Ancona (AN) - Spain (SPA)

Tab 2.4.: Explanation on how values are assigned to the variables CL_{ij} through the use of CEPII's database

2.3. Italian provinces

Currently, the territory of the Italian Republic is administratively divided into 107 provinces. Since the following study focuses on Italian provinces from 2008 to 2019, it is important to assess every internal change during these years.

In 2005, regional legislation in Sardinia resulted in the creation of four new provinces: Carbonia-Iglesias, Ogliastra, Medio Campidano, and Olbia-Tempio. This expansion aimed to improve local governance and brought the total number of provinces to 107. Later, in 2009, the number increased to 110 with the establishment of three additional provinces: Monza e Brianza in Lombardy, Fermo in Marche, and Barletta-Andria-Trani in Puglia. These new provinces were formed to enhance administrative efficiency and address regional needs.

Significant changes occurred again in 2017 when Sardinia underwent an administrative reorganization. The number of provinces on the island was reduced from eight to five. The provinces of Carbonia-Iglesias, Ogliastra, Medio Campidano, and Olbia-Tempio were dissolved, and the new Province of South Sardinia was established. This reorganization aimed to streamline administrative processes and returned the total number of Italian provinces to 107.

These discontinuities are managed in the database, and Table 2.5 provides a detailed overview of these changes and the corresponding years during which each province was present or restructured.

Province	**Code**	**Years Present in the Database**
Ogliastra	OG	2008–2017
Olbia-Tempio	OT	2008–2017
Carbonia-Iglesias	CI	2008–2017
Monza e Brianza	MB	2010–2019
Fermo	FM	2010–2019
Barletta-Andria-Trani	BT	2010–2019
Province of South Sardinia	SU	2018–2019

Tab 2.5: Discontinuities of Italian provinces during the year 2008-2019 in the database due to administrative changes

Additionally, the database includes four entries listed under the province code “99,” which represent unknown provinces. These entries will be excluded from the analysis, as they lack sufficient geographical data for proper assessment.

Taking into account these discontinuities, the total number of distinct provinces considered in the database is 112, reflecting the administrative shifts and ensuring that the analysis is comprehensive and accurate in addressing the dynamics of Italian provincial data from 2008 to 2019.

2.4. Gravity model

As an econometric problem, the objective is to obtain estimates of the unknown β parameters.

$$ln\ Offshoring_{ijt} = \beta_0 + \beta_1 GDPPC_{it} + \beta_2 GDPPC_{jt} + \beta_3 ln\ D_{ij} +$$

$$\beta_4 TZ_i + \beta_4 C_{it} + \beta_5 EEA_{it} + \beta_6 PEEA_{it} + \beta_7 RTA_{it} + \beta_8 CL_{ij} + \varepsilon$$

Equation 2.1.: Gravity model

where has been added ε as a stochastic disturbance term (error), representing unobserved factors or random shocks affecting offshoring activity that are not captured by the included variables.

β_i are the coefficients to be estimated through econometric methods. Each coefficient represents the effect of the corresponding variable on offshoring activity.

$ln\ Offshoring_{ijt}$ is the natural logarithm of the total value of import that respects the definition of offshoring by ATECO code between countries i and province j at time t.

2.5. Database construction

The construction of the database for the present analysis relies on Stata as the primary software tool. The final database, used to perform the linear regression analysis, is the result of merging multiple datasets while excluding any unused or irrelevant data to ensure accuracy and efficiency.

To standardize the classification of countries, the analysis employs the ISTAT ISO3 coding system, which is a widely recognized international standard for identifying countries with three-letter codes.

The process of database construction involves several stages, including data cleaning, merging, and the exclusion of variables not relevant to the research objectives. Each dataset input has been carefully selected to ensure that the variables align with the model's requirements. For further details on the specific steps and operations performed during the construction of the database, please refer to the Appendix, where the actions are outlined in greater detail.

2.5.1. Offshoring values: Trade_Sample_IT_Import_Export

The starting point of this analysis is the database containing import and export transactions of manufacturing firms in Italy from 2008 to 2019. For each Italian firm engaged in international trade, the dataset records the values of different imported and exported goods in each given year. The products are categorized using the NC8 code, an 8-digit Combined Nomenclature classification, which allows for precise identification of the product types involved in the transactions.

To specifically analyze offshoring activities, new variables were introduced into the dataset. One such variable is same_sector, a binary indicator showing whether a particular transaction occurs within the same sector as the enterprise. Another key variable is offshoring, which captures the total offshoring amount related to each transaction. Additionally, a binary variable named offshorer was created to indicate whether a given transaction qualifies as an offshoring operation.

Following these updates, the dataset contained 113,085 entries that were identified as offshoring transactions within the manufacturing sector. This volume represents the offshoring operations, allowing for a detailed analysis of Italy's offshoring trends in the manufacturing domain.

To simplify the structure for analysis, the dataset was further aggregated by country, year, and province. The offshoring transactions were summed across these combinations, providing a broader view of offshoring activities on both a regional and national scale. This aggregation helps capture significant trends in offshoring at multiple levels, ensuring the study focuses on meaningful patterns across geographic areas and timeframes (see Appendix 1 for detailed Stata commands used in this process).

In this dataset, countries are classified according to a system defined by ISTAT (Italy's National Institute of Statistics), which updates the classification of countries and territories annually to reflect geopolitical changes. The countries are grouped by continent and geopolitical area, with further distinctions such as membership in the European Union. Each country is assigned both a statistical code and the ISO3 code, used in standard international nomenclature systems. To ensure consistency across different datasets and for the purpose of subsequent analyses, it became necessary to standardize the classification using the ISO3 alphabetic code. For this reason, a new database was constructed to facilitate the conversion from the ISTAT country codes to ISO3 codes, ensuring uniformity and compatibility with international standards in trade data analysis. This homogenous classification is essential for conducting accurate international comparisons and for integrating the dataset with external sources.

2.5.2. GDP per capita foreign countries: World Bank’s data

The dataset was obtained from the World Bank's GDP per capita data repository. The database was first imported into Stata, where unnecessary columns and rows were removed, retaining only the relevant columns such as country name, country code, and GDP per capita for the years 2008-2019. The year columns were appropriately renamed, and the cleaned dataset was saved for subsequent analysis.

As the GDP per capita values in the dataset were expressed in US dollars, it became necessary to convert these values into euros to align with the context of the analysis. To perform this conversion, an additional step was required. A new dataset containing

historical Euro-Dollar conversion rates was obtained. The conversion rates for each year were calculated as the average of the daily exchange rates for that specific year.

Following this, the cleaned GDP per capita dataset was merged with the conversion rates dataset. This merger enabled the calculation of GDP per capita values in euros by dividing the US dollar values by the corresponding yearly average Euro-Dollar conversion rates. The resulting dataset, which now includes GDP per capita values converted to euros, was saved and prepared for the final analysis (see Appendix 2 for details on the merging process and calculations).

2.5.3. Timezone: CEPII's database

To examine the influence of time zone alignment on trade, the CEPII database was utilized to determine whether a country shares the same time zone as Italy. A binary variable named *same_tz* was created for this purpose. This variable is assigned a value of 1 if the country in question is in the same time zone as Italy, and 0 if it is not (for detailed methodology, refer to Appendix 3).

2.5.4. Currency: self made database from EU institution information

A self-made database was constructed based on information from EU institutions to analyze the impact of currency usage on trade. This database includes countries that use the Euro as their official currency. For each year and for each country in the Eurozone, the variable currency was assigned a value of 1 if the country was part of the Eurozone during that specific year, and 0 otherwise. This variable allows for the examination of the influence of shared currency on trade and offshoring decisions (see Appendix 4 for further details).

2.5.5. GDP per Capita by region

The dataset containing regional GDP per capita was imported as a CSV file, where the date information was stored in a YYYY-MM-DD format. The year component was extracted by isolating the first four characters of the date string, which was subsequently converted from a string format to a numeric format for analytical purposes. Following this,

the dataset was refined to include only the relevant columns: year, province, and value added per inhabitant. The value added per inhabitant column was then renamed to GDP per capita (`gdppc`) to ensure consistency in terminology throughout the analysis (refer to Appendix 7 for detailed code and procedures).

To translate the NUTS3 codes into the corresponding Italian provincial abbreviations, a mapping process was used based on the official correspondences between the NUTS coding system (Nomenclature of Territorial Units for Statistics) and the Italian provincial administrative abbreviations.

2.5.6. European Economic Area: self made database

The creation of the EEA database was carried out following the methodology outlined in section 2.2.5. The process was relatively straightforward, as all EEA member countries, with the exception of Croatia, were assigned a value of 1 for every year under consideration. Croatia's status was adjusted to reflect its accession to the EEA in 2014, ensuring accurate representation in the database.

The data was first imported into Stata from the CSV file created. Upon importing, the variables representing countries and years were appropriately renamed to align with the specific years from 2008 to 2019. A new variable was generated to represent the ISO country codes, which mirrored the country names initially provided. The dataset was then reshaped into a long format, facilitating easier merging in subsequent analytical phases (refer to Appendix 6 for the complete code).

A similar procedure was followed in creating a separate database that tracked the effects of Croatia’s pre-accession period, represented by the variable *peea*, ensuring that the gradual alignment of Croatia with the EEA prior to its official accession was also accounted for.

2.5.7. Regional Trade Agreement: CEPII’s Database

To create a focused dataset for analyzing Italy's trade agreements related to the import of goods, a specific data processing workflow was implemented. The dataset was first

filtered to include only the years 2008 to 2019 and restricted to entries where Italy was the origin country. The variable *rta_coverage* was used to identify the scope of the trade agreements, where 0 indicates no agreement, 1 indicates agreements covering goods only, 2 indicates services only, and 3 indicates goods and services.

Given the focus on goods, the dataset was refined to retain only those RTAs that involve goods, specifically those with *rta_coverage* values of 1 (goods only) or 3 (goods and services). A binary variable rta was then generated, where entries meeting these criteria were assigned a value of 1, and others were set to 0. Finally, unnecessary variables were dropped, resulting in a streamlined dataset that is well-suited for analyzing Italy's RTAs related to goods (see Appendix 9 for detailed steps and code).

2.5.8. Common Language: CEPII's Database

The creation of the CLij database required a more complex approach due to the linguistic diversity in certain Italian provinces, as detailed in section 2.2.7. Valle d'Aosta and Alto Adige, where bilingualism in French and German is significant, required a manipulation of the CEPII's database. Additionally, the region of Friuli-Venezia Giulia, which has a significant Slovenian-speaking minority, was also taken into account.

The process began with filtering the dataset to include only Italy (ITA), France (FRA), Germany (DEU.1, DEU.2), and Slovenia (SVN) as origin countries, aiming to identify Italian, French, German, and Slovenian-speaking regions that correspond to the multilingual Italian provinces. The initial filtering step retained only records with a common official language (comlang_off) shared between these countries and the relevant Italian regions.

A new variable, *lang*, was introduced to codify these linguistic relationships, assigning "ITA" for Italian, "FRA" for French, "TED" for German, and "SLO" for Slovenian-speaking areas, based on the origin country. This ensured that provinces such as Friuli-Venezia Giulia, which shares linguistic ties with Slovenia, were appropriately represented.

As the dataset needed to represent France, Germany, and Slovenia not only as origin but also as destination countries, it was further expanded by generating new entries with these countries as destinations (for the complete code, see Appendix 8).

2.5.9. Final merge

The final merge process integrates all the datasets built from different key variables, ensuring the creation of a unified database on which to perform the regression analysis. Several types of merges were performed, each corresponding to specific combinations of variables:

Country level Merges
Datasets involving national-level information are merged based on country codes. In this analysis, the countries involved are those relevant to variables like common language, national indicator and time zone.

Province -Year Merges
For data that is both time-specific and region-specific, merges are performed using province and year as key variables. This is particularly relevant for datasets capturing regional GDP per capita over time. One issue was the presence of observations where no offshoring transactions occurred for specific provinces in certain years. In these cases, despite the availability of GDPpc data, the lack of corresponding offshoring values rendered the GDPpc data unusable for the analysis. Additionally, some mismatches occurred due to discrepancies in the coding of provinces, where the NUTS codes provided did not correspond to recognized provincial identifiers. As a result, the merge process flagged these provinces as unmatched entries, leading to the exclusion of potentially relevant GDPpc information from certain years. These limitations highlight the importance of ensuring consistency between datasets, particularly when integrating information from diverse sources.

Country - Year Merges
When datasets include both country and year variables, such as national economic data over multiple years, merging by country and year ensures proper alignment. This is the approach used for variables like GDP per capita, currency, EEA membership, and PEEA. (See Appendix 11)

Province - Country Merges
In certain cases, datasets require both province and country as key variables. An example is the distance variable, which depends on both the province and the country.

Province - Country - Year Merges
The most complex merges involve province, country, and year. This final merge is

performed once all the subsets have been prepared, representing the last step in the integration process.

While integrating data from various sources, such as the CEPII Gravity Database, presents significant opportunities for enriching trade analysis, it also introduces notable challenges. One of the primary issues in merging such datasets is the alignment of country identifiers. The Gravity Database utilizes unique country identifiers that combine ISO3 codes with numeric identifiers, accounting for changes in territorial boundaries or political entities over time. For instance, countries that undergo territorial changes without altering their ISO codes, like the reunification of Germany or the division of Sudan, may create discrepancies when aligning with datasets that do not capture such nuances.

Moreover, inconsistencies in how country data is recorded across different time periods or across datasets can complicate the merging process. Discrepancies often arise in cases where regional divisions (such as provinces within Italy or Germany) must be reconciled with national-level data. These alignment challenges are critical to address, as inconsistencies in merging datasets by key variables like ISO3 or year can lead to incorrect conclusions or data loss. This requires careful attention to both the structure and content of the datasets being integrated.

A further complication arises from missing data, especially when merging variables that are not available across all observations. For example, the *comlang_off* variable in the Gravity Database, which tracks whether two countries share an official language, is not consistently available for every country pair. This creates gaps in the dataset, which, if not handled appropriately, could introduce bias into the analysis. The presence of such gaps necessitates robust strategies to ensure that missing data does not distort the results of the trade flow analysis.

2.6. Expected results

Before performing the regression it's important to outline the anticipated outcomes in order to compare it with the result obtained after.

Firstly, a positive association is expected between the GDP per capita of both the exporting country and importing province ($GDPPC_{it}$ and $GDPPC_{jt}$) and offshoring activity. This positive relationship reflects the notion that higher economic prosperity in both nations stimulates offshoring, as stated by the Gravity theory. Conversely, a negative

coefficient is anticipated for the natural logarithm of distance $ln\,D_{ij}$, as greater geographical distance typically poses obstacles to offshoring. Additionally, the presence of trade zones or agreements TZ_{ij} is expected to exert a positive impact on offshoring activity by facilitating trade between member countries. Similarly, currency C_{ijt} and trade agreements TA_{ijt} are predicted to positively influence offshoring, indicating the significance of shared currency and trade relationships in fostering offshoring relationships. Lastly, language proximity L_{ij} is expected to yield a positive coefficient, suggesting that linguistic similarities enhance offshoring activities.

These projected outcomes are consistent with the literature reviewed in the previous paragraph and empirical findings in the field of international economics.

If nearshoring is indeed occurring, the results should reveal some additional trends. Specifically, the coefficient for distance $ln\,D_{ij}$ might show a stronger negative impact, indicating that companies are increasingly favoring geographically closer locations for their offshoring activities. This would reflect a shift from traditional offshoring to regions further distant, such as East Asia, toward nearer regions like Eastern Europe or the Mediterranean, where geographical proximity reduces logistical complexities and costs. Additionally, if nearshoring is present, the effects of trade zones and agreements might be more pronounced in nearby regions, further promoting intra-regional trade and collaboration. The presence of cultural and linguistic ties, as captured by language proximity, could also play a more significant role in nearshoring decisions, as businesses seek to mitigate the risks associated with cultural differences.

Observing these patterns would suggest that nearshoring is not just a theoretical trend but a practical shift in global trade dynamics, with significant implications for the future of international production and supply chain strategies.

2.6.1. Implication of divergent results

In case the results deviate from the expected outcomes, it would open up several alternative interpretations, each clarifying on evolving dynamics within global trade and offshoring practices.

1. Weak Influence of Distance

If the distance variable does not show a significantly negative impact on offshoring, this could signal a shift in how companies view geographical proximity. Traditionally, distance has been seen as a major barrier to trade due to increased transportation costs, longer delivery times, and greater logistical challenges. However, the advent of advanced transportation and communication technologies may have diminished these barriers, making it easier for companies to maintain efficient supply chains even over long distances.

This potential finding could imply that other factors, such as the availability of specialized resources, lower labor costs, or more favorable regulatory environments, are increasingly taking precedence over geographical proximity. Companies might be willing to offshore production to distant regions if these areas offer significant advantages in terms of cost savings, access to raw materials, or favorable business conditions. The influence of global logistics companies, improvements in supply chain management, and digital platforms facilitating remote operations could further explain this trend.

2. Limited Impact of Trade Agreements

If the presence of trade zones and agreements does not significantly influence offshoring, it could suggest a broader evolution in the structure of international trade. Traditionally, trade agreements have been crucial in reducing tariffs, streamlining customs procedures, and facilitating cross-border trade. However, the limited impact of these factors in your model might indicate that companies are increasingly relying on other forms of international cooperation.

This shift could reflect the growing importance of digital trade agreements, which govern online transactions and digital services rather than physical goods. Moreover, informal trade relationships, such as long-standing business partnerships or networks of trusted suppliers, might be playing a more significant role in shaping trade flows than formal agreements. The rise of global value chains, where production is spread across multiple countries, could also be reducing the relevance of bilateral or regional trade agreements in favor of more flexible and dynamic forms of collaboration.

3. Reduced Importance of Language Proximity

If language proximity shows little to no impact, it may suggest that global companies have developed advanced mechanisms to overcome language barriers. In the past, linguistic

similarities were crucial for facilitating communication, building trust, and ensuring smooth business operations. However, in today's globalized world, companies might be less constrained by language differences due to the availability of translation services, multilingual staff, and digital communication tools that can bridge linguistic gaps.

This finding could also indicate that companies are prioritizing other cultural or economic factors over language when deciding where to offshore production. For example, similarities in business practices, legal frameworks, or market conditions might be more important considerations than language. Alternatively, the rise of English as the global business lingua franca might be reducing the importance of local languages in international trade relationships.

4. Indicators of New Global Trends

If the results reveal unexpected patterns, it might indicate emerging trends in the global trade of goods that differ from traditional models. One potential trend could be the increasing importance of new logistics hubs, particularly in regions that are rapidly developing their infrastructure. This shift could lead to a reorganization of global trade routes, with goods being sourced from or passing through these new hubs more frequently than established centers.

Furthermore, geopolitical developments, such as trade wars, tariffs, or protectionist measures, could be exerting a stronger influence on the trade of goods than previously anticipated. Companies might be responding to these factors by diversifying their supply chains, sourcing goods from a wider range of countries to mitigate risks associated with geopolitical instability. This could result in more fragmented and less predictable trade patterns, as businesses adapt to the new realities of global commerce.

The implications of these unexpected results go beyond the immediate findings of this study and open up new paths for future research. One interesting direction could be exploring how digital trade agreements and informal trade networks are shaping global supply chains. As companies increasingly rely on digital platforms and less formal partnerships to conduct international trade, understanding these aspects could offer valuable insights into how offshoring and nearshoring strategies are evolving.

Moreover, looking into how businesses manage language barriers in a world that's becoming more multilingual could provide a fresh perspective on international operations. Language is crucial for communication and building relationships in global markets, so how companies adapt to linguistic diversity could play a significant role in their success.

Beyond the research, these findings could have practical implications for policymakers and business leaders. By understanding what drives or hinders offshoring and nearshoring, they can develop more effective trade policies, investment strategies, and business models. For example, if the study reveals that geographical distance is less of an obstacle than previously thought, policymakers may choose to prioritize investments in infrastructure and logistics capabilities, thereby enhancing the efficiency and competitiveness of trade networks.

In the end, while the expected results provide a solid foundation for this analysis, being open to different outcomes allows for a richer and deeper understanding of the complex dynamics in global trade. This not only strengthens the study but also adds to the broader conversation about the future of international business and economic growth.

3. Elaboration of results

3.1 Overview of Results

This section provides an overview of the results obtained from the empirical analysis. The primary objective is to examine the determinants of offshoring and nearshoring for Italian firms between 2008 and 2019, using the gravity model of trade discussed in previous chapters. The regression estimates will be presented in tabular form, displaying the estimated coefficients for each variable, standard errors, t-values, and significance levels.

These results will be interpreted in light of the theoretical expectations outlined in Section 2.6, in order to verify whether the data supports the hypotheses and to explore any divergences from expectations. Special attention will be given to variables such as geographic distance, time zone, common currency, trade agreements, and linguistic proximity to assess their impact on offshoring and nearshoring decisions for Italian companies.

` 3.1.1 Summary Table of Key Statistics

The Tab 3.1. presents the key descriptive statistics for the main variables used in the analysis. These statistics provide an initial overview of the distribution and characteristics of both the dependent and independent variables in the model.:

Variable	Mean	Standard Deviation	Minimum	Maximum
Offshoring (total value)				
Distance (km)				
Time zone				
Common currency				
EEA				
PEEA				
RTA				
Common language				

Tab 3.1.: Summary Statistics of Key Variables Used in the Offshoring and Nearshoring Analysis (2008-2019)

3.1.2 Presentation of Regression Results

The results of the regression analysis are presented in the following table, where each row corresponds to one of the independent variables included in the model. The table provides critical information about the estimated relationship between each independent variable and the dependent variable (offshoring). Specifically, the table contains the following columns:

- **Estimated Coefficient**: This value represents the strength and direction of the relationship between each independent variable and offshoring. A positive coefficient suggests that an increase in the independent variable leads to an increase in offshoring, while a negative coefficient suggests the opposite.

- **Standard Error**: This measures the accuracy of the coefficient estimate. Smaller standard errors indicate more precise estimates.

- **T-statistic**: The t-statistic is used to determine whether the estimated coefficient is statistically significant. It is the ratio of the estimated coefficient to its standard error.

- **P-value**: The p-value indicates the probability that the observed results could have occurred by chance under the null hypothesis (i.e., no effect). Lower p-values (typically below 0.05) suggest that the coefficient is statistically significant.

Variable	Coefficient	Standard Error	T-statistic	P-value
Geographic distance (ln)				
Time zone (same zone)				
Common currency (Euro)				
Trade agreements (RTA)				

Linguistic proximity				

Table 3.2: Regression Results for the Determinants of Offshoring and Nearshoring (2008-2019)

3.1.1 Positive Drivers of Offshoring

This section should focus on the variables that have shown a significant positive impact on offshoring activities. For example:

- **GDP per capita (foreign country and Italian province)**:

3.1.2 Negative Drivers of Offshoring

This section would focus on the factors that are negatively associated with offshoring:

- **Geographic Distance**:

3.2 Detailed Interpretation of Key Variables

For each of the core variables in the gravity model, provide a deeper interpretation of the findings:

3.2.1 Geographic Distance (D_ij)

3.2.2 Time Zone (TZ_i)

.

3.2.3 Currency (C_it)

3.2.4 Trade Agreements (EEA_it and RTA_it)

3.2.5 Common Language (CL_ij)

3.3 Comparison with Expected Results

In this section, compare the actual results with the expected results outlined in Section 2.6:

- **Geographic Distance**:
- **GDP per Capita**:
- **Currency and Trade Agreements**:

Conclusion

References

- Altman, Steven A., and Caroline R. Bastian. DHL Global Connectedness Report 2024: An In-depth Analysis of the State of Globalization.

- Bayar, Omer. “Treatment of Endogenous Monadic Variables in Gravity Equations.” International Review of Economics & Finance, vol. 52, 2017, pp. 21–28.

- Baier, Scott L., and Jeffrey H. Bergstrand. "Do Free Trade Agreements Actually Increase Members' International Trade?" Journal of International Economics, vol. 71, no. 1, 2007, pp. 72-95.

- Dow, Douglas, and Amal Karunaratna. "Developing a Multidimensional Instrument to Measure Psychic Distance Stimuli." Journal of International Business Studies, vol. 37, 2006, pp. 578-602. doi:10.1057/palgrave.jibs.8400221.

- Economist. "The Rise of Nearshoring." The Economist, 1 December 2005, pp. 61-63.

- Feenstra, Robert C., and Gordon H. Hanson. "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979-1990." The Quarterly Journal of Economics, vol. 114, no. 3, 1999, pp. 907-940.

- Ghemawat, Pankaj. "Distance Still Matters: The Hard Reality of Global Expansion." Harvard Business Review, 2001.

- Head, Keith, and Thierry Mayer. "Gravity, Market Potential and Economic Development." Journal of Economic Geography, vol. 11, no. 2, 2010, pp. 281-294. doi:10.1093/jeg/lbq037.

- Head, Keith, and Thierry Mayer. "Gravity Equations: Workhorse, Toolkit, and Cookbook." Journal of Economic Literature.

- Hutchinson, William K. "Does Ease of Communication Increase Trade? Commonality of Language and Bilateral Trade." International Trade, 2002.

- Jahns, Christopher, Evi Hartmann, and Lydia Bals. "Offshoring: Dimensions and Diffusion of a New Business Concept." Journal of Purchasing and Supply Management, vol. 12, no. 4, 2006, pp. 218-231. doi:10.1016/j.pursup.2006.10.001.

- Lee, J. M., Ibarra-Olivo, J. E., Lavoratori, K., & Li, L. Inequality, Geography and

Global Value Chains. The Academy of International Business, 2023. doi:10.1007/978-3-031-24090-4.

- "What is Friendshoring? Western Policymakers Want to Move Supply Chains to Friendly Countries." The Economist, 30 August 2023.

- Slepniov, D., Brazinskas, S., & Vejrum Wæhrens, B. "Nearshoring Practices." Baltic Journal of Management, vol. 8, no. 1, 2013, pp. 5-26. doi:10.1108/17465261311291632.

- "The European Union and the World Trade Organization." European Parliament Factsheets. www.europarl.europa.eu/factsheets/en/sheet/161/the-european-union-and-the-world-trade-organization.

- "2022 Strategic Foresight Report - Twinning the Green and Digital Transitions in the New Geopolitical Context." Communication from the Commission to the European Parliament and the Council.

- "Benetton Group Supply Chain Map." www.benettongroup.com/it/sostenibilita/catena-fornitura/mappa-e-lista.

- "Digital Operational Resilience Act (DORA)." European Central Bank. www.ecb.europa.eu/ecb/history-arts-culture/history/html/index.en.html.

- "Estimating the Gravity Model." UNESCAP. www.unescap.org/sites/default/d8files/5%20-%203.%20Estimating%20the%20Gravity%20Model_0.pdf.

- "Historical Data on Exchange Rates." European Central Bank. data.ecb.europa.eu/data/datasets/EXR/EXR.D.USD.EUR.SP00.A.

- "Istat Report 2020." www.istat.it/it/files//2020/12/C01.pdf.

- https://www.iban.com/country-codes

- https://www.iese.edu/media/research/pdfs/DI-0792-E.pdf

- https://www.istat.it/classificazione/classificazione-degli-stati-esteri/

- https://www.esf.de/portal/SharedDocs/PDFs/DE/FP-2007-2013/nuts_codes_eu.pdf?__blob=publicationFile&v=1

Appendix

Appendix 1 - Offshoring DB

```
* Drop rows where import is 0
drop if missing(import_val)


* Generate the same_sector variable
gen same_sector = (ateco2007_4d == ateco2007impr_4d)

* Sort the data by firmid, year, and country
sort paese anno firmid

* Generate the offshoring variable
gen offshoring = import_val * same_sector

* Generate the offshorer variable
gen offshorer = (offshoring > 0) if !missing(offshoring)

* Drop rows where offshorer is not 1
drop if offshorer != 1

* Drop the unused columns
drop ateco2007 ateco2007_str ateco2007impr ateco2007impr_str developed
ateco2007impr_5d_str ateco2007impr_5d ateco2007impr_4d_str ateco2007_4d_str
ateco2007_4d ateco2007impr_3d_str ateco2007impr_3d manufacturing export_val
ateco2007_3d ateco2007_3d_str same_sector NC8 firmid movim offshorer
ateco2007impr_4d import_val

rename paese iso3num

collapse (sum) offshoring, by(iso3num provincia anno)
```

Appendix 2 - GDP Countries

```
import delimited
"C:\Users\ilari\Downloads\Tesi\ORIG_GDPpc_World_Bank\API_NY.GDP.PCAP.CD_DS2_en_csv_v2_144080.csv",
 clear
* Remove the first row (which now contains the redundant column names)
drop in 1

* Drop unnecessary columns
drop if v1 == ""

* Keep only relevant columns: Country Name, Country Code, and years 2008-2019
keep v2 v53 v54 v55 v56 v57 v58 v59 v60 v61 v62 v63 v64

* Rename the year columns

rename v2 iso3
rename v53 GDP_pc2008
rename v54 GDP_pc2009
rename v55 GDP_pc2010
rename v56 GDP_pc2011
rename v57 GDP_pc2012
rename v58 GDP_pc2013
rename v59 GDP_pc2014
rename v60 GDP_pc2015
rename v61 GDP_pc2016
rename v62 GDP_pc2017
rename v63 GDP_pc2018
rename v64 GDP_pc2019

reshape long GDP_pc, i(iso3) j(year)

save "C:\Users\ilari\Downloads\Tesi\UPD_GDPpc_doll.dta", replace

clear
import delimited "C:\Users\ilari\Downloads\Tesi\ORIG_GDPpc_World_Bank\Conversion_eur_doll.csv"

* Rename the columns for ease of use
rename usdollareuroexrdusdeursp00a usd_eur

gen stata_date = date(date, "MDY")
format stata_date %td
gen year = year(stata_date)


* Calculate the average conversion rate for each year
collapse (mean) usd_eur, by(year)

* Keep only the years 2008 to 2019
keep if year >= 2008 & year <= 2019

save "C:\Users\ilari\Downloads\Tesi\ORIG_GDPpc_World_Bank\Conversion_dol_euro.dta", replace

use "C:\Users\ilari\Downloads\Tesi\UPD_GDPpc_doll.dta", clear
merge m:1 year using "C:\Users\ilari\Downloads\Tesi\ORIG_GDPpc_World_Bank\Conversion_dol_euro.dta"
gen GDP_pc_eur = GDP_pc / usd_eur

drop GDP_pc _merge usd_eur
save "C:\Users\ilari\Downloads\Tesi\UPD_GDPpc_eur.dta", replace
```

Appendix 3 - Countries

```
* Load the time zones and country files
import delimited
"C:\Users\ilari\Downloads\Tesi\ORIG_TimeZoneDB.csv\time_zone.csv",clear

drop v1 v4 v5 v6

rename (v2 v3) ( country_code timezone)

* Identify the relevant time zones for Italy
gen same_tz = ( timezone == "CET" | timezone == "CEST")

* Remove duplicates in the country_code
bysort country_code: keep if _n == 1

drop timezone

merge m:1 country_code using
"C:\Users\ilari\Downloads\Tesi\ORIG_TimeZoneDB.csv\Country_code.dta"

drop _merge

rename country_name country_txt
drop country_code

generate iso = " "

replace iso="ABW" if country_txt=="Aruba"
replace iso="AND" if country_txt=="Andorra"
replace iso="AIA" if country_txt=="Anguilla"
replace iso="AIA" if country_txt=="Anguila"
replace iso="ANT" if country_txt=="Netherlands Antilles (Former)"
replace iso="ASM" if country_txt=="American Samoa"
replace iso="BMU" if country_txt=="Bermuda"
replace iso="BOL" if country_txt=="Bolivia (Plurinational State of)"
```

Appendix 4 - Currency

```
import delimited "C:\Users\ilari\Downloads\Tesi\Database costruiti da
me\ZonaEuro.csv"


* Rename the columns for clarity (assuming the columns are named accordingly)
rename v1 country_code
rename v2 country_txt
rename v3 v2008
rename v4 v2009
rename v5 v2010
rename v6 v2011
rename v7 v2012
rename v8 v2013
rename v9 v2014
rename v10 v2015
rename v11 v2016
rename v12 v2017
rename v13 v2018
rename v14 v2019

* Reshape the dataset to long format to work with individual years
reshape long v, i(country_code) j(year)
rename v currency

* Generate the iso variable
gen iso = ""

* Replace the iso variable with specific country codes
replace iso="ABW" if country_txt=="Aruba"
replace iso="AND" if country_txt=="Andorra"
```

Appendix 5 - GDP Italian Province

```
import delimited "C:\Users\ilari\Downloads\Tesi\ORIG_Province_GDPpc.csv"

* Extract the year from the date string

gen year_only = substr(year, 1, 4)

* Convert the extracted year to a numeric variable

destring year_only, replace

 * Keep only the relevant columns: year, province, and GDP per capita (value added per capita)

keep year_only province value_added_per_habitant

* Rename the GDP per capita column for clarity

rename value_added_per_habitant gdppc

rename year_only year
```

Appendix 6 - EEA

```
import delimited "C:\Users\ilari\Downloads\Tesi\Database costruiti da me\EEA.csv"

*Rename the variable

rename v1 country_txt
rename v2 Y2008
rename v3 Y2009
rename v4 Y2010
rename v5 Y2011
rename v6 Y2012
rename v7 Y2013
rename v8 Y2014
rename v9 Y2015
rename v10 Y2016
rename v11 Y2017
rename v12 Y2018
rename v13 Y2019
generate iso=country_txt

drop in 1

do "C:\Users\ilari\Downloads\Tesi\country_codes_EN.do"

order iso
drop country_txt

* Reshape the dataset to long format to work with individual years
reshape long Y, i(iso) j(year)
rename Y eea
```

A***ppendix 7 - RTA***

```
use "C:\Users\ilari\Downloads\Tesi\ORIG_Gravity_dta_CEPII\Gravity_V202211.dta"

keep year country_id_d country_id_o rta_coverage
drop if year < 2008 | year > 2019
keep if country_id_o == "ITA"
* Create a new variable 'rta' and initialize it to 0.
generate rta = 0
* Set 'rta' to 1 for entries where the RTA covers goods or goods and services.
replace rta = 1 if rta_coverage == 1 | rta_coverage == 3

drop rta_coverage country_id_o
```

Appendix 8 - Common Language

```
use "C:\Users\ilari\Downloads\Tesi\ORIG_Gravity_dta_CEPII\Gravity_V202211.dta"

* Keep only the years between 2008 and 2019
drop if year < 2008 | year > 2019

* Keep only the specified origin countries
keep if (country_id_o == "FRA" | country_id_o == "DEU.1" | country_id_o ==
"DEU.2" | country_id_o == "ITA")

* Keep only the specified variables
keep year country_id_o country_id_d comlang_off

* Keep only observations with a common official language
keep if comlang_off == 1

* Create the variable 'lang' with empty values
generate lang = ""

* Assign the correct languages based on the origin country
replace lang = "FRA" if country_id_o == "FRA"
replace lang = "TED" if (country_id_o == "DEU.1" | country_id_o == "DEU.2")
replace lang = "ITA" if country_id_o == "ITA"

* Drop the comlang_off variable since it's no longer needed
drop comlang_off year country_id_o

duplicates drop country_id_d, force
rename country_id_d iso3

* Store the number of observations before expanding
local old_n = _N

* Add exactly 6 new rows
set obs `=`old_n' + 6'

* Manually input the new values into the new rows
replace iso3 = "FRA" in `=`old_n' + 1'
replace lang = "FRA" in `=`old_n' + 1'

replace iso3 = "DEU.1" in `=`old_n' + 2'
replace lang = "TED" in `=`old_n' + 2'

replace iso3 = "DEU.2" in `=`old_n' + 3'
replace lang = "TED" in `=`old_n' + 3'

replace iso3 = "DEU" in `=`old_n' + 4'
replace lang = "TED" in `=`old_n' + 4'

replace iso3 = "SVN" in `=`old_n' + 5'
replace lang = "SLO" in `=`old_n' + 5'

replace iso3 = "CHE" in `=`old_n' + 6'
replace lang = "ITA" in `=`old_n' + 6'
```

Appendix 9 - Merge Country

```
use "C:\Users\ilari\Downloads\Tesi\FINAL_MERGE\UPD_currency_iso3.dta", clear

rename iso iso3

* Perform the merge

merge 1:1 year iso3 using "C:\Users\ilari\Downloads\Tesi\UPD_GDPpc_doll.dta"

* Assign 0 to the 'currency' variable when observations are only found in the 'using' file

replace currency = 0 if _merge == 2

drop _merge country_txt

save "C:\Users\ilari\Downloads\Tesi\UPD_Merge_Countries.dta", replace

use "C:\Users\ilari\Downloads\Tesi\FINAL_MERGE\UPD_timezone_iso3.dta"

rename iso iso3

replace iso3 = trim(iso3)

drop if missing(iso3) | iso3 == ""

* Perform the merge

merge 1:m iso using "C:\Users\ilari\Downloads\Tesi\UPD_Merge_Countries.dta"

* Replace missing same_tz with 0 for unmatched records

replace same_tz = 0 if _merge == 2

drop if _merge == 1

* Optionally, drop the _merge variable if no longer needed

drop _merge country_txt
```

```
order year iso3

* Save the result

save "C:\Users\ilari\Downloads\Tesi\UPD_Final_Merged.dta", replace

rename iso3 iso

* Merge with the new dataset based on iso3 = iso and year

merge 1:1 iso year using "C:\Users\ilari\Downloads\Tesi\UPD_EEA_iso.dta"

* For the unmatched entries in the master dataset, assign eea = 0

replace eea = 0 if _merge == 1

drop _merge

save "C:\Users\ilari\Downloads\Tesi\UPD_Final_Merged.dta", replace

* Merge with the new dataset based on iso3 = iso and year

merge 1:1 iso year using "C:\Users\ilari\Downloads\Tesi\UPD_PEEA_iso.dta"

* For the unmatched entries in the master dataset, assign eea = 0

replace peea = 0 if _merge == 1

drop _merge country_txt

save "C:\Users\ilari\Downloads\Tesi\UPD_Final_Merged.dta", replace
```