REGIONAL TRANSPORT INFRASTRUCTURE AND TRADE FLOWS IN THE EU

A Master's Thesis

by

BERRİN ÖZCAN

Department of Economics

İhsan Doğramacı Bilkent University

Ankara

August 2021

BERRİN ÖZCAN

REGIONAL TRANSPORT INFRASTRUCTURE AND TRADE FLOWS IN THE EU

Bilkent University 2021

To my loving family

REGIONAL TRANSPORT INFRASTRUCTURE AND TRADE FLOWS IN THE EU

The Graduate School of Economics and Social Sciences

of

İhsan Doğramacı Bilkent University

by

BERRİN ÖZCAN

In Partial Fulfilment of the Requirements for the Degree of

MASTER OF ARTS IN ECONOMICS

THE DEPARTMENT OF

ECONOMICS

İHSAN DOĞRAMACI BİLKENT UNIVERSITY

ANKARA

August 2021

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Economics.

Assoc. Prof. Dr. Fatma Taşkın

Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Economics.

Asst. Prof. Banu Demir Pakel

Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Economics.

Prof. Dr. Serdar Sayan

Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

Prof. Dr. Refet Gürkaynak

Director

Y. Tarık Kara

ESBE Müdür V.

ABSTRACT

REGIONAL TRANSPORT INFRASTRUCTURE AND TRADE FLOWS IN THE EU

Özcan, Berrin

M.A., Department of Economics

Supervisor: Assoc. Prof. Dr. Fatma Taşkın

August 2021

How does regional transport infrastructure affect bilateral trade flows? An extensive literature on infrastructure and trade flows has attempted to answer this question by using country level or regional data. This current thesis focuses on the European Union (EU) and investigates the effect of transport infrastructure on international and intranational trade flows using NUTS 2 (Nomenclature of Territorial Units for Statistics) level data from 200 EU regions between the years 2000-2010. It is the first study to focus on flows and infrastructure at regional level in a multi-country setting. As in the previous studies in the infrastructure literature, the gravity equation is used to explain the relationship between the regional transport infrastructure and trade in the EU. Various alternative estimation methods such as Fixed Effects, PPML, lagged variables, instrumental variables (IV) and Hausman-Taylor IV method are used in order to overcome the issues related to heteroskedasticity, reverse causality and biased estimates that are frequently encountered with gravity equation. In the presence of bilateral and time fixed effects, the results suggest an increase of 0.05 to 0.13 per cent bilateral trade as infrastructure measures increase by 1 per cent. The robustness check follows that the estimates are not sensitive to the choice of unit of measure for the infrastructure variables.

Keywords: Bilateral Trade, Gravity Model, Infrastructure, Transport

ÖZET

AB'DE BÖLGESEL ULAŞIM ALTYAPISI VE TİCARET

Özcan, Berrin

Yüksek Lisans, İktisat Bölümü

Tez Danışmanı: Doç. Dr. Fatma Taşkın

Ağustos 2021

Bölgesel ulaşım altyapısı ticareti nasıl etkiler? Kapsamlı bir uluslararası ticaret literatürü bu soruyu ülke ve bölge bazındaki verilerle cevaplamaya çalışmıştır. Bu tez, Avrupa Birliği'ndeki (AB) 200 NUTS 2 (İstatistiki Bölge Birimleri Sınıflandırması, İBBS) seviyesindeki bölgenin 2000-2010 yılları arasındaki verilerini kullanarak ulaşım altyapısının uluslararası ve iç ticaret üzerindeki etkisini araştırmaktadır. Bu analiz, birden çok ülke verisini kullanarak ulaşım altyapısı ve ticaret konusunu bölgesel bazda inceleyen literatürdeki ilk çalışmadır. Altyapı ve uluslararası ticaret literatüründeki önceki çalışmalara benzer şekilde çekim modelini kullanarak AB'deki bölgesel ulaşım altyapısı ve ticaret arasındaki ilişki açıklanmıştır. Sabit zaman modelleri, PPML, gecikmeli değişkenler, araç değişkenler ve Hausman-Taylor IV metotları heteroskedastisite, ters nedensellik ve sapmalı tahminler gibi çekim modelinde sıkça rastlanan içsellik sorunlarının çözümü için kullanılmıştır. Zaman ve çift yönlü sabit zaman etkisi altında ulaşım altyapısında görülecek yüzde birlik artış sonucunda ikili ticarette yüzde 0,05 ila 0,13 oranında artış sağlanmaktadır. Yapılan dayanıklılık testi sonucunda çalışmanın sonuçlarının kullanılan ulaşım altyapısının ölçü birimine hassas olmadığı görülmüştür.

Anahtar Kelimeler: Altyapı, Çekim Modeli, İkili Ticaret, Ulaşım

ACKNOWLEDGEMENTS

I would like to thank my Thesis Supervisor, Assoc. Prof. Fatma Taşkın, for her constant support and encouragement. She was the best advisor that I could have wished for and is one of the best people I have known in my life. I must also acknowledge the valuable comments and suggestions that my thesis defense jury members, Asst. Prof. Banu Demir Pakel and Prof. Serdar Sayan has provided. I would like to thank my friends from my cohort at Bilkent University who made the stressful times during our master's degree more bearable. I am eternally grateful to have such a loving and supportive family. I could not have achieved most of the things in my life without my mother, father, and sister. I am especially indebted to the most powerful women I know, my mother and sister, who thought me to challenge myself and inspired me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

A vast literature reveals important evidence on the tradeoff between integration to intranational and international markets and protectionism that has faced countries for centuries. One of the most striking examples of this include how the Australian states' competitive railway freight rates that aimed to protect their own industries hampered interstate trade (Hytten, 1931). The standardization of the railway gauge across the US and the costs related to the intraregional adaptation to the standard railway width (Puffert, 2000) and the railways' contribution to the industrialization of the Midwest and thus increase in demand for Midwestern products in the US (Meyer, 1989) also tell the two different sides of the same story on trade costs and integration. So, to what degree transport costs affect a region's integration to the intranational or international markets?

The impact of trade costs on a country's exports has been a prominent aspect of international trade models and literature. As the survey paper of Anderson and van Wincoop (2004) shows, trade costs that are incurred until a product is delivered to the final consumer are very large. More importantly, these costs differ across countries: developing countries suffer from higher trade costs relative to developed countries. Thus, it is a challenging question to answer as to why trade costs are very high and why they change across countries. This line of empirical research focuses the trade costs generated by sources such as tariffs, geography, institutional factors, and transportation and their effect on international trade.

Studies on infrastructure and international trade try to answer the relationship between trade costs and international trade by focusing on the effects of different levels of physical and institutional infrastructure in countries, regions, and cities. Although conclusions of the previous research are similar, the analyses differ in terms of their scope and coverage. In the literature, some studies examine the impact of infrastructure on trade flows using country level data whereas other studies look at differences in infrastructure at regional level in a particular country. There is, however, no study that investigates the effects of sub-national differences in infrastructure on trade in a multi country setting.

The current thesis attempts to fill this gap in the literature by examining regional trade flows and transport infrastructure for 21 EU countries for the period 2000-2010. It is the first study to analyze international and intranational trade flows and regional inland transport infrastructure in 200 EU NUTS 2 level regions. The extent to which regional transport infrastructure affects international and intranational trade at region level is investigated by employing a gravity approach. The potential estimation issues such as simultaneity problem between trade and infrastructure, endogeneity bias, and heteroskedasticity in log-linear form are also addressed by proposing alternative estimation methods such as Fixed Effects (FE) estimation, lagged variables, instrumental variables (IV), and Hausman-Taylor IV approach.

The thesis is organized as follows: next section reviews the relevant literature. The third section presents an overview of the modes of transport that are used for intra-EU freight. Fourth section describes the data and explains the variables that will be used in the analysis. The fifth section presents the main gravity equation and modifications to the gravity model considering the endogeneity problem that might be caused by the potential reverse causality between infrastructure and trade. In the sixth section, the estimation results are presented, and the findings are discussed. The final section concludes.

2. LITERATURE REVIEW

How international trade is affected by transport costs and geography has been a highly discussed research question in the literature. Many important studies have put forward evidence on how distance, transportation costs and barriers to trade can create discrepancies in trade costs and how trade costs in turn affect trade flows (Mundell, 1957; Moneta, 1959; Hummels, 1999; Anderson & van Wincoop, 2004). Elaborating on this relationship, the literature on infrastructure focuses on how gravity equation incorporates the impact of infrastructure on international trade. In the literature, the impact of infrastructure on international trade flows has been explored by many studies in different settings. The question of how level of infrastructure affects the trade flows was answered by empirical and theoretic analysis using global or regional data. Here, a selection of the studies examining the role of infrastructure on trade flows is organized in terms of their level of analysis (global or regional) and their contributions to the international trade literature are presented.

In their global analysis, Limão and Venables (2001) estimate that a 10% increase in transport costs results in 20% lower trade flows, indicating the positive impact of high-quality hard infrastructure on trade flows using the transport costs of one exporting company in Baltimore, Maryland to its trade partners around the world. As some of the other most important contributions to the infrastructure literature, both Bougheas et al. (1999) and Nordas and Piermartini (2004) provide evidence to support the determinant role of the infrastructure in trade performance. In their analysis, Nordas and Piermartini use an aggregated measure for infrastructure indicators such as number of airports, number of telephone lines per people, port efficiency, etc. and find that port efficiency provides the greatest contribution to the total bilateral trade. In addition, the theoretical findings of Martin and Rogers (1995) follow that the firms locate their plants in regions with high-quality infrastructure which have important implications for multinational firms and countries that want to attract international investment.

As for trade flows in a particular region or area, the literature focuses mainly on Africa and Asia. For instance, in their study examining the effects of cross-border transport infrastructure in the Greater Mekong Subregion (GMS), Fujimura and Edmonds (2006) argue that the enhancement of the transport infrastructure domestically and at the borders contributes to regional trade, attract Foreign Direct Investment (FDI), and create “mutually reinforcing effects” which in turn have implications on growth and poverty reduction. In another analysis focusing on the GMS, Stone and Strutt (2010) investigate the impact of the Cross-Border Transport Agreement in the region. They show that substantial gains can be achieved in the region in terms of trade facilitation with quantitatively higher estimated impact of infrastructure in the GMS compared to earlier studies (Stone & Strutt, 2010). There are also studies that investigate how soft infrastructure, namely the socioeconomic factors that improve a population’s living standards such as access to health and education, affect trade flows along with physical or hard infrastructure. In their analysis of selected economies in Asia, for example, Ismail and Mahyideen (2015) find that both soft and physical infrastructure play an important role in trade performance and quality of infrastructure has more sustainable impacts on economic growth than the quantity of infrastructure.

For Africa, Buys, Deichmann, and Wheeler (2010) find that enhancement of the road infrastructure between African cities would increase the trade flows by $254 billion at a cost of $32 billion. Another study that points out the heterogeneity among African regions estimate the effect of improving road infrastructure in ECOWAS countries to the level of South Africa as an increase of $356.06 million in intraregional trade, which translates into nearly 5% (Akpan, 2014). In his study of network analysis containing 198 countries and Sub-Saharan economies, Shepherd (2016) finds that a country's trade performance is as important as its neighbors' to connect to the global value chains and emphasizes the pivotal role of infrastructure. Francois and Manchin (2013) also point out that low-income countries should improve their transport and communications infrastructure as well as their institutional quality to participate in global trade.

There are other important studies at regional scale which investigate the implications of regional infrastructure on regional trade performance and development. For example, Coşar and Demir (2016) report the impacts of a high-quality domestic road infrastructure on declining trade costs, high returns to road infrastructure investment and higher reductions in transport costs for industries that are more reliant on transportation by using data on Turkish provinces. In another study focusing on the road infrastructure, Duranton, Morrow, and Turner (2014) investigate the impact of highways on regional trade in US cities. They find that the highway distance between US cities decrease the bilateral trade significantly and emphasize the role of bilateral distance in trade flows and regional welfare. Similarly, using bilateral trade data between Spanish regions and 45 countries, Bensassi, Márquez-Ramos, Martínez-Zarzoso, and Suárez-Burguet (2015) underscore the importance of logistics and regional infrastructure as factors that influence competitiveness. In their analysis of regional growth and infrastructure in the European Union, Crescenzi and Rodriguez-Pose (2012) investigate the extent to which infrastructure in the EU contributed to the regional growth. Their results show that regional growth in the EU is not driven by infrastructure, instead, by region's innovation capacity, and attractiveness for migrants which may cast doubt on the degree of infrastructure expenditure in the EU (Crescenzi & Rodriguez-Pose, 2012). Utilizing data from Chinese provinces, the results of Démurger (2001) suggest that the disparities in infrastructure account for the heterogeneity between regions in terms of development and trade flows. Again, for Chinese regions, Banerjee, Duflo, and Qian (2020) find positive but small effects of access to transport infrastructure on regional per capita income across sectors during China's rapid expansion period. Their results show that there was no difference between Chinese provinces in terms of regional per capita income growth, contrary to Démurger (2001), which they link to the lack of factor mobility during this period.

The theory and evidence on the positive impact of infrastructure have been provided and suggested by the literature. Most of the literature deals with the infrastructure and trade relation by using country level data as in Limão & Venables (2001) and Nordås & Piermartini (2004). On the other hand, the studies that are at regional level focus on an area or country specifically such as Africa (Buys et al., 2010; Akpan, 2014) and Asia (Fujimura & Edmonds, 2006; Ismail & Mahyideen, 2015) using country level data and Turkey (Coşar & Demir, 2016), Spain (Bensassi et al., 2015), and US

(Duranton et al., 2014) using city or state level data. This thesis aims to contribute to the literature by focusing on intranational and international trade between the 200 NUTS 2 regions of 21 EU countries. Unlike the previous studies, differentials in regional trade flows with respect to different levels of transport infrastructure will be investigated using data from multiple countries. The purpose of this thesis is to fill the gap in the infrastructure and trade literature by taking on a regional perspective in a multi-country setting.

Before introducing the data and the econometric model, next section elaborates on the modes of freight transport in the EU. The prominence and performance of inland, airway and maritime freight transports are discussed in detail and this study's focus on road and railway infrastructure is also motivated in the following section.

3. MODES OF FREIGHT TRANSPORT IN THE EU

This section provides an overview of the modes of intra-EU freight transport. To provide a better survey on the transport infrastructure in the EU and its role in bilateral trade flows in the region, the modal split and its consequences for the analysis presented in this paper is discussed.

Based on the report by Eurostat, between the years 2008 and 2019, freight transport by road consistently constitutes nearly 75 per cent of the entire inland freight transport[1] in the EU where share of railway transport ranges between 17.6 to 19.2 per cent (2021). For Baltic Countries, the importance of railway transport is quite substantial: the share of railways in freight transport stays around 70 to 85 per cent over the years (Eurostat, 2021). For many EU member states, the share of railway networks in intra-EU freight transport decreased between 2008-2019 and this change has reflected itself as an increase in the share of roads (Eurostat, 2021). For countries such as Denmark, the situation was the opposite, where the upward change in railways has affected the use of roads for freight transportation negatively (Eurostat, 2021).

According to the same report by Eurostat, most of the intra-EU freight transport has been realized in Germany with 28.7 per cent of the entire EU international freight transport being performed in the country's roads whereas France, Poland, Spain, and Italy follow Germany's lead (2021).

The modal split of intra-EU freight, i.e., the share of the transport modes relative to the overall freight transport, reveals that transport by road is the most frequently used mode of transport by 53.4 per cent (Eurostat, 2021). Following road, maritime transport with 29.6 per cent and railways with 12.3 per cent contribute to intra-EU

[1] Inland transport modes include road, railways, and inland waterways.

freight transport whereas transport by inland waterways and air are less extensively used compared to the others (Eurostat, 2021). Among all five modes of transport, the amount of freight transported by sea increased the most over the years 2008-2019, the share of airways did not change and the use of the other three modes declined (Eurostat, 2021). As freight transport has increased in the EU at international level by 6.2 per cent, different modes of transport such as maritime, air and road have become more popular with increases 16.4, 7.9 and 5.2 per cent in shares, respectively, whereas transport by air and inland waterways has declined between 2008-2019 (Eurostat, 2021).

Even though the statistics given above do not cover the entire sample period of this paper, this information is relevant for the purposes of the analysis since these changes in the modes of transport may translate into interesting results regarding the role of infrastructure variables chosen. The prominence of different transport modes change through the years and the state of transport infrastructure may reflect this fact. Therefore, in the analysis, except for the inland waterways, all of the modes of transport mentioned here will be used as a robustness check to determine whether and which transport infrastructure affect trade flows in the EU. However, the specific focus of the paper will be on roads and railways since the level of infrastructure can be measured more clearly for these modes compared to airway and maritime transport. The next section clarifies the variables of interest as well as their sources.

4. DATA

The analysis in this thesis covers trade flows and infrastructure in NUTS 2 regions of the EU countries. NUTS regions refer to the statistical regions that are set up by Eurostat to establish a consistent definition for member states' divisions (European Commission, 2021). The borders and levels of these divisions and subdivisions are defined by Eurostat and the respective member country. Depending on the centralized or decentralized structure of countries in the EU, NUTS 2 regions may refer to states, provinces, or regions. For the purposes of this study, the 2010 version of the NUTS 2 classifications are used.

Based on transport infrastructure data's availability, the data that will be used in the analysis covers 21 countries that are members of the EU and their 200 NUTS 2 regions between the years 2000-2010[2]. Some of the current member states of EU (Cyprus, Slovenia, Bulgaria, Romania, and Croatia) are excluded from the sample due to lack of international trade data at NUTS 2 level. United Kingdom, who was a member during the sample period is also excluded because of the absent transport infrastructure data.

The following subsections lay out the data used in the analysis and their sources in detail.

4.1. Transport Infrastructure and Bilateral Distance

Transport infrastructure data are provided by the Eurostat and consist of motorways, other road networks, total railway lines and railway lines with double or more lines. These transport infrastructure measures are available both in kilometers and motorways and total railways are available also in per thousand square kilometers.

[2] Included countries are Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Slovakia, Spain, Sweden, and Northern Ireland. Names and NUTS 2 codes of the EU regions are provided in Appendix B.

Since total railway lines include railways with more than two tracks, a railway variable where double railways are subtracted from total railways is generated. Road and railway infrastructure variables will be normalized by using their logarithms in line with the previous studies in the literature.

Bilateral distance between the European NUTS 2 regions is provided by Persyn et al. (2020). The dataset contains arithmetic and harmonic averages of pairwise distance and time-related transport cost measures in the EU. In this analysis, the geodesic distance between two NUTS 2 regions which is averaged over many centroids for each region-pair is used as the bilateral distance measure.

4.2. Regional Trade Flows and Regional Income

Regional trade flows will be the dependent variable the gravity model. Data on regional exports is provided by the PBL EUREGIO dataset which is available for 249 EU NUTS 2 regions for the 2000-2010 period (Thissen et al., Thissen, Lankhuizen, van Oort, Los, & Diodato, 2018). EUREGIO dataset consists of estimated input-output tables for European regions and their intranational and international trade partners based on and consistent with the World Input-Output Database's (WIOD) 2013 release[3]. WIOD reports the imports and exports in FOB (free on board) prices and since not all countries report their exported product prices in FOB terms, Thissen et al. (2018) adjust the different products to present them in FOB prices at the product level by using a correction factor.

There are some aspects of international trade data that are not accounted for in this analysis with the use of EUREGIO dataset. Since the bilateral trade between two regions is aggregated over 15 product categories, the product composition of the goods transported by using certain modes of transport cannot be analyzed separately. For example, some valuable goods that are composed of light materials can only be transported by air. In this case, the measurement of freight by air may be misleading when measure of thousand tons of freight transported by air is used.

[3] The PBL EUREGIO database combines international trade data from WIOD, national accounts, and estimates for interregional trade that were presented by PBL Netherlands Environmental Assessment Agency. Thissen et al. (2018) uses the survey data available and constructs regional input-output tables without making any behavioral assumptions related to changes in economic, firm and policy decisions. This quality of the dataset allows for econometric analysis on these behavioral changes since they are excluded during the construction of the input-output tables. For the detailed discussion of these issues and estimation of interregional trade flows, please refer to Thissen et al. (2018).

Inability to track the routes of exports from the source region to the destination is another shortcoming of bilateral trade data that should be acknowledged. However, EUREGIO dataset corrects for re-exports, which refers to the products that are exported again by the importer region after a short period of time without any further industrial processing. The authors compute the complete origin-destination matrices and replaces the re-exporters with the source region where the exports were originating from (Thissen et al., 2018).

For the analysis, regional GDP (in million Euros) by NUTS 2 regions will be taken from Eurostat's Regional Economic Accounts database.

Table 1 below shows the summary statistics of the variables introduced above. Definitions and sources of each variable are presented in Appendix A. The minimum amount of trade has taken place between the two different regions of Spain whereas the maximum is between the same region of France, Île de France, which has also the highest GDP in the entire panel. The lowest GDP is realized in Åland, Finland with nearly 807.5 million Euros. In the panel, the largest bilateral distance is between the Canarias, Spain and Västsverige, Sweden.

As for the infrastructure measures, Andalucía, Spain is the region with the most advanced motorway infrastructure with 2391 kilometers. French region Rhône-Alpes has the largest network of railways with 2843 kilometers whereas the longest railways with double or more tracks are located in Lombardia, Italy with 1671 kilometers. The most extensive road infrastructure other than motorways belongs to Rhône-Alpes, France and Ciudad de Melilla in Spain hosts the sparsest network of other types of roads. For the other unit of measure of infrastructure which is per thousand square kilometers, German regions Bremen and Berlin lead with 186 kilometers of motorways and 708 kilometers of railways, respectively.

Summary statistics for the mountain index which will be used as an instrument for the infrastructure variables are also presented. Polish region Lódzkie is the most mountainous region in the entire panel, where many regions have non-mountain characteristic.

In addition to the variables of interest, airway and maritime transport efficiency measures as suggested by earlier studies (Bensassi et al., 2015) freight by air and sea are also summarized in Table 1. There are many regions that do not provide data on these variables simply because they do not have an airport or port in their region. Among those who have an airport, the region with most freight loaded and unloaded is Darmstadt, Germany. In Ciudad de Melilla, Spain 482,000 thousand tons of freight has been loaded and unloaded, which is the minimum in the panel. Zuid-Holland, Netherlands hosts the port where the maximum amount of intra-EU freight by sea is taken place with 407 million tons.

Table 1. Summary statistics

Variable	Number of observations	Mean	Standard Deviation	Min	Max
Trade flows (million US Dollars)	439,989	193.78	2910.51	$5.01 \cdot 10^{-7}$ (Ciudad de Melilla, ES - Andalucía, ES)	503806.6 (Île de France, FR - Île de France, FR)
GDP (million Euros)	440,000	45072.83	51327.22	807.46 (Åland, FI)	445517.6 (Île de France, FR)
Bilateral distance (km)	440,000	1120.068	677.84	1 (Ciudad de Ceuta, ES - Ciudad de Ceuta, ES)	4863 (Canarias, ES - Västsverige, SE)
Motorways (km)	358,000	288.691	306.61	0 (Multiple regions)	2391 (Andalucía, ES)
Total railways (km)	285,800	929.818	635.17	0 (Multiple regions)	2843 (Rhône-Alpes, FR)
Railways (km)	262,000	579.198	441.334	0 (Multiple regions)	2028 (Latvia, LT)
Railways with double or more lines (km)	262,200	348.9847	348.02	0 (Multiple regions)	1671 (Lombardia, IT)
Other roads (km)	352,800	16894.1	17552.84	26 (Ciudad de Melilla, ES)	92114 (Rhône-Alpes, FR)
Motorways (per thousand km)	347,000	30.31	29.39	0 (Multiple regions)	186 (Bremen, DE)
Railways (per thousand km)	226,600	71.03	82.09	0 (Multiple regions)	708 (Berlin, DE)
Mountain index (rated from 1 to 4)	440,000	3.541	0.64	1 (Lódzkie, PL)	4 (Multiple regions)
Freight by air (thousand tons)	263,800	77.249	264.137	0 (Multiple regions)	2270 (Darmstadt, DE)
Freight by sea (thousand tons)	181,200	$3.18 \cdot 10^{8}$	$4.51 \cdot 10^{7}$	482,000 (Ciudad de Melilla, ES)	$4.07 \cdot 10^{8}$ (Zuid-Holland, NL)

Since they are the indicators of a region's level of physical infrastructure all together, four main transport infrastructure variables are expected to be correlated among each other. The next table lays out the levels of correlation between each variable.

The highest correlation is observed between railway measures and other roads with 0.53 and 0.52. The correlation between motorways and double railways follows with 0.45.

Table 2. Correlations of the infrastructure variables

	Double railways$_i$	*Motorways$_i$*	*Other roads$_i$*	*Railways$_i$*
Double railways$_i$	1.00			
Motorways$_i$	0.454	1.00		
Other roads$_i$	0.534	0.264	1.00	
Railways$_i$	0.246	0.316	0.517	1.00

Next section introduces the econometric model that will be employed to explain trade flows with transport infrastructure in the EU.

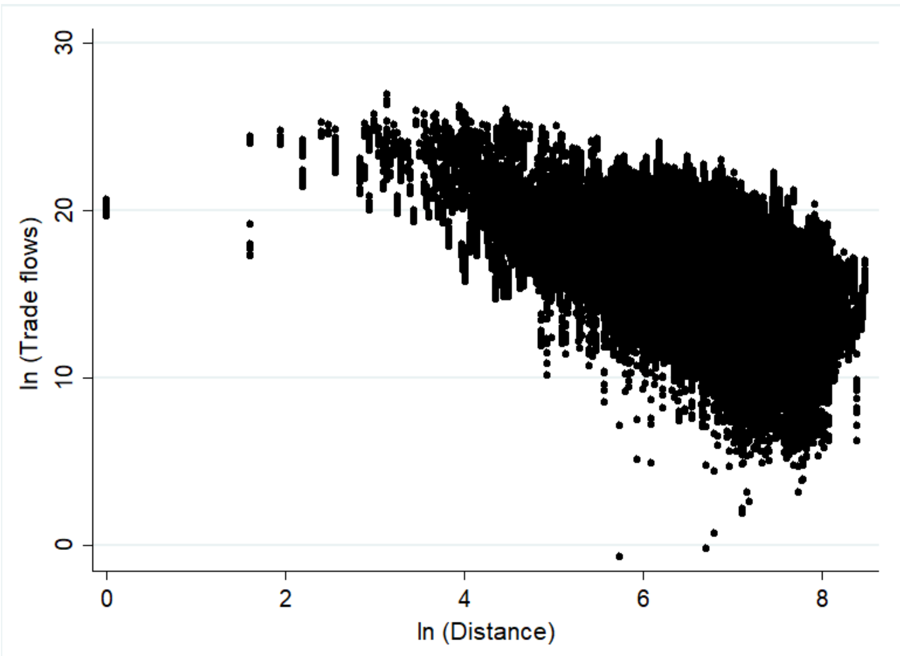
5. ECONOMETRIC MODEL

The current section explains and specifies the econometric model that will be used. In addition, potential estimation issues such as endogeneity bias are discussed, and possible solutions are proposed following the remedies adopted in the literature.

5.1. Baseline Model

Following the literature, the gravity model is used to estimate the extent to which infrastructure affects bilateral trade in European regions. To show that employing gravity model is a right strategy for studying trade flows and infrastructure, Figure 1 plots bilateral trade against bilateral distance in logarithms using the present data. As predicted by the gravity equation, distance and trade flows are negatively correlated when the EU NUTS 2 level data are used. In sum, the present data demonstrates features that are suitable for gravity method.

Figure 1. Bilateral trade flows and bilateral distance



Since Tinbergen (1962), gravity equation has been the main econometric model used by economists to explain the trade flows between countries or regions. Theoretical formulations of the atheoretical gravity model of Tinbergen (1962) was done by various studies such as Anderson (1979), Deardorff (1998) and Anderson and van Wincoop (2003). The gravity equation associates bilateral trade flows between two countries or regions with the variables capturing the characteristics of the respective regions, and the bilateral distance or other factors that can affect trade flows between the two regions. The gravity equations from various international trade models can all be expressed as a version of the following,

$$X_{i,j} = EX_i \, IM_j \, \tau_{i,j}^{-\sigma} \tag{1}$$

where, EX_i and IM_j may include source and destination region characteristics such as size or income of the region generally measured with GDP, whereas $\tau_{i,j}$ includes trade barriers or trade-enhancing factors such as bilateral distance, sharing the same religion, language, currency, border, etc. and σ is the trade elasticity measure which can be interpreted differently according to different international trade models (Arkolakis, Costinot, & Rodríguez-Clare, 2012).

Taking the logs of both sides of equation (1), gravity equation can be written in log-linear form for estimation purposes as the following,

$$ln\, X_{i,j} = ln\, EX_i + ln\, IM_j - \sigma\, ln\, \tau_{i,j} \tag{2}$$

To observe the effect of infrastructure on bilateral trade in the EU better and to make the comparison with the standard gravity models easier, the estimation results of equation (2) using the methodologies of Tinbergen (1962), McCallum (1995) and many authors such as Eaton and Kortum (2002) and Feenstra (2004) will also be presented. Tinbergen (1962) regresses the bilateral trade flows of countries on their respective incomes and bilateral distance with Ordinary Least Square (OLS) method. Whereas McCallum (1995) adds dummies to control for the border effect and home bias. Lastly, studies such as Feenstra (2004) adds exporter and importer fixed effects to control for the multilateral resistance terms that are a result of structural gravity equations (Head & Mayer, 2014).

In this study, the gravity approach is adopted to explain the relationship regional trade flows of EU and regional infrastructure. As explained in the previous section, the panel data has been defined using 200 EU NUTS 2 regions for the 2000-2010 period. In its most general form, the gravity model to be estimated in this study will be the following equation in order to explain the pairwise trade between the 200 regions in the EU,

$$lnTrade\ Flows_{i,j,t} = \alpha + \beta_0 lnGDP_{i,t} + \beta_1 lnGDP_{j,t} + \beta_2 Distance_{i,j} + \beta_3 Infrastructure_{i,t} + \beta_4 Infrastructure_{j,t} + \beta_5 Border_{i,j} + \beta_6 Same_{i,j} + \varepsilon_{i,j,t} \quad (3)$$

where, $Trade\ Flows_{i,j,t}$ is the exports from a particular region i to a region j at a given year t, $GDP_{i,t}$ and $GDP_{j,t}$ are the income of regions i and j for a given year t, and $Distance_{i,j}$ is the average bilateral distance between the centroids of the two regions. $Infrastructure_{i,t}$ is the vector containing transport infrastructure variables for region i and defined similarly for region j. $Border_{i,j}$ stands for the binary variable which takes the value of 1 when two regions share a border and 0 otherwise. $Same_{i,j}$ is a dummy variable and is equal to 1 when two regions belong to the same country and 0 when trade flows are at international level.

Hence, this formulation uses regional properties that are both time variant and time invariant. Time variant variables are $Trade\ Flows_{i,j,t}$, $GDP_{i,t}$, $GDP_{j,t}$, $Infrastructure_{i,t}$ and $Infrastructure_{j,t}$ and time invariant ones are $Distance_{i,j}$, $Border_{i,j}$, and $Same_{i,j}$.

The panel data properties of the data allow for two main alternative estimation methods, which are Fixed Effect (FE) Estimation and Random Effects (RE) Estimation depending on the assumption regarding the intercept term which are explained below in detail.

5.2. Specifications with Fixed Effects

The intercept term in equation (3) can be formulated as an individual-specific effect model with α_i for each region where α_i's are the characteristics unobserved to the researcher. Then two different formulations of the intercept are FE and RE models. In FE model, α_i's are permitted to be correlated with the regressors, $X_{i,j}$. Then the error in equation (3) becomes $u_{i,j,t} = \alpha_i + \varepsilon_{i,j,t}$, which will lead to a limited form of endogeneity, where $X_{i,j}$ is correlated with the time-invariant component of the error α_i, but uncorrelated with the idiosyncratic error $\varepsilon_{i,j,t}$. OLS estimates of the model will lead to biased and inconsistent coefficients. One possible estimation method will be to use FE Model, where fixed effects α_i of the model are eliminated by subtracting the corresponding group means, i.e., mean differencing. By choosing appropriate fixed effects, it is possible to alleviate these biases and obtain consistent estimates of the coefficients. In this study, the gravity equation will be formulated by defining alternative fixed effects.

The estimation will be conducted by employing exporter, importer, and time fixed effects, δ_i, δ_j and δ_t in equation (3)[4]. Including source region fixed effects, δ_i, absorbs the effects of time invariant determinants of trade flows. However, the effect of any variable that is time-invariant and representing a regional property of one region will be eliminated as well. For example, the determinants of trade flows due to a specific geographical characteristic such as being a coastal region cannot be identified separately in the presence of source and destination region fixed effects.

In the alternative RE model, on the other hand, α_i is assumed to be uncorrelated with the regressors. The estimation is done by Feasible Generalized Least Square (FGLS), which will lead to efficient and consistent coefficient estimates. In this model, it is possible to estimate coefficients for the time-invariant regressors. The appropriateness of the RE models will be statistically tested by using the Durbin-Wu-Hausman test since the violation of the above assumption will lead to inconsistent coefficient estimates (Cameron & Trivedi, 2005).

In line with the literature, to control for the common shocks and time invariant characteristics of the exporter, importer, and time, fixed effects δ_i, δ_j and δ_t will be added to equation (2), as explained above. However, there are studies that include these time invariant regressors, but use paired exporter and importer effects instead to identify the fixed effects.

The gravity equation is then specified as follows,

[4] For example, Donaubauer et al. (2018) include exporter-importer fixed effects and time fixed effects in their analysis for the similar reasons.

$$ln\,Trade\,Flows_{i,j,t} = \varphi_0 lnGDP_{i,t} + \varphi_1 lnGDP_{j,t} + \varphi_2 Infrastructure_{i,t} + \varphi_3 Infrastructure_{j,t} + \delta_{i,j} + \delta_t + \epsilon_{i,j,t} \quad (4)$$

where, $\delta_{i,j}$ is the pair fixed effects which eliminates the part of the error term in the gravity regressions. The region-pair fixed effects will absorb the time invariant bilateral covariates, such as $Border_{i,j}$, $Same_{i,j}$ and the bilateral distance term (Yotov et al., 2016).[5]

It is possible and necessary to extent the use of FE with alternative specifications such as pairwise, origin-time, and destination-time fixed effects. The use of exporter-time and importer-time fixed effects control for the multilateral resistance terms as suggested by Anderson and van Wincoop (2003). For cross-section data, adding exporter and importer region fixed effects is sufficient to control for the multilateral resistance terms. However, in the presence of panel data, fixed effects at the level of exporter-time and importer-time should be added to account for the multilateral resistance terms. In this specification, every variable that is at bilateral, exporter-time and importer-time level drop, therefore only the interactions of infrastructure variables can be identified. Aside from the econometric reasoning behind adding region-time fixed effects and interactions, it should also be recognized that bulk of the trade costs is generated in the origin and destination regions, domestically (Anderson & van Wincoop, 2004). Therefore, region-time fixed effects as well as the interactions between source and destination regions' transport infrastructure together should be added to account for these characteristics.

In the section where results of the above-mentioned specifications are presented, the fixed effects alternatives are incorporated into the estimations with different combinations to make the comparison between models more convenient.

The next subsection deals with the potential issues pertinent to the gravity model estimation.

5.3. Other Potential Estimation Issues

This subsection undertakes potential threats to the estimation strategy proposed above. The use of gravity specification to study the effects of infrastructure on trade

---

[5] In case of data where one expects problems in the bilateral distance measure, this equation will be the preferred specification of the analysis following Cheng and Wall (2005).

flows might include further pitfalls. On such issue is that the infrastructure variables may not be exogenous since the regions where there is relatively more trade may be attracting more investment compared to other regions, and thus have better infrastructure. This creates a reverse causality problem underlying the estimation. This potential simultaneity problem is addressed by researchers in earlier studies and some remedies have been proposed. For example, Bensassi et al. (2015) use Hausman-Taylor IV approach and add external excluded instruments and their lags as regressors. Coşar and Demir (2016) estimate an IV model, using the initial share of expressways along a province through gateway routes as an instrument. Crescenzi and Rodriguez-Pose (2012), aim to minimize the endogeneity by means of Generalized Method of Moments (GMM) estimators that use appropriate lags of the explanatory variables as instruments of their own current values. Lastly, Donaubauer et al. (2018) exclude major trade partners and extend the lags of the infrastructure index to account for reverse causality.

Since the current bilateral trade data is at NUTS 2 level and small regions may not be investing in a region's transport infrastructure they are trading with extensively, reverse causality may not be a serious issue in this case. However, following the previous studies, to deal with the potential problem of reverse causality, the econometric specifications will be estimated using an IV approach where a geographic typology index is used as an instrument for the endogenous time-variant infrastructure variables. In another specification, lagged versions of the explanatory variables will be added to overcome the endogeneity arising from reverse causality. Lastly, the Hausman-Taylor IV approach will also be adopted to control for the endogeneity.

Another problem with the use of gravity equation in explaining magnitude of international trade is the prominence of zeros in the bilateral trade matrix. In most gravity model estimations in the literature, the number of pairs of countries/regions that are not trading at all and hence has zero trade amount in the trade matrix are quite large. This, in the log-linear form, leads to heteroskedasticity in the error term causing biased and inconsistent estimates. The most common approach to overcome these issues is the PPML method as laid out by Santos-Silva and Tenreyro (2006). This issue is not a prevalent problem in the present data since there are only 11 missing values of bilateral trade in the entire panel. However, PPML methodology is employed to estimate the main gravity equations to control for small number of zero trade amounts and possible heteroskedasticity in the error term and to allow for comparisons.

The next section presents the results of the standard gravity equations, baseline, and fixed effects specifications together for a better comparison of the findings.

6. RESULTS

In the current panel, there are 40,000 distinct region pairs which create 440,000 observations covering 11 years. Based on the findings in the previous gravity literature, high levels of regional income are expected to affect trade positively whereas higher bilateral distance is expected to deter bilateral trade. The coefficients of transport infrastructure measures mentioned above are also expected to be positive after reviewing the previous studies related to regional infrastructure and trade.

6.1. Tinbergen, McCallum and Standard Structural Gravity Estimations

This section reports the results of the log-linearized version of the standard gravity model given in equation (2). First of these models is the main gravity model of Tinbergen (1962), home bias and border effect is included into the second one as suggested by McCallum (1995). The third and fourth are the models that include the suggestions of Feenstra (2004) and many others in the literature which add bilateral and time fixed effects and exporter, importer, and time fixed effects, respectively. The results are presented in Table 3.

First column where trade flows are regressed on regional GDP and bilateral distance using OLS method bears resemblance to the results of the standard gravity models in earlier studies. As the statistically robust gravity equation suggests, the coefficients of exporter and importer regions' income are close to one, positive and highly significant throughout different specifications. Bilateral distance has the expected negative sign where 1 per cent increase in distance between the EU regions decreases trade by 1.3 to 1.6 per cent at 1 per cent significance level. Sharing a border and intranational trade are also favorable characteristics for trade flows: the coefficient of $Border_{ij}$ ranges between 0.12 and 0.46, varying in significance, whereas $Same_{ij}$ has a coefficient of 1.18 or 1.27, depending on the specification.

In sum, as Table 3 shows, findings of the traditional gravity estimations from previous studies can be replicated in our sample which uses region level data from the EU. Since the focus of this thesis is transport infrastructure and its effects on intra-EU trade, next subsections elaborate on this relationship in detail. In order to capture consistent and unbiased estimates for the included variables, various methodologies are also employed and remedies for endogeneity issues are discussed..

Table 3. Traditional gravity estimation results

Dependent variable *ln Trade Flows*$_{ijt}$	(1) Tinbergen	(2) McCallum	(3) FE estimation	(4) FE estimation
ln GDP$_{it}$	0.962***	0.958***	0.800***	0.801***
	(0.005)	(0.005)	(0.012)	(0.012)
ln GDP$_{jt}$	0.795***	0.790***	0.715***	0.715***
	(0.006)	(0.006)	(0.012)	(0.012)
ln Distance$_{ij}$	-1.611***	-1.327***	-1.531***	
	(0.009)	(0.009)	(0.012)	
Border$_{ij}$		0.462***	0.124**	
		(0.054)	(0.052)	
Same$_{ij}$		1.269***	1.181***	
		(0.030)	(0.028)	
Constant	8.908***	6.941***	10.728***	0.420**
	(0.121)	(0.108)	(0.199)	(0.181)
Observations	439,989	439,989	439,989	439,989
Number of region pairs	40,000	40,000	40,000	40,000
Number of regions	400	400	400	400
R-squared	0.744	0.765	0.844	0.979
Durbin-Wu-Hausman test (prob $> \chi^2$)	-	-	0.000	0.000
Exporter FE	N	N	Y	N
Importer FE	N	N	Y	N
Time FE	N	N	Y	Y
Exporter-time FE	N	N	N	N
Importer-time FE	N	N	N	N
Bilateral FE	N	N	N	Y

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Standard errors are clustered at bilateral level.

6.2. Main Results

Results of the main analysis of the paper is presented below in tables 4, 5, 6, 7 and 8. Table 4 demonstrates the OLS results for the baseline specification. Table 5 shows the results of the gravity equation and its variations by adding combinations of fixed effects and interactions. Table 6 and 7 include alternative estimation methods some of which are used to address the heteroskedasticity and endogeneity problem due to possible simultaneity. By adopting the IV method, Table 8 attempts to solve the endogeneity bias pertinent to the infrastructure variables.

Table 4 adds infrastructure variables gradually and report the OLS estimates of the coefficients. Signs and significance levels of the three main gravity variables i.e., regional income and bilateral distance measures do not change throughout the table. However, the significance and sign of motorway infrastructure change as other infrastructure variables are included. For instance, when trade flows are regressed on the main gravity variables and source and destination region's motorway network structure, only destination region's motorway infrastructure affects trade positively. When logarithm of kilometers of total railways is included, the impact of the destination region's motorways continued to be statistically significant and with a positive contribution to amount of trade flows. Railway infrastructure, regardless of belonging to source or destination region, seems to have a positive impact on trade. In this set of estimations, the other two infrastructure variables, double railways and other roads are not accordant with the predictions at the beginning of this section both of which have negative and significant effect on trade.

As was explained above, α_i's that include the time invariant unobserved heterogeneity may be correlated with the regressors, $X_{i,j}$. Then the error in equation (3) becomes $u_{i,j,t} = \alpha_i + \varepsilon_{i,j,t}$, which will lead to a limited form of endogeneity, where $X_{i,j}$ is correlated with the time-invariant component of the error α_i but uncorrelated with the idiosyncratic error $\varepsilon_{i,j,t}$. OLS estimates of the model will lead to biased and inconsistent coefficients. Hence, one has to be careful when interpreting the results in Table 4.

This form of endogeneity has been addressed by using FE estimations in Table 5. When exporter, importer and time fixed effects are included in the first specification, the source and destination regions' motorway infrastructure and railway network affect trade positively at a significance level of 1 per cent. This effect ranges between 0.05 and 0.13. The second column adds bilateral and time fixed effects, therefore all the coefficients of variables at bilateral level such as $Distance_{ij}$, $Border_{ij}$ and $Same_{ij}$ cannot be identified. Here, the infrastructure variables for source and destination take on positive and significant values except for the other roads measure. In column 2, the effect of transport measures on trade flows ranges between 0.05 and 0.13.

In the third column of Table 5, source and destination region interactions of each infrastructure variable are included. In this specification, when exporter-time and importer-time fixed effects are included in addition to bilateral fixed effects to control for the multilateral resistance terms, every time-variant and bilateral variable drops. Therefore, only the interaction variables are identified. Here, only the interaction of railway networks is significantly positive whereas interactions of double railways and other roads are significantly negative, with a smaller coefficient than railways interaction variable. To interpret the coefficients in column 3, as the railway infrastructure of both source and destination region are larger, change in trade flows is higher. For railways with more than two tracks and other roads, the percentage change in bilateral trade is lower when these two measures are higher for a region pair.

After each specification in the following tables where FE estimation is used, the result of the Durbin-Wu-Hausman test for FE or RE specification is presented. In this section, for every specification, the null hypothesis of the test which states that there is no systematic difference between the coefficients of the two estimations is rejected. Therefore, alternative specifications that use RE estimation are not included where Durbin-Wu-Hausman test rejects the null hypothesis.

The estimation results of the regressions in Table 5 might suffer from two possible weaknesses. As was mentioned above, one issue is the zero trade flows and the second issue is the possibility that the infrastructure variables may not be strictly exogenous, which may be the cause of endogeneity and bias in the estimations. Especially if the regions that export and/or receive large trade amounts might have higher income growth and possibly larger financial capacity to finance improvement and expansion of transportation infrastructures. These issues are addressed in the regression estimations of tables 6 and 7.

Table 4. Ordinary Least Squares (OLS)					
	Dependent variable: *ln Trade Flows*$_{ijt}$				
	(1)	(2)	(3)	(4)	(5)
ln GDP$_{it}$	0.899***	0.850***	0.903***	0.834***	0.901***
	(0.009)	(0.013)	(0.017)	(0.013)	(0.017)
ln GDP$_{jt}$	0.699***	0.565***	0.636***	0.533***	0.630***
	(0.009)	(0.013)	(0.017)	(0.013)	(0.017)
ln Distance$_{ij}$	-1.162***	-1.369***	-1.480***	-1.346***	-1.452***
	(0.011)	(0.021)	(0.026)	(0.021)	(0.026)
ln Motorways$_{it}$	-0.018**	0.015	0.060***	0.045***	0.072***
	(0.008)	(0.011)	(0.013)	(0.012)	(0.013)
ln Motorways$_{jt}$	0.050***	0.096***	0.160***	0.151***	0.186***
	(0.007)	(0.011)	(0.012)	(0.011)	(0.012)
ln Railways$_{it}$			0.099***		0.161***
			(0.012)		(0.012)
ln Railways$_{jt}$			0.150***		0.285***
			(0.011)		(0.012)
ln Double railways$_{it}$			-0.123***		-0.087***
			(0.013)		(0.013)
ln Double railways$_{jt}$			-0.177***		-0.103***
			(0.012)		(0.012)
ln Total railways$_{it}$		0.042***		0.151***	
		(0.012)		(0.014)	
ln Total railways$_{jt}$		0.113***		0.308***	
		(0.012)		(0.013)	
ln Other roads$_{it}$				-0.116***	-0.124***
				(0.009)	(0.010)
ln Other roads$_{jt}$				-0.204***	-0.259***
				(0.009)	(0.011)
Border$_{ij}$	0.609***	0.423***	0.342***	0.451***	0.369***
	(0.061)	(0.084)	(0.092)	(0.082)	(0.090)
Same$_{ij}$	1.356***	1.439***	1.289***	1.482***	1.343***
	(0.038)	(0.059)	(0.064)	(0.059)	(0.065)
Constant	7.305***	9.102***	9.154***	9.945***	10.676***
	(0.136)	(0.224)	(0.264)	(0.223)	(0.267)
Observations	240,777	109,727	81,453	109,727	81,453
R-squared	0.704	0.706	0.719	0.717	0.732
Durbin-Wu-Hausman test (prob > χ^2)	-	-	-	-	-
Exporter FE	N	N	N	N	N
Importer FE	N	N	N	N	N
Time FE	N	N	N	N	N
Exporter-time FE	N	N	N	N	N
Importer-time FE	N	N	N	N	N
Bilateral FE	N	N	N	N	N

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at region pair level.

Table 5. Fixed Effects (FE) Estimation			
	Dependent variable: *ln Trade Flows*$_{ijt}$		
	(1)	(2)	(3)
ln GDP$_{it}$	0.987***	0.841***	
	(0.037)	(0.022)	
ln GDP$_{jt}$	0.680***	0.682***	
	(0.038)	(0.017)	
ln Distance$_{ij}$	-1.747***		
	(0.029)		
ln Motorways$_{it}$	0.111***	0.080***	
	(0.017)	(0.009)	
ln Motorways$_{jt}$	0.054***	0.053***	
	(0.016)	(0.008)	
ln Railways$_{it}$	0.125***	0.083***	
	(0.020)	(0.010)	
ln Railways$_{jt}$	0.061***	0.129***	
	(0.021)	(0.012)	
ln Double railways$_{it}$	0.047***	0.051***	
	(0.012)	(0.007)	
ln Double railways$_{jt}$	-0.003	0.053***	
	(0.014)	(0.008)	
ln Total railways$_{it}$			
ln Total railways$_{jt}$			
ln Other roads$_{it}$	-0.043***	-0.013	
	(0.015)	(0.008)	
ln Other roads$_{jt}$	-0.027*	-0.045***	
	(0.014)	(0.007)	
Border$_{ij}$	-0.124		
	(0.081)		
Same$_{ij}$	1.288***		
	(0.056)		
ln Motorways$_{it}$ * *ln Motorways*$_{jt}$			-0.007
			(0.004)
ln Railways$_{it}$ * *ln Railways*$_{jt}$			0.018***
			(0.007)
ln Double railways$_{it}$ * *ln Double railways*$_{jt}$			-0.009**
			(0.003)
ln Other roads$_{it}$ * *ln Other roads*$_{jt}$			-0.037***
			(0.005)
Constant	9.982***	-1.512***	19.853***
	(0.632)	(0.330)	(0.533)
Observations	81,453	80,795	80,795
R-squared	0.814	0.977	0.984
Durbin-Wu-Hausman test (prob > χ^2)	0.000	0.000	0.000
Exporter FE	Y	N	N
Importer FE	Y	N	N
Time FE	Y	Y	N
Exporter-time FE	N	N	Y
Importer-time FE	N	N	Y
Bilateral FE	N	Y	Y

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Standard errors are clustered at region pair level.

Even though there are only 11 missing values of bilateral trade flows, in line with the literature, the PPML approach is adopted in Table 6 to control for the zeros in the trade matrix and the heteroskedasticity problem in the error term. As Santos-Silva and Tenreyro (2006) show, logarithm of bilateral trade flows in the gravity model causes a serious estimation problem because of the prominence of zero trade flows in a trade matrix. The reason is that when bilateral trade flows are used in log-linearized form, zeros in the trade matrix become undefined and these observations are lost. Also, when there is heteroskedasticity in the error terms, log-linearization and OLS estimation together result in biased estimates of true trade elasticities (Santos-Silva & Tenreyro, 2006). Therefore, estimating the multiplicative form of the gravity equation as in equation (1) with PPML is proposed. This way, zeros in the trade matrix and inconsistent estimates in the presence of heteroskedasticity is taken care of.

First column gives the results of the PPML method which also adds time, exporter, and importer fixed effects. Second column alternates the region and time fixed effects with bilateral and time fixed effects. Next one controls for the multilateral resistance terms and bilateral fixed effects. Comparing with the previous table, estimation with PPML alters the previous findings significantly and almost all of the coefficients lose their significance. When a similar method of estimation, Generalized Linear Models (GLM) where trade flows are assumed to be distributed with Poisson is used as a robustness check, similar results are observed, therefore it is not reported here.

One simple method to avoid the simultaneity bias will be to include the explanatory variables in terms of their lagged values. This method will avoid the reverse causation of trade determining the amount of infrastructure. The results are presented in Table 7. In the regressions in columns one and two, one period and two period lags of the transport infrastructure variables are included, respectively. This approach will alleviate the endogeneity if it allows sufficient time lag for increased exports to cause significant change in transport infrastructure. Except for other roads, all of the infrastructure measures have their expected positive signs and highly significance coefficients. The effect of various transport infrastructure measures on bilateral trade flows ranges between 0.05 percent to 0.12 percent.

Lastly, the third column of Table 7 gives the Hausman-Taylor IV estimation results where the lagged values of the infrastructure variables become instrumental variables for the time-variant and potentially endogenous infrastructure variables. The advantage of the Hausman-Taylor IV method is that it provides estimates that are between FE and RE estimation methods. In this case, it is a hybrid of these two methods since endogenous and time-variant infrastructure variables are assumed to be correlated with the individual effects in the error term whereas the other variables are assumed to be exogenous. The results of this estimation suggest positive and significant effect of railway network for the region pairs. Motorways in the source region also affects bilateral trade positively. Rest of the variables are either negative and significant or insignificant.

As a final endogeneity control, in Table 8, an IV estimation is run. Here, the variable chosen as an instrument for the infrastructure variables is the mountain index for source and destination region. This index is higher for regions with mostly plain areas. It is lower for regions that either have a large population that live in mountain areas or when the region consists mostly of mountainous areas. The regions score between 1 and 4 in this index. As a geographic characteristic which is exogenous, mountains are expected to hinder trade by making construction of new infrastructure more difficult. Also, this variable has an acceptable level of correlation with the endogenous infrastructure variables[6] and high F-statistics for first stage results therefore it is used as instrument for the time-variant endogenous variables. Figure 2 demonstrates the linear relationship between the mountain index and the infrastructure variables.

[6] The correlations between mountain index and infrastructure measures are all between 11-24 per cent in absolute value.

Figure 2. Correlation between the main infrastructure variables and mountain index

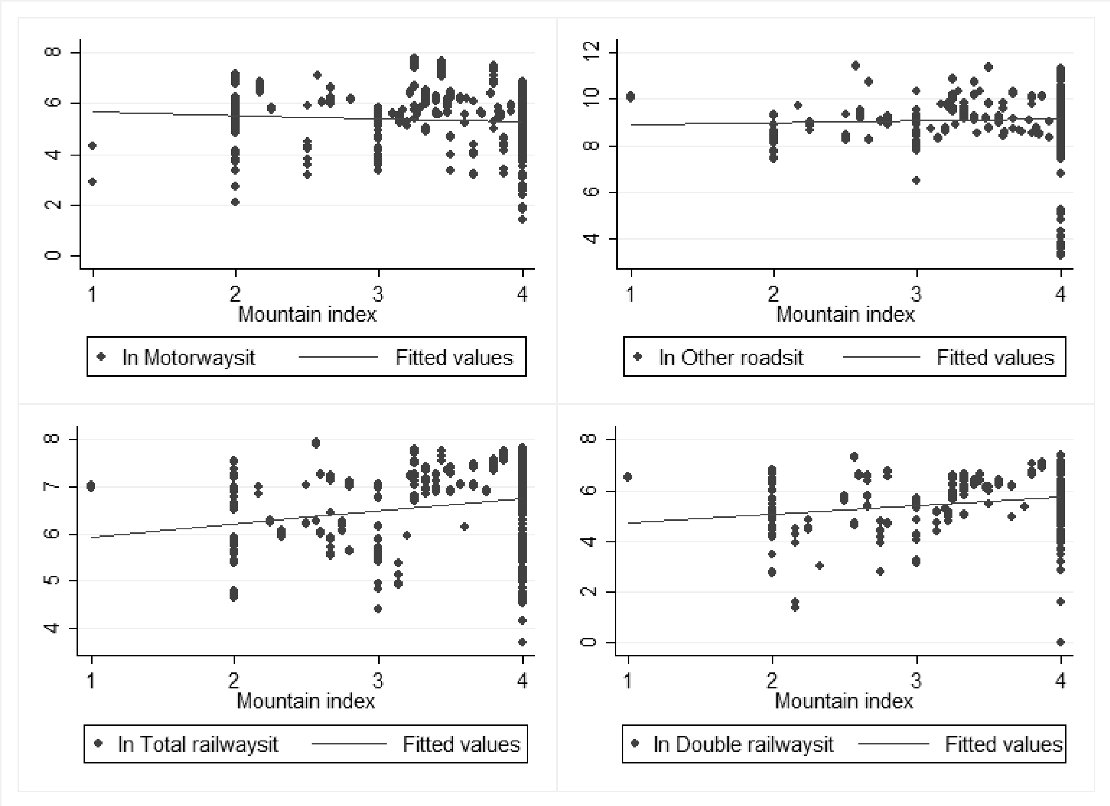


Table 6. Poisson Pseudo Maximum Likelihood (PPML)

Dependent variable: *Trade Flows*$_{ijt}$			
	(1)	(2)	(3)
ln GDP$_{it}$	0.713***	0.720***	
	(0.109)	(0.102)	
ln GDP$_{jt}$	0.172*	0.167*	
	(0.099)	(0.090)	
ln Distance$_{ij}$	-1.875***		
	(0.094)		
ln Double railways$_{it}$	0.016	0.006	
	(0.019)	(0.018)	
ln Double railways$_{jt}$	-0.012	-0.005	
	(0.017)	(0.016)	
ln Motorways$_{it}$	0.007	0.003	
	(0.041)	(0.042)	
ln Motorways$_{jt}$	0.054	0.071*	
	(0.040)	(0.041)	
ln Railways$_{it}$	0.017	0.026	
	(0.049)	(0.043)	
ln Railways$_{jt}$	0.047	0.058	
	(0.045)	(0.040)	
ln Other roads$_{it}$	-0.036	-0.041	
	(0.039)	(0.025)	
ln Other roads$_{jt}$	-0.016	-0.011	
	(0.038)	(0.024)	
Border$_{ij}$	-0.209		
	(0.135)		
Same$_{ij}$	1.618***		
	(0.102)		
ln Motorways$_{it}$ * *ln Motorways*$_{jt}$			0.054***
			(0.012)
ln Railways$_{it}$ * *ln Railways*$_{jt}$			-0.003
			(0.018)
ln Double railways$_{it}$ * *ln Double railways*$_{jt}$			0.002
			(0.002)
ln Other roads$_{it}$ * *ln Other roads*$_{jt}$			0.028
			(0.018)
Constant	20.121***	12.950***	18.890***
	(1.447)	(1.335)	(1.821)
Observations	81,453	80,795	80,795
Number of region pairs	12,699	12,041	12,041
Pseudo R-squared	0.914	0996	0.998
Durbin-Wu-Hausman test (prob > χ^2)	-	-	-
Exporter FE	Y	N	N
Importer FE	Y	N	N
Time FE	Y	Y	N
Exporter-time FE	N	N	Y
Importer-time FE	N	N	Y
Bilateral FE	N	Y	Y

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Standard errors are clustered at region pair level.

Table 7. Endogeneity controls			
Estimation method	FE 1 period lag	FE 2 period lag	Hausman-Taylor IV
Dependent variable	*ln Trade* $Flows_{ijt+1}$	*ln Trade* $Flows_{ijt+2}$	*ln Trade* $Flows_{ijt}$
ln GDP_{it}	0.524***	0.605***	0.461***
	(0.023)	(0.026)	(0.019)
ln GDP_{jt}	0.626***	0.439***	0.286***
	(0.017)	(0.020)	(0.016)
ln $Distance_{ij}$			-2.135***
			(0.037)
ln Double $railways_{it}$	0.073***	0.052***	0.005
	(0.007)	(0.009)	(0.010)
ln Double $railways_{jt}$	0.050***	0.018*	0.004
	(0.009)	(0.010)	(0.010)
ln $Motorways_{it}$	0.115***	0.101***	0.031**
	(0.009)	(0.010)	(0.014)
ln $Motorways_{jt}$	0.052***	0.107***	0.003
	(0.007)	(0.008)	(0.012)
ln $Railways_{it}$	0.098***	0.082***	0.162***
	(0.011)	(0.015)	(0.014)
ln $Railways_{jt}$	0.096***	0.082***	0.209***
	(0.013)	(0.017)	(0.016)
ln Other $roads_{it}$	-0.035***	-0.042***	0.007
	(0.008)	(0.009)	(0.013)
ln Other $roads_{jt}$	-0.060***	-0.052***	-0.026**
	(0.007)	(0.007)	(0.012)
$Border_{ij}$			-0.732***
			(0.093)
$Same_{ij}$			0.957***
			(0.071)
Constant	2.609***	4.059***	20.742***
	(0.332)	(0.375)	(0.395)
Observations	67,529	54,361	81,453
Number of region pairs	10,816	8,464	12,699
R-squared	0.979	0.979	-
Durbin-Wu-Hausman test (prob $> \chi^2$)	0.000	0.000	-
Exporter FE	N	N	N
Importer FE	N	N	N
Time FE	Y	Y	N
Exporter-time FE	N	N	N
Importer-time FE	N	N	N
Bilateral FE	Y	Y	N

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Standard errors are clustered at region pair level.

Table 8. IV regressions for trade flows using mountain index as instrument

	Panel A. Two-Stage Least Squares			
Dependent variable *ln Trade Flows*$_{ijt}$	(1)	(2)	(3)	(4)
ln GDP$_{it}$	1.017***	0.878***	0.844***	0.787***
	(0.056)	(0.016)	(0.024)	(0.053)
ln GDP$_{jt}$	0.244***	0.807***	0.910***	1.156***
	(0.054)	(0.016)	(0.022)	(0.051)
ln Distance$_{ij}$	-1.245***	-1.577***	-1.618***	-1.225***
	(0.024)	(0.018)	(0.024)	(0.019)
Same$_{ij}$	1.227***	1.454***	1.353***	1.381***
	(0.044)	(0.048)	(0.055)	(0.054)
Border$_{ij}$	0.465***	0.245***	0.210**	0.748***
	(0.076)	(0.073)	(0.083)	(0.105)
ln Motorways$_{it}$	-0.166**			
	(0.082)			
ln Motorways$_{jt}$	0.753***			
	(0.080)			
ln Total railways$_{it}$		0.264***		
		(0.050)		
ln Total railways$_{jt}$		-0.389***		
		(0.047)		
ln Double railways$_{it}$			0.113***	
			(0.041)	
ln Double railways$_{jt}$			-0.466***	
			(0.037)	
ln Other roads$_{it}$				0.272**
				(0.131)
ln Other roads$_{jt}$				-0.967***
				(0.121)
Constant	8.952***	10.714***	11.469***	11.294***
	(0.381)	(0.349)	(0.244)	(0.861)
Observations	240,777	169,608	118,466	283,905
R-squared	0.627	0.709	0.711	0.347

Panel B. First stage results

Dependent variable	(1) *ln Motor* $ways_{it}$	(2) *ln Motor* $ways_{jt}$	(3) *ln Total* $railways_{it}$	(4) *ln Total* $railways_{jt}$	(5) *ln Double* $railways_{it}$	(6) *ln Double* $railways_{jt}$	(7) *ln Other* $roads_{it}$	(8) *ln Other* $roads_{jt}$
ln GDP_{it}	0.676***	-0.006	0.268***	-0.000	0.556***	0.010	0.396***	0.004
	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)	(0.001)	(0.010)	(0.007)
ln GDP_{jt}	-0.006	0.673***	-0.000	0.268***	0.000	0.556***	0.004	0.396***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.008)	(0.007)	(0.007)	(0.010)
ln $Distance_{ij}$	0.124***	0.124***	0.023	0.023	-0.184***	-0.183***	0.065***	0.065***
	(0.009)	(0.009)	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.014)
$Same_{ij}$	0.189***	0.187***	-0.002***	-0.002***	-0.099***	-0.099***	-0.151***	-0.152***
	(0.017)	(0.017)	(0.021)	(0.021)	(0.029)	(0.029)	(0.033)	(0.033)
$Border_{ij}$	0.226***	0.238***	0.180***	0.180***	-0.136***	-0.132**	0.456***	0.461***
	(0.029)	(0.029)	(0.033)	(0.033)	(0.045)	(0.046)	(0.045)	(0.044)
$Mountain_i$	-0.160***	0.020**	0.257***	0.001	0.367***	-0.010	0.124***	0.005
	(0.009)	(0.008)	(0.008)	(0.008)	(0.015)	(0.011)	(0.011)	(0.013)
$Mountain_j$	0.019**	-0.160***	-0.001	0.257***	-0.010	0.387***	0.005	0.124***
	(0.008)	(0.009)	(0.008)	(0.008)	(0.011)	(0.015)	(0.013)	(0.011)
Constant	-1.930***	-1.939***	3.035***	3.035***	0.1	0.096	4.165***	4.162***
	(0.108)	(0.108)	(0.116)	(0.115)	(0.162)	(0.162)	(0.158)	(0.158)
Observations	240,777	240,777	169,608	169,608	118,466	118,466	283,905	283,905
F-statistic	1906.01	1907.37	449.25	449.39	537.50	537.09	323.83	337.87
R-squared	0.396	0.396	0.161	0.161	0.311	0.311	0.099	0.099

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at region pair level. Each specification includes time fixed effects.

Since there is only one instrument (for each source and destination infrastructure measure) and endogenous variables, each specification in Table 8 regresses each infrastructure variable on the mountain index to satisfy at least just-identification. First stage regressions in Panel B show expected coefficients for mountain index in each column similar to Figure 2. In the second stage results in Panel A, each infrastructure variable seems to explain trade flows significantly however with inverted signs for source and destination regions. For example, given its negative correlation with the mountain index, motorways contribute to trade flows with -0.17 per cent increase in the source region whereas this effect is 0.75 for the destination region. As mentioned above, since the mountain index is a time-fixed variable that is constant over the entire panel, it may not be a perfect instrument for time-variant infrastructure measures. However, given the data availability and restrictions regarding the regional level analysis, this instrument is found to be the only option for the IV method.

The next section runs the robustness checks and discusses the sensitivity of the main results to different specifications.

6..3. Robustness Checks

To ensure that the results of the estimation are robust to different definitions of the infrastructure variables and different modes of transportation, this section provides the necessary checks. As the first robustness check, motorway, and railway infrastructure endowment per thousand square kilometers will be added as regressors based on data availability. Secondly, to better understand the determinant role of road and railway infrastructure in the intra-EU trade, airport and port efficiency parameters will be included following Bensassi et al. (2015). The results are presented below in Table 9 and 10.

First column of Table 9 includes source and destination regions' motorways and railways infrastructure and adds exporter, importer, and time fixed effects. In this specification, only destination region's motorway network affects trade flows significantly, and its coefficient is 0.084. Both source and destination regions' railway infrastructure affect trade flows significantly with coefficients 0.11 and 0.13. For the second column, with the addition of bilateral fixed effects, only the significance level of $Railways_{it}$ change. One per cent increase in the source region's motorway network affects trade flows with a highly significant 0.085 per cent increase. The effect of destination region's railway infrastructure increases in magnitude in column 2 with 0.17. To control for the multilateral resistance terms, the third specification adds bilateral, exporter-time and importer-time fixed effects and interactions of the infrastructure variables. Similar to the main results, the interaction of motorway networks is not significant whereas for railways, the interaction of source and destination region affects trade flows significantly and positively.

Since the result of Durbin-Wu-Hausman test in column 3 fails to reject H_0, which states that coefficients produced by FE and RE estimations are not systematically different, column 4 shows the results of the RE estimation. Last column suggests negative but significant role of motorways and railways in source and destination regions, contrary to the previous expectations.

Table 9. Robustness check – Alternative infrastructure unit of measures				
	(1)	(2)	(3)	(4)
Dependent variable:	FE	FE	FE	RE
ln Trade Flows$_{ijt}$	estimation	estimation	estimation	estimation
ln GDP$_{it}$	0.933***	0.760***		0.766***
	(0.047)	(0.025)		(0.013)
ln GDP$_{jt}$	0.382***	0.516***		0.593***
	(0.05)	(0.022)		(0.013)
ln Distance$_{ij}$	-2.112***			-1.297***
	(0.011)			(0.020)
ln Motorways$_{it}$	0.084***	0.085***		-0.140***
	(0.02)	(0.009)		(0.021)
ln Railways$_{it}$	0.111***	0.039**		-0.506***
	(0.037)	(0.016)		(0.064)
ln Motorways$_{jt}$	-0.002	-0.009		-0.244***
	(0.019)	(0.008)		(0.023)
ln Railways$_{jt}$	0.128***	0.169***		-0.484***
	(0.035)	(0.017)		(0.064)
Border$_{ij}$	-0.349***			0.459***
	(0.033)			(0.091)
Same$_{ij}$	0.422***			1.391***
	(0.027)			(0.073)
ln Motorways$_{it}$ * *ln Motorways*$_{jt}$			0.003	0.059***
			(0.005)	(0.007)
ln Railways$_{it}$ * *ln Railways*$_{jt}$			0.035***	0.122***
			(0.013)	(0.015)
Constant	15.969***	2.289***	15.854***	13.650***
	(0.861)	(0.401)	(0.226)	(0.328)
Observations	67,360	65,193	65,193	67,360
R-squared	0.809	0.973	0.981	0.712
Durbin-Wu-Hausman test (prob > χ^2)	0.000	0.000	0.199	-
Exporter FE	Y	N	N	N
Importer FE	Y	N	N	N
Time FE	Y	Y	N	N
Exporter-time FE	N	N	Y	N
Importer-time FE	N	N	Y	N
Bilateral FE	N	Y	Y	N

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. For the random effects estimation, reported R^2 is the overall.

In Table 10, in addition to the standard gravity variables, motorway and railway infrastructure, logarithms of freight transported by air and sea are included in the FE estimation. The results in the first column indicate positive and significant effect of port efficiency in the source region, measured by thousand tons of loaded and unloaded freight in ports. From other types of transport, only the coefficient of railways in the exporter region is significant with 0.39. When bilateral fixed effects replace exporter and importer fixed effects in the next column, *ln Total railways*$_{it}$ preserves its significant and positive coefficient where *ln Total railways*$_{jt}$ gains significance and affects trade flows with 0.16 with 10 per cent significance level. Maritime transport efficiency in source region affects bilateral trade significantly with an increase of 0.13. Airport efficiency in exporter and importer regions and exporter's motorway network contribute to trade flows negatively. However, as mentioned in the fourth section, since valuable and light materials are generally transported by air, the effect of tons of freight by air on trade flows should be interpreted carefully. In the third column, bilateral and region-time fixed effects are added to control for the multilateral resistance terms. Therefore, as in the previous tables, interactions of the infrastructure and efficiency variables are used as regressors since every other variable at region-time and bilateral level drop. The results of this FE estimation follow that only interaction term with the expected positive and significant sign is total railways where others either have unexpected signs or do not have a significant effect on trade flows.

Thus, based on the results in column 2 and 3 of Table 9, it can be said that the main estimation results are not sensitive to the choice of unit of measure for infrastructure variables. Findings summarized in Table 10 can be interpreted as a reflection of the statistics presented in section 3 to some extent. As mentioned above, maritime transport has become more popular between 2008-2019 and the extent to which how efficient ports operate seems to be important for bilateral trade flows in the EU. As for the airports, since they account for a very small portion of intra-EU freight transportation and their share has declined over the years 2008-2019 and considering the frequent transport of light materials by air, an insignificant or negative coefficient may have been expected. It is negative and significant for source and destination regions when bilateral and time fixed effects are absorbed. On the other hand, impact of motorway infrastructure becomes negative and significant for source region when bilateral and time fixed effects are added. So, it is difficult to interpret whether the roles of these four modes of transport is in line with the findings in section three.

Table 10. Robustness check – All modes of freight transport			
	(1)	(2)	(3)
Dependent variable: *ln Trade Flows*$_{ijt}$	FE estimation	FE estimation	FE estimation
ln GDP$_{it}$	1.423***	0.777***	
	(0.204)	(0.080)	
ln GDP$_{jt}$	1.129***	0.600***	
	(0.202)	(0.071)	
ln Distance$_{ij}$	-1.611***		
	(0.034)		
ln Freight by air$_{it}$	-0.029	-0.046**	
	(0.042)	(0.019)	
ln Freight by air$_{jt}$	-0.008	-0.037**	
	(0.039)	(0.017)	
ln Freight by sea$_{it}$	0.181*	0.125**	
	(0.101)	(0.055)	
ln Freight by sea$_{jt}$	0.004	0.046	
	(0.101)	(0.040)	
ln Motorways$_{it}$	0.304	-0.237***	
	(0.195)	(0.091)	
ln Motorways$_{jt}$	0.122	-0.043	
	(0.192)	(0.090)	
ln Total railways$_{it}$	0.390***	0.388***	
	(0.126)	(0.064)	
ln Total railways$_{jt}$	-0.166	0.156*	
	(0.142)	(0.092)	
Border$_{ij}$	0.089		
	(0.085)		
Same$_{ij}$	1.646***		
	(0.049)		
ln Freight by air$_{it}$ * *ln Freight by air*$_{jt}$			-0.012*
			(0.006)
ln Freight by sea$_{it}$ * *ln Freight by sea*$_{jt}$			0.005
			(0.023)
ln Motorways$_{it}$ * *ln Motorways*$_{jt}$			-0.095*
			(0.054)
ln Total railways$_{it}$ * *ln Total railways*$_{jt}$			0.112**
			(0.050)
Constant	-6.925*	-2.457	14.173*
	(3.916)	(2.234)	(7.584)
Observations	9,164	8,952	8,952
R-squared	0.838	0.981	0.987
Durbin-Wu-Hausman test (prob > χ^2)	0.000	0.000	0.07
Exporter FE	Y	N	N
Importer FE	Y	N	N
Time FE	Y	Y	N
Exporter-time FE	N	N	Y
Importer-time FE	N	N	Y
Bilateral FE	N	Y	Y

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Standard errors are clustered at region pair level.

7. CONCLUSION

Using NUTS 2 level data for 200 EU regions between the years 2000-2010, this thesis investigates the extent to which bilateral trade flows are responsive to the level of transport infrastructure. The literature on trade and infrastructure provides an extensive overview of the issue and finds positive impacts of infrastructure on trade. However, there is no study to date that takes on a regional perspective whilst focusing on more than one country. This study aims to fill this gap in the literature by analyzing bilateral trade between the 200 regions of 21 EU member states.

To assess the effect of infrastructure on trade, following the literature, a gravity approach is used. As widely known, the use of this method is prone to many estimation problems. To overcome the endogeneity problems stemming from reverse causality, heteroskedasticity in error terms pertaining to the use of logarithms, potentially endogenous bilateral terms and omitted variable bias, numerous controls are presented. These include FE estimation, PPML approach, using lagged versions of the endogenous infrastructure variables, Hausman-Taylor IV method, and estimation with region-time and bilateral fixed effects.

Main results of the paper show varying but positive and significant impact of transport infrastructure on trade flows. The sign and significance of the variables of interest change as the estimation method is switched from OLS to FE, as the literature suggests. In the presence of fixed effects, the coefficients of motorways, railways, and double railways of source and destination regions are positive and significant at 1 per cent level and range from 0.05 to 0.13. When multilateral resistance terms are controlled for, only the interaction of the infrastructure variables can be identified and among them only the railways interaction bears the expected significant and positive sign. As for the heteroskedasticity controls, PPML method does not give estimates that are in line with the expectations stated earlier in the paper. When lagged versions of the variables are used, the results are very similar to the main results where the coefficients of infrastructure measures range between 0.05 and 0.12. In addition, the use of Hausman-Taylor IV approach and mountain typology index as an instrument for endogenous infrastructure variables shows that railways of source and destination region affect trade flows significantly.

Lastly, a check for robustness regresses trade flows on motorways and railways networks using a different measure of unit, per thousand square kilometers, based on data availability. The results of FE estimation are in line with the main results where motorways and railways significantly and positively contribute to trade flows. In sum, this robustness check suggests that the main findings are not a result of the choice of unit of measure for the variables of interest. Also, the most popular mode of transport in intra-EU trade, maritime transport efficiency is included as a control in addition to airport efficiency. Freight by sea is found to effect intra-EU trade positively, vindicating the facts presented in the third section.

To conclude, as different transport networks are included in the gravity model in addition to the standard gravity variables, international trade in the EU can be explained by the level of infrastructure. The analysis in this thesis brings about policy recommendations for emerging markets such as Turkey that would like to engage more in international and intranational trade. Given their geographical characteristics, Turkish regions may find it desirable to invest in quality railway infrastructure in addition to quality road infrastructure as suggested by the findings of this thesis and Coşar and Demir (2016) to increase their international and intraregional trade flows. Although small in magnitude, the results of this paper infer gains from infrastructure investment for the EU. Moreover, given the conclusions of the other analyses in the literature which focus on developing regions such as Asia and Africa, quality transport infrastructure can benefit Turkey and other developing countries to secure a better position in domestic and global trade. When the EU's geographically, economically, socially, and culturally integrated structure is taken into consideration, the results of this paper can be used as a reference point for developing countries to improve their transport infrastructure, especially railway networks if their geography allows.

REFERENCES

Akpan, U. (2014). Impact of Regional Road Infrastructure Improvement on Intra-Regional. *African Development Review, 26*(S1), 64-76.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review, 69*(1), 106-116.

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

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

Arkolakis, C., Costinot, A., & Rodríguez-Clare, A. (2012). New Trade Models, Same Old Gains? *American Economic Review, 102*(1), 94-130.

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics, 145*, 1-36.

Bensassi, S., Márquez-Ramos, L., Martínez-Zarzoso, I., & Suárez-Burguet, C. (2015). Relationship between logistics infrastructure and trade: Evidence from Spanish regional exports. *Transportation Research Part A, 72*, 47-61.

Bougheas, S., Demetriades, P. O., & Morgenroth, E. L. (1999). Infrastructure, transport costs and trade. *Journal of International Economics, 47*(1), 169-189.

Buys, P., Deichmann, U., & Wheeler, D. (2010). Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa. *Journal of African Economies, 19*(3), 399-432.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications.* New York: Cambridge Press.

Cheng, I. H., & Wall, H. J. (2005). Controlling for Heterogeneity in Gravity Models of Trade and Integration. *Federal Reserve Bank of St. Louis Review, 87*(1), 49-63.

Coşar, A., & Demir, B. (2016). Domestic road infrastructure and international trade: Evidence from Turkey. *Journal of Development Economics, 118*, 232-244.

Crescenzi, R., & Rodríguez-Pose, A. (2012). Infrastructure and regional growth in the European Union. *Papers in Regional Science, 91*(3), 487-513.

Deardorff, A. (1998). Determinants of bilateral trade: Does gravity work in a neoclassical world? *NBER Chapters, in: The Regionalization of the World Economy*, 7-32.

Démurger, S. (2001). Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China? *Journal of Comparative Economics, 29*, 95-117.

Donaubauer, J., Glas, A., Meyer, B., & Nunnenkamp, P. (2018). Disentangling the impact of infrastructure on trade using a new index of infrastructure. *Review of World Economics, 154*, 745-784.

Duranton, G., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. *Review of Economic Studies, 81*(2), 681-724.

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

European Commission. (2021). *History of NUTS.* Retrieved July 10, 2021, from Eurostat: https://ec.europa.eu/eurostat/web/nuts/history

Eurostat. (2021, July 7). *Freight transport statistics - modal split.* Retrieved July 18, 2021, from Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split#Modal_split_based_on_five_transport_modes:_road_competes_with_maritime_at_intra-EU_level

Feenstra, R. C. (2004). *Advanced International Trade: Theory and Evidence.* Princeton: Princeton University Press.

Francois, J., & Manchin, M. (2013). Institutions, Infrastructure, and Trade. *World Development, 46*, 165-175.

Fujimura, M., & Edmonds, C. (2006). Impact of Cross-border Transport Infrastructure on Trade and Investment in the GMS. *ADB Institute Discussion Paper, 48*, 1-35.

Head, K., & Mayer, T. (2014). Gravity Equations: Workhorse, Toolkit and Cookbook. *Handbook of International Economics, 4*, 131-195.

Hummels, D. L. (1999, January). *Toward a Geography of Trade Costs.* Retrieved November 2020, from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=160533

Hytten, T. (1931). Railway Policy as an Obstacle to Interstate Free Trade in Australia. *Weltwirtschaftliches Archiv, 34*, 195-211.

Ismail, N., & Mahyideen, J. (2015, December). *The Impact of Infrastructure on Trade and Economic Growth in Selected Economies in Asia.* Retrieved December 20, 2020, from ADBI Working Paper Series: https://www.think-asia.org/bitstream/handle/11540/9671/adbi-wp553.pdf?sequence=1

Lessmann, C., & Seidel, A. (2017). Regional inequality, convergence, and its determinants – A view from outer space. *European Economic Review, 92*(C), 110-132.

Limão, N., & Venables, A. J. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review, 15*(3), 451-479.

Martin, P., & Rogers, C. A. (1995). Industrial location and public infrastructure. *Journal of International Economics, 39*, 335-351.

McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *American Economic Review, 85*(3), 615-623.

Meyer, D. R. (1989). Midwestern Industrialization and theAmerican Manufacturing Belt in the Nineteenth Century. *The Journal of Economic History, 49*(4), 921-937.

Moneta, C. (1959). The Estimation of Transport Costs in International Trade. *Journal of Political Economy, 67*, 41-58.

Mundell, R. A. (1957). Transport costs in international trade theory. *The Canadian Journal of Economics and Political Science, 23*(3), 331-348.

Nordås, H. K., & Piermartini, R. (2004, August 10). *Infrastructure and Trade.* Retrieved November 2020, from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=923507

Persyn, D., Diaz-Lanchas, J., Barbero, J., Conte, A., & Salotti, S. (2020). A new dataset of distance and time related transport costs for EU regions. *Territorial Development Insights Series, JRC119412, European Commission*.

Puffert, D. J. (2000). The Standardization of Track Gauge on North American Railways, 1830-1890. *The Journal of Economic History, 60*(4), 933-960.

Santos-Silva, J. M., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics, 88*(4), 641-658.

Shepherd, B. (2016). Infrastructure, trade facilitation, and network connectivity in Sub-Saharan Africa. *Journal of African Trade, 3*(1-2), 1-22.

Stone, S., & Strutt, A. (2010). Transport Infrastructure and Trade Facilitation in the Greater Mekong Subregion. In D. Brooks, & S. Stone, *Trade Facilitation and Regional Cooperation in Asia* (pp. 156-192). Cheltenham: Edward Elgar.

Thissen, M., Lankhuizen, M., van Oort, F., Los, B., & Diodato, D. (2018). EUREGIO: The construction of a global IO database with regional detail for Europe for 2000-2010. *Tinbergen Institute Discussion Paper TI 2018-084/VI.*

Tinbergen, J. (1962). An Analysis of World Trade Flows. In *Shaping the World Economy.* New York, NY: Twentieth Century Fund.

Yotov, Y. V., Piermartini, R., Monteiro, J. A., & Larch, M. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model.* Geneva: World Trade Organization.

APPENDICES

Appendix A: Data Sources and Variable Definitions

The appendix introduces the data sources and variable definitions in further detail.

Variable name	Definition	Source
Trade flows$_{i,j,t}$	Total amount of sales from 15 categories of goods and services from region i to j at time t in million Dollars.	Thissen et al. (2018)
GDP$_{i,t}$	Gross Domestic Product realized in region *i* at time *t* both in Euros and million Euros.	Lessmann and Siedel (2017) for French regions, Eurostat for the rest of the countries
Bilateral distance$_{i,j}$	Arithmetic average of geodesic distance between the centroids of regions *i* and *j*.	Persyn et al. (2020)
Motorways$_{i,t}$	Motorways in region *i* at time *t* in kilometers or per thousand square kilometers depending on the analysis.	Eurostat
Railways$_{i,t}$	Total railway lines in region *i* at time *t* in kilometers or per thousand square kilometers depending on the analysis.	Eurostat
Double railways$_{i,t}$	Railway lines with double or more tracks in region *i* at time *t* in kilometers.	Eurostat

Other roads$_{i,t}$	Road infrastructure other than motorways or highways in region *i* at time *t* in kilometers.	Eurostat
Freight by air$_{i,t}$	Thousand tons of freight and mail transported by air in region *i* at time *t*.	Eurostat
Freight by sea$_{i,t}$	Thousand tons of freight transported by sea in region *i* at time *t*.	Eurostat
Border$_{i,j}$	=1 if regions *i* and *j* share a border also if *i* and *j* are the same regions, =0 if regions *i* and *j* do not share a border.	Author's own elaboration
Same$_{i,j}$	=1 if regions *i* and *j* are in the same country, =0 if regions *i* and *j* are in different countries.	Author's own elaboration
Mountain index$_{i}$	Mountain regions are defined as regions in which more than 50% of the surface is covered by topographic mountain areas, or in which more than 50% of the regional population lives in these topographic mountain areas. Lower scores correspond to more mountainous regions whereas non-mountain regions score a higher mountain typology index.	Eurostat - Regions and Cities Illustrated (RCI)

Appendix B: Name and Codes of EU Regions in NUTS 2 Level

The names and NUTS 2 codes of 200 EU regions covered in this thesis are presented in the table below.

NUTS 2 code	Region name	NUTS 2 code	Region name
AT11	Burgenland	FR52	Bretagne
AT12	Niederösterreich	FR53	Poitou-Charentes
AT13	Wien	FR61	Aquitaine
AT21	Kärnten	FR62	Midi-Pyrénées
AT22	Steiermark	FR63	Limousin
AT31	Oberösterreich	FR71	Rhône-Alpes
AT32	Salzburg	FR72	Auvergne
AT33	Tirol	FR81	Languedoc-Roussillon
AT34	Vorarlberg	FR82	Provence-Alpes-Côte d'Azur
BE10	Région de Bruxelles-Capitale	FR83	Corse
BE21	Prov. Antwerpen	GR11	Anatoliki Makedonia, Thraki
BE22	Prov. Limburg (BE)	GR12	Kentriki Makedonia
BE23	Prov. Oost-Vlaanderen	GR13	Dytiki Makedonia
BE24	Prov. Vlaams-Brabant	GR14	Thessalia
BE25	Prov. West-Vlaanderen	GR21	Ipeiros
BE31	Prov. Brabant wallon	GR22	Ionia Nisia
BE32	Prov. Hainaut	GR23	Dytiki Ellada
BE33	Prov. Liège	GR24	Sterea Ellada
BE34	Prov. Luxembourg	GR25	Peloponnisos
BE35	Prov. Namur	GR30	Attiki
CZ01	Praha	GR41	Voreio Aigaio
CZ02	Strední Cechy	GR42	Notio Aigaio
CZ03	Jihozápad	GR43	Kriti
CZ04	Severozápad	HU10	KozepMagyarorszag
CZ05	Severovýchod	HU21	Közép-Dunántúl
CZ06	Jihovýchod	HU22	Nyugat-Dunántúl
CZ07	Strední Morava	HU23	Dél-Dunántúl
CZ08	Moravskoslezsko	HU31	Észak-Magyarország
DE11	Stuttgart	HU32	Észak-Alföld
DE12	Karlsruhe	HU33	Dél-Alföld
DE13	Freiburg	ITC1	Piemonte
DE14	Tübingen	ITC2	Valle d'Aosta/Vallée d'Aoste
DE21	Oberbayern	ITC3	Liguria
DE22	Niederbayern	ITC4	Lombardia
DE23	Oberpfalz	ITD1	Provincia Autonoma di Bolzano

DE24	Oberfranken	ITD2	Provincia Autonoma di Trento
DE25	Mittelfranken	ITD3	Veneto
DE26	Unterfranken	ITD4	Friuli-Venezia Giulia
DE27	Schwaben	ITD5	Emilia-Romagna
DE30	Berlin	ITE1	Toscana
DE50	Bremen	ITE2	Umbria
DE60	Hamburg	ITE3	Marche
DE71	Darmstadt	ITE4	Lazio
DE72	Gießen	ITF1	Abruzzo
DE73	Kassel	ITF2	Molise
DE80	Mecklenburg-Vorpommern	ITF3	Campania
DE91	Braunschweig	ITF4	Puglia
DE92	Hannover	ITF5	Basilicata
DE93	Lüneburg	ITF6	Calabria
DE94	Weser-Ems	ITG1	Sicilia
DEA1	Düsseldorf	ITG2	Sardegna
DEA2	Köln	LT00	Lithuania
DEA3	Münster	LU00	Luxembourg
DEA4	Detmold	LV00	Latvija
DEA5	Arnsberg	MT00	Malta
DEB1	Koblenz	NL11	Groningen
DEB2	Trier	NL12	Friesland (NL)
DEB3	Rheinhessen-Pfalz	NL13	Drenthe
DEC0	Saarland	NL21	Overijssel
DED1	Chemnitz	NL22	Gelderland
DED2	Dresden	NL23	Flevoland
DED3	Leipzig	NL31	Utrecht
DEF0	Schleswig-Holstein	NL32	Noord-Holland
DEG0	Thüringen	NL33	Zuid-Holland
DK01	Hovedstaden	NL34	Zeeland
DK02	Sjælland	NL41	Noord-Brabant
EE00	Eesti	NL42	Limburg (NL)
ES11	Galicia	PL11	Lódzkie
ES12	Principado de Asturias	PL21	Malopolskie
ES13	Cantabria	PL22	Slaskie
ES21	País Vasco	PL31	Lubelskie
ES22	Comunidad Foral de Navarra	PL32	Podkarpackie
ES23	La Rioja	PL33	Swietokrzyskie
ES24	Aragón	PL34	Podlaskie
ES30	Comunidad de Madrid	PL41	Wielkopolskie

ES41	Castilla y León	PL42	Zachodniopomorskie
ES42	Castilla-la Mancha	PL43	Lubuskie
ES43	Extremadura	PL51	Dolnoslaskie
ES51	Cataluña	PL52	Opolskie
ES52	Comunitat Valenciana	PL61	Kujawsko-Pomorskie
ES53	Illes Balears	PL62	Warminsko-Mazurskie
ES61	Andalucía	PL63	Pomorskie
ES62	Región de Murcia	PT11	Norte
ES63	Ciudad de Ceuta	PT15	Algarve
ES64	Ciudad de Melilla	PT16	Centro (PT)
ES70	Canarias	PT17	Área Metropolitana de Lisboa
FI19	Länsi-Suomi	PT18	Alentejo
FI20	Åland	SE11	Stockholm
FR10	Île de France	SE12	Östra Mellansverige
FR21	Champagne-Ardenne	SE21	Sydsverige
FR22	Picardie	SE22	Norra Mellansverige
FR23	Haute-Normandie	SE23	Mellersta Norrland
FR24	Centre - Val de Loire	SE31	Övre Norrland
FR25	Basse-Normandie	SE32	Småland med öarna
FR26	Bourgogne	SE33	Västsverige
FR30	Nord-Pas-de-Calais	SK01	Bratislavský kraj
FR41	Lorraine	SK02	Západné Slovensko
FR42	Alsace	SK03	Stredné Slovensko
FR43	Franche-Comté	SK04	Východné Slovensko
FR51	Pays-de-la-Loire	UKN0	Northern Ireland